

**ALGEBRA 2 &**

**PRE-CALCULUS:**

**The Secrets**  
**behind the Steps**

## Contents

Update Notice .....	7
Introduction .....	8
Exponents.....	11
Conversion Factors and Exponents .....	13
Scientific Notation .....	15
Fractional Exponents.....	16
Exponent Practice .....	17
Roots .....	18
Square Roots .....	19
Simplifying Square Roots.....	21
Cube Roots .....	23
More Roots.....	24
Imaginary Numbers .....	24
Equations Containing Radicals .....	27
Factoring .....	30
Quadratic Equations.....	31
Simple Factoring Review .....	32
Finding Two Numbers that Multiply to P and Add to S .....	33
The Difference of Two Squares .....	35
The ac Method and Grouping .....	38
The Quadratic Formula .....	43
The Quadratic Formula and Factors.....	51
The Sum or Difference of Two Cubes.....	53
Factoring with Fractional Exponents.....	56
Calculator Programs .....	56

Functions.....	59
Domain, Range and Intercepts.....	60
Finding the Domain and Range .....	61
One-to-One Functions.....	64
Onto Functions.....	65
More About Functions .....	67
Even and Odd Functions.....	67
Inverse Functions .....	68
Composing Functions .....	71
Absolute Value Equations.....	73
Absolute Value Functions and Piecewise Functions .....	79
Inequalities.....	82
Graphing Inequalities .....	82
Solving Inequalities .....	84
Inequalities Involving Absolute Value .....	86
Tolerance Levels.....	89
Linear Inequalities .....	90
Function Graphs.....	95
Function Transformations .....	95
Finding the Vertex and the Axis of Symmetry.....	100
Order of Operations for Function Transformations.....	103
Polynomial Functions.....	107
Graphing Polynomial Functions .....	108
Number of Turns .....	109
The Intermediate Value Theorem .....	110
Long Division and Synthetic Division.....	110
<b>Finding the Value of <math>P(a)</math> Using Synthetic Division</b> .....	113
Finding the Zeros.....	114
Descartes' Rule of Signs .....	117

Rational Expressions and Functions.....	121
Adding, Subtracting, and Simplifying Rational Expressions .....	122
Rational Equations .....	125
Vertical Asymptotes .....	127
Range of a Rational Function .....	129
Horizontal Asymptotes (Limits at Infinity) .....	131
Slant Asymptotes .....	134
Graphing Rational Functions .....	135
Rational Inequalities.....	137
Partial Fraction Decomposition .....	140
Exponential Functions and Logarithms.....	144
Exponential Functions .....	145
Modified Exponential Functions .....	145
The Number $e$ .....	146
Compound Interest and Other Increases (and Decreases) .....	150
Exponential Equations.....	153
Logarithms.....	154
Change of Base Formula.....	158
Graphing Logarithmic Functions .....	159
Half-Life .....	162
Parabolas.....	165
Quadratic Equations and Parabolas .....	165
Basic Graphs .....	169
Shifted Parabolas.....	172
Side-Opening Parabolas .....	176
Circles and Ellipses .....	178
The Foci of an Ellipse .....	182
Equations of Circles and Ellipses .....	183
Eccentricity of an Ellipse.....	185

Deriving the Equation of an Ellipse (Optional) .....	186
The Area of an Ellipse (Optional) .....	187
Hyperbolas .....	189
Asymptotes of a Hyperbola .....	193
Shifting Hyperbolas .....	195
Deriving the Equation of a Hyperbola (Optional) .....	195
Reciprocal Functions .....	196
Sequences and Series .....	197
Arithmetic Sequences .....	197
Geometric Sequences .....	202
Arithmetic Series .....	206
Geometric Series .....	207
Probability and Statistics .....	211
Basic Statistics Terms .....	211
The Binomial Theorem .....	213
The Binomial Distribution .....	215
The Normal Distribution .....	219
Standard Deviation .....	220
Z-Scores .....	222
Margin of Error .....	224
Systems of Equations .....	226
Linear Programming .....	229
Matrices .....	231
Solving Systems of Linear Equations with Matrices .....	231
Matrix Multiplication .....	233
Determinants .....	236
Cramer's Rule .....	240
Identity Matrices .....	246
Inverse Matrices .....	247

Solving Systems with an Infinite Number of Solutions .....	250
Statistics .....	252
The Normal Curve and Z-Scores .....	252
Trigonometry .....	253
Inverse Trigonometric Functions .....	271
Secant, Cosecant, and Cotangent .....	274
The Law of Cosines .....	277
The Law of Sines .....	280
Trigonometric Identities .....	283
Trigonometric Equations .....	290
Modeling Change with Sine (or Cosine) Functions .....	293
Graphing a Modified Trigonometric Function .....	296
Heron's Formula .....	298
Equations of the Form $A \sin x + B \cos x$ .....	300
Complex Numbers .....	303
De Moivre's Theorem: Proof by Induction .....	305
Parametric Equations .....	309
Falling Objects .....	310
Projectiles .....	313
Polar Coordinates .....	315
Polar Graph Shapes .....	322
Polar Equations of Conics .....	323
Vectors .....	327
Adding or Subtracting Vectors .....	327
Multiplying a Vector by a Scalar .....	329
The Dot Product: Vectors and Work .....	329
Limits and Continuity .....	331
The Squeeze Theorem and a Special Limit .....	337
Trigonometric Limits .....	340

## Update Notice

Use only recently saved copies of this file. Download the latest version at <http://www.guidedcourses.com>.

If download is not available, or to report errors or typos in this document, contact [ylani@outlook.com](mailto:ylani@outlook.com)

Textbooks differ in the order in which they cover material. This document contains a navigation pane so you can look for your topic. Many document readers have an option to show the table of contents as a navigation pane. Alternatively, use the search function of your pdf reader to quickly find what you are looking for.

Last updated: September 2, 2020.

Don't see your topic? Check the Algebra 1 topics at <http://www.guidedcourses.com>.

## Introduction

Young children ask a lot of “why” questions. Why can’t we have candy for dinner? Why do I have to go to school? Why do I get boogers in my nose? Eventually they learn that they live in a society where things are done a certain way, and they should be asking less disruptive “how” questions. How questions are civilization-maintaining questions, and they depend on the civilization that you live in. How do you make fire? How do you turn wheat into bread? How do you download and install an app? These important questions guide the transfer of skills from one generation to the next, but not always in the best way. There is an old story about a daughter who asks her mother why she should cut off the end of a ham before roasting it. The mother replies that this is how she has always done it, and that was the way her mother used to do it. When they ask grandma about it, she says that she always cut off the end of the ham because she had a small roasting pan and the ham wouldn’t fit otherwise!

Why do you do it this way? Is there a better way to do it? Why do it at all? What if...? These are civilization-advancing questions that promote understanding and independent thinking. Unfortunately mathematics sometimes gets transferred from one generation to the next using only a “how to” approach. Even in college the only questions students ask in a calculus class are along the lines of “How do you do problem 98 in section 3.2?” One statistics professor actually told her students, “On the test, try to determine if it is a chapter 4 problem or a chapter 5 problem. If it is a chapter 4 problem, use the formulas from chapter 4 on your formula sheet. Otherwise, use the chapter 5 formulas.” Now that we can program computers to do math problems for us, this kind of thing is not just boring; it’s rather pointless. You can memorize steps, but how long will you remember them? And what will you really have learned after you do all 98 problems in section 3.2?

Yes, this book will show you how to do math problems. It will also encourage you to go behind the scenes to learn the secrets of why we can solve problems that way. One way to do that is by asking a lot of questions. It also helps to realize that creating math problems is not the exclusive domain of math textbook writers. It can actually be done by untrained people in their own home (provided they wear safety goggles). You can take a difficult math problem and make your own, easier version of it. Then you can experiment with this simpler problem to see how a method works, and why it works.

**How will I know if I have the right answer to a problem?** This is an important question. Look for the green color coding in the text to help you figure it out for different types of problems.



## When things get tough:

1. DO NOT PANIC! This stuff may seem hard, but it's math, not magic. There is always a logical reason for what is going on, and this book is here to help you find it.
2. Read slowly. Imagine us having a very leisurely conversation about the topic, and backtracking when things don't seem clear.
3. Replace "I don't get it" with "I don't believe it". That's what I always do because it puts the burden of proof on the textbook or the teacher rather than implying that I'm not smart 😊. Also, questioning and trying things out on your own is much better than just blindly believing anything that is printed.
4. For each topic, read only the sections you need. Courses vary a lot in what they contain and adding extra detail will rarely help you to understand the subject better. Spend your time learning the basics.
5. Still don't believe it? E-mail  
ylani@outlook.com 😊.

## Graphing

Please note that it is not necessary to buy a graphing calculator for Algebra 2 and Pre-Calculus. There are many free graphing apps for smartphones and computers. For your PC, you can download MathGV graphing software at <http://www.mathgv.com/> so you can experiment with different function graphs. It provides nicer and larger pictures than a calculator. To graph functions using MathGV, select File → New 2D Cartesian Graph. Then select Graph → New 2D Function. Alternatively, you can use one of many free online graphing calculators so there is nothing to download. Try searching for "geometry graphing" to find the most flexible program.

To graph functions using a TI-84 calculator, press "Y=" and enter your function. To insert an  $x$ , press the button labeled "X,T,θ,n". Press "GRAPH" to see the graph. If necessary, adjust the viewing rectangle by pressing "WINDOW".

## Text Color Coding:

**Summary** (Appears at the start for easy reference, read it later)

**Caution, or really important stuff**

**Examples** (you may have to create your own)

**Check your answers**

Within the text, negative numbers are indicated by a small minus sign, as in  $-4$ . Subtractions are indicated by a larger minus sign, as in  $x - 4$ . This is due to properties of the text editor rather than a choice I would prefer to make. A single minus sign is sufficient in mathematics to indicate either negative numbers or a subtraction. However, calculators may require you to use a separate minus sign for negative numbers, usually indicated by (  $-$  ) printed on the key.

192 0 0

51 51 255

0 153 102

102 0 204

## Exponents

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

$$(x^a)^b = x^{ab}$$

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

$$(xy)^a = x^a y^a$$

$$x^{-a} \text{ is } \frac{1}{x^a}$$

$$x^0 = 1$$

$$x^{\frac{1}{n}} = \sqrt[n]{x}$$

$$x^{\frac{a}{n}} = \sqrt[n]{x^a} \text{ or } \left(\sqrt[n]{x}\right)^a$$

Conversion factors may be raised to a power as needed, to match up with the units in your problem.

Scientific notation puts numbers in the form  $a \times 10^b$ , where  $1 \leq a < 10$ , and  $b$  may be a negative number.

Exponents indicate that something should be multiplied by itself a given number of times. While  $2x$  means 2 times  $x$ ,  $x^2$  means  $x$  times  $x$ . Exponents are the second item in “Please Excuse My Dear Aunt Sally”. This means that they have a very high priority in the order of mathematical operations. An exponent on something generally belongs exclusively to that one thing, unless there are parentheses present. So,  $3x^2$  is just  $3x \cdot x$ , because the exponent applies only to the  $x$ . However, once you place some parentheses they get priority, and the exponent applies to the entire thing inside those parentheses.  $(3x)^2$  means  $3x \cdot 3x$  which is  $9x^2$ .

First we will do some multiplication involving exponents, which is the easiest part.

$x^3 \cdot x^4 = ?$  This really means  $x \cdot x \cdot x$  times  $x \cdot x \cdot x \cdot x$ , which equals  $xxxxxxx$ . Now you have 7  $x$ 's in a row, which is  $x^7$ . Notice that the exponents simply add up.  $x^3 \cdot x^4 = x^7$ . This also means that  $x^5 \cdot x^5 = x^{10}$ .

Once mathematicians figure something like this out, they can't resist creating a formula, like  $x^a \cdot x^b = x^{a+b}$ . Don't be intimidated by such formulas that simply restate what was already

explained. Say something like, "Yeah, you just said that, duhh!" Then look it over carefully, just to... uh... make sure the mathematician hasn't made a mistake.

Division with exponents is not much harder. We already know that  $\frac{5x}{x} = 5$ , because if you multiply by  $x$  and then divide by  $x$ , you have effectively done nothing.  $\frac{5xx}{xx}$  is also 5. Just cross off one  $x$  above the line and one below the line, until all of the useless  $x$ 's are gone.  $\frac{xxx}{xx} = x$ . We

can also write this as  $\frac{x^3}{x^2} = x^1$ , which is just  $x$ . Another example:

$\frac{x^5}{x^3} = x^2$ , because we can remove 3  $x$ 's above the division line and 3  $x$ 's below the division line. Because of how this works, we can quickly divide by subtracting the exponents.

$\frac{x^7}{x^4} = x^{7-4} = x^3$ . Write this out completely and make sure that you believe it is correct. The general formula for this is of course [😊]:  $\frac{x^a}{x^b} = x^{a-b}$ .

Notice that using this method,  $\frac{x^1}{x^1}$  is  $x^{1-1}$  which is  $x^0$ . Since any number divided by itself is 1,  $x^0$  must mean 1. However  $x$  cannot be 0 here since we can't divide by 0.  $0^0$  is undefined.

Another thing that happens when you start subtracting exponents is that you end up with negative exponents. Consider  $\frac{x^2}{x^5}$ . You can write that as  $\frac{xx}{xxxxx}$ . Once you have removed two sets of  $x$ 's that are not doing anything, you end up with 3  $x$ 's below the line. Note that when all the  $x$ 's on top are gone there is still a 1 rather than just nothing:

$$\frac{xx}{xxxxx} = \frac{xx \cdot 1}{xxxxx} = \frac{1}{xxx}$$

If you just subtract the exponents that would leave you with  $x^{-3}$ . This tells you that  $x^{-3}$  means  $\frac{1}{x^3}$ . That's just the way things work out. So  $x^{-a}$  is  $\frac{1}{x^a}$ .

Because an expression like  $x^{-4}$  is really a fraction, you should treat it like one.  $\frac{x^{-4}}{5}$  means  $x^{-4}$  divided by 5:  $\frac{1}{x^4} \div 5 = \frac{1}{x^4} \div \frac{5}{1} = \frac{1}{x^4} \cdot \frac{1}{5} = \frac{1}{5x^4}$ .

$$\frac{6}{x^{-2}} \text{ means } 6 \div \frac{1}{x^2} = \frac{6}{1} \cdot \frac{x^2}{1} = \frac{6x^2}{1}.$$

As a shortcut, you might notice that you can move the part with the negative exponent to the other side of the division line to get a positive exponent. For example,  $\frac{x^{-3}}{2}$  is really  $\frac{1}{2x^3}$ , and if you actually do the division you will end up with  $\frac{1}{2x^3}$ . The net result is that  $x^{-3}$  moves to the bottom and gets a positive exponent. It works the same way if  $x^{-3}$  is on the bottom to start with:  $\frac{2}{x^{-3}}$  is really  $\frac{2}{\frac{1}{x^3}}$ , which works out to  $\frac{2x^3}{1}$  or just  $2x^3$ .

This not a substitute for understanding what is going on, but it does save a lot of time to just move the expression with the negative exponent to either the top or the bottom and remove the minus sign.

---

Multiplying with exponents causes us to have to add the exponents. You can also [trust mathematicians to think of this] put an exponent on something that already has an exponent. This is called raising a power to a power, and it looks like  $(x^2)^3$ . Thinking carefully about what this means, we rewrite it as  $x^2 \cdot x^2 \cdot x^2 = xx \cdot xx \cdot xx$  or  $x^6$ . Basically, we end up with  $x^{2 \cdot 3} = x^6$ . So, when you raise a power to another power, you multiply the exponents.  $(x^2)^3 = x^6$ .

Be careful when there are multiple things inside the parentheses.  $(xy)^3$  means  $xy \cdot xy \cdot xy = x \cdot x \cdot x \cdot y \cdot y \cdot y = x^3y^3$ .

## Conversion Factors and Exponents

Normally we don't include units like pounds or dollars in our algebra notation when there is only one type of unit, since it just clutters things up. But when you need to convert between units, you'll want to keep them around and work with them using regular algebra.

Suppose you have a value of 330 minutes, and you need to know how many hours that is. Start by creating a simple equation:

$$1 \text{ hour} = 60 \text{ minutes}$$

Divide both sides by 60 minutes:

$$\frac{1 \text{ hour}}{60 \text{ minutes}} = \frac{60 \text{ minutes}}{60 \text{ minutes}}$$

$$\frac{1 \text{ hour}}{60 \text{ minutes}} = 1$$

The fraction  $\frac{1 \text{ hour}}{60 \text{ minutes}}$  is called a conversion factor. You can read it as 1 hour per 60 minutes or 1 hour divided by 60 minutes. Note that we could also have divided both sides of the equation by 1 hour, to get  $1 = \frac{60 \text{ minutes}}{1 \text{ hour}}$ . Sixty minutes per hour is also a valid conversion factor, but it is not as useful in this particular situation. The value of a conversion factor is always 1. As you know, you can multiply any number, or even any unknown, by 1 without creating a change. Take the original value of 330 minutes and multiply it by the first conversion factor:

$$330 \text{ minutes} \times \frac{1 \text{ hour}}{60 \text{ minutes}}$$

$$\frac{330 \text{ minutes}}{1} \times \frac{1 \text{ hour}}{60 \text{ minutes}} = \frac{330 \text{ minutes} \cdot 1 \text{ hour}}{60 \text{ minutes}}$$

Now the unit minutes cancels out:

$$\frac{330 \cdot 1 \text{ hour}}{60} = \frac{330 \text{ hours}}{60} = \frac{330}{60} \text{ hours} = 5.5 \text{ hours}$$

If you arrange your equation so that you are both multiplying and dividing by the unit you want to get rid of, it will magically disappear. When you actually go to do this yourself, you might feel slightly confused if you accidentally use the wrong conversion factor:

$$330 \text{ minutes} \times \frac{60 \text{ minutes}}{1 \text{ hour}} = ??$$

Oops, that doesn't look right. We don't have something above and below the division line that we can cross off. In fact, we end up with a mess of  $19800 \text{ min}^2/\text{hr}$  🤪. This problem can easily be fixed by turning the conversion factor the other way around. Just as there are 60 minutes per hour, there is also 1 hour per 60 minutes, so that is the one you want:

$$330 \text{ minutes} \times \frac{1 \text{ hour}}{60 \text{ minutes}} = \frac{330 \text{ minutes} \times 1 \text{ hour}}{60 \text{ minutes}} = 5.5 \text{ hours}$$

You can also convert more complex units, such as miles per hour. If a high-speed train is travelling at 240 miles per hour, how fast is that per minute? Well, we can just use the conversion factor:

$$\frac{240 \text{ miles}}{1 \text{ hour}} \times \frac{1 \text{ hour}}{60 \text{ minutes}} = \frac{240 \text{ miles}}{60 \text{ minutes}} = \frac{4 \text{ miles}}{1 \text{ minute}}$$

So, the train is covering a distance of 4 miles in just 1 minute, and its speed is 4 miles/minute.

Thanks to the Internet, conversion factors are always readily available to you.

Conversion factors also work when you need more complicated units like square feet or cubic yards. For example, to find out how many square feet are in 3 square yards, we can use the conversion factor  $\frac{3 \text{ feet}}{1 \text{ yard}}$  and simply square it:

$$3 \text{ yards}^2 \times \left( \frac{3 \text{ feet}}{1 \text{ yard}} \right)^2 = 3 \text{ yards}^2 \times \frac{9 \text{ feet}^2}{1 \text{ yard}^2} = 27 \text{ ft}^2$$

Don't forget that when you square 3 feet you get 3 ft times 3 ft which equals **9** ft<sup>2</sup>, not 3 ft<sup>2</sup>. Draw a picture if necessary to see how that works. If the side of a square is 3 ft, the area is 9 ft<sup>2</sup>.

## Scientific Notation

Visit this website for a thorough explanation of scientific notation:

<http://ieer.org/resource/classroom/scientific-notation>

Scientific notation not only helps with very large numbers; it also gets rid of those potentially confusing zeroes at the beginning of numbers. If you write 0.03880 in scientific notation, it looks like this:

$$0.03880 = 3.880 \times \frac{1}{100}$$

Because  $\frac{1}{100}$  can be written as  $10^{-2}$ , the actual notation used is  $3.880 \times 10^{-2}$ . Now it is clear that 0.03880 has 4 significant figures. The zeros at the beginning of the number are not involved in showing how precise the measurement was.

If you know how to work with exponents, you can easily multiply scientific numbers. For example,  $(2 \times 10^{-5}) \times (6 \times 10^{-3}) = 2 \times 6 \times 10^{-5} \times 10^{-3} = 12 \times 10^{-8}$ . Because the first number is 10 or larger, we have to make it 10 times smaller, which means that  $10^{-8}$  has to be 10 times larger to keep the number the same:  $12 \div 10 = 1.2$ , and  $10^{-8} \times 10 = 10^{-8} \times 10^1 = 10^{-7}$ . The final answer is  $1.2 \times 10^{-7}$ .

## Fractional Exponents

What about exponents that are fractions? Do they exist? Let's consider  $x^{\frac{1}{2}}$ . We know that when we multiply things with exponents, we can just add the exponents. So,

$$x^{\frac{1}{2}} \cdot x^{\frac{1}{2}} = x^{\frac{1}{2} + \frac{1}{2}} = x^1.$$

This means that  $x^{\frac{1}{2}}$  has to represent a number that, when multiplied by itself, is  $x$ . The only candidate for this is  $\sqrt{x}$ , since  $\sqrt{x} \cdot \sqrt{x} = x$ . There is no other choice than to conclude that  $x^{\frac{1}{2}} = \sqrt{x}$ .

So what is  $x^{\frac{1}{3}}$ ? There would have to be a number such that  $x^{\frac{1}{3}} \cdot x^{\frac{1}{3}} \cdot x^{\frac{1}{3}} = x$ . This is the number we call the cube root of  $x$ , or  $\sqrt[3]{x}$ . By now you can probably guess that  $x^{\frac{1}{4}}$  is the fourth root of  $x$ ,  $\sqrt[4]{x}$ . Creating a general formula:  $x^{\frac{1}{n}} = \sqrt[n]{x}$  or the  $n^{\text{th}}$  root of  $x$ .

At some point you are going to see some very scary looking fractional exponents, like  $x^{\frac{2}{3}}$ . Do not despair! Just use your knowledge of fractions to see how such an exponent could have been created. We know that when we raise a power to a power, the two numbers are multiplied. That means that  $x^{\frac{2}{3}}$  could have been created in two ways:  $(x^2)^{\frac{1}{3}}$  or  $\left(x^{\frac{1}{3}}\right)^2$ .

Therefore,  $x^{\frac{2}{3}} = \sqrt[3]{x^2} = \left(\sqrt[3]{x}\right)^2$ . Both these forms mean the same thing. Because it is more trouble to use parentheses, you'll usually see  $x^{\frac{2}{3}}$  written as  $\sqrt[3]{x^2}$ . Notice that it is the denominator of the fractional exponent that determines what kind of root is involved.  $x^{\frac{3}{5}}$  means  $\sqrt[5]{x^3}$ , and  $x^{\frac{a}{n}}$  can be written as  $\sqrt[n]{x^a}$ .



Just remember that scary-looking exponents follow the same rules as regular exponents! If you have a complex problem like: simplify  $\frac{x^{-\frac{1}{4}}}{x^{\frac{1}{4}}}$ , go ahead and apply the regular rule for division which means subtracting exponents:  $-\frac{1}{4} - \frac{1}{4} = -\frac{1}{2}$ . Our answer is  $x^{-\frac{1}{2}}$  which means  $\frac{1}{x^{\frac{1}{2}}}$  or  $\frac{1}{\sqrt{x}}$ .

## Exponent Practice

1. Simplify:  $\frac{x^{-3}}{x^{-5}}$
2. Show that  $\frac{32(2x)^{-5}}{x^{-1}x^{-4}} = 1$
3. Show that  $27^{-1}x^{-14}y^{-12}(3x^4y^3)^4 = 3x^2$
4.  $5 \times 10^0 =$
5.  $25^4 = 5^?$
6. Complete the worksheets available at <http://ieer.org/resource/classroom/scientific-notation>

Check your answers to questions 4 and 5 with a calculator.

## Roots

$$\sqrt{x} \cdot \sqrt{x} = x \quad x^{1/2} \text{ means } \sqrt{x} \text{ and } x^{1/3} = \sqrt[3]{x}$$

$$x^{1/n} = \sqrt[n]{x} \quad x^{a/n} = \sqrt[n]{x^a}$$

$$\sqrt{ab} = \sqrt{a}\sqrt{b} \quad \text{and} \quad \sqrt{\frac{a}{b}} = \frac{\sqrt{a}}{\sqrt{b}}$$

To find the simplified square root of a number, break it down into its prime factors.

$$\sqrt[n]{ab} = \sqrt[n]{a}\sqrt[n]{b} \quad \text{and} \quad \sqrt[n]{\frac{a}{b}} = \frac{\sqrt[n]{a}}{\sqrt[n]{b}}$$

Caution:  $\sqrt{x^2} = |x|$

Use an absolute value sign for all even values of  $n$  in the expression  $\sqrt[n]{x^n} = |x|$

You can remove the square root from the denominator of a fraction:  $\frac{a}{\sqrt{b}} \cdot \frac{\sqrt{b}}{\sqrt{b}} = \frac{a\sqrt{b}}{b}$

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

Complex numbers obtained from the quadratic formula always come in pairs:  $a + bi$  and  $a - bi$ .

The magnitude (modulus) of a complex number,  $|a + bi|$ , is  $\sqrt{a^2 + b^2}$

### Radical Equations

Isolate the radical on one side of the equation, and then eliminate it.

Whenever you are taking a square root, or an even root on both sides, those roots could be either positive or negative. Account for that by putting  $\pm$  on the right side.

Check for extraneous solutions! Always put your solutions back into the original equation.

If the radical equation looks somewhat similar to a quadratic, you may be able to substitute  $x = (\sqrt{x})^2$ .

Get rid of confusing fractional exponents by raising both sides to the reciprocal power. Don't forget to use a  $\pm$  sign when the denominator of the fraction you are using is even!

The Latin word for root is “radix”, as in radishes which are the edible roots of a plant, so you’ll see roots referred to as radicals.

## Square Roots

By definition, the square root of a number is something that can be multiplied by itself to give that number:  $\sqrt{25}$  is 5, because  $5 \cdot 5 = 25$ . That also means that  $\sqrt{25} \cdot \sqrt{25} = 25$ , and in general:  $\sqrt{x} \cdot \sqrt{x} = x$ . If that doesn’t seem obvious, just try it out with a few real numbers and you’ll see that it always works like that.

While some numbers like 9 and 25 have nice integer square roots, most numbers do not. Just check some random numbers with your calculator to see that their square roots have a lot of numbers after the decimal point. In fact, if a number does not have an integer square root then its square root is irrational, which means that it cannot be written as a ratio of two integers. You see as many numbers after the decimal point as your calculator can display, but they really go on forever.

All positive numbers have both a positive and a negative square root. By convention  $\sqrt{9}$  means the positive square root of 9 which is 3. If we want to indicate the negative square root of 9 we write  $-\sqrt{9}$ . Negative numbers do not have real square roots. It is possible to take the square root of a negative number by using the imaginary number  $i$ , which is defined as  $\sqrt{-1}$ . The square of the imaginary number  $i$  is  $-1$ :  $i \cdot i = -1$ . We’ll take a closer look at square roots of negative numbers at the end of this chapter.

The first important equation for manipulating square roots is:

$$\sqrt{ab} = \sqrt{a}\sqrt{b} \quad (\text{for non-negative values of } a \text{ and } b)$$

By definition,  $\sqrt{ab} \cdot \sqrt{ab} = ab$ . We can easily verify that  $\sqrt{a}\sqrt{b}$  is actually the same as  $\sqrt{ab}$ , since  $\sqrt{a}\sqrt{b} \cdot \sqrt{a}\sqrt{b} = \sqrt{a}\sqrt{a}\sqrt{b}\sqrt{b} = ab$ .

Caution: this simple relationship leads some students to think that  $\sqrt{a+b} = \sqrt{a} + \sqrt{b}$ . You can confirm that this is not true by using some real numbers:  $\sqrt{9+16}$  is  $\sqrt{25}$ , which is not equal to  $\sqrt{9} + \sqrt{16}$ . You can also prove that it is not true in a general way.  $\sqrt{a+b} \cdot \sqrt{a+b} = a+b$ . However,  $(\sqrt{a} + \sqrt{b}) \cdot (\sqrt{a} + \sqrt{b}) = \sqrt{a} \cdot \sqrt{a} + \sqrt{a}\sqrt{b} + \sqrt{b}\sqrt{a} + \sqrt{b}\sqrt{b} = a + 2\sqrt{a}\sqrt{b} + b$ . Not the same thing.  $\sqrt{a+b} \neq \sqrt{a} + \sqrt{b}$ !

The second important equation involves fractions and square roots:

$$\sqrt{\frac{a}{b}} = \frac{\sqrt{a}}{\sqrt{b}} \quad (\text{for non-negative values of } a \text{ and } b)$$

Again, we can verify this relationship easily.  $\frac{\sqrt{a}}{\sqrt{b}} \cdot \frac{\sqrt{a}}{\sqrt{b}} = \frac{\sqrt{a}\sqrt{a}}{\sqrt{b}\sqrt{b}} = \frac{a}{b}$ . We can conclude that  $\frac{\sqrt{a}}{\sqrt{b}}$  must be the same as the square root of  $\frac{a}{b}$ ,  $\sqrt{\frac{a}{b}}$ . This last fact means that you can rewrite something like  $\frac{\sqrt{125}}{\sqrt{5}}$  as  $\sqrt{\frac{125}{5}} = \sqrt{25} = 5$ .

When we multiply, we add exponents:  $x^3 \cdot x^2 = x \cdot x \cdot x \cdot x \cdot x = x^5$ . Just keep that in mind when you are looking for something that multiplied by itself is equal to  $x$ , or  $x^1$ :

$$x^? \cdot x^? = x^1$$

There is only one thing that we can put for the question mark so that the exponents add to 1.  $x^{1/2} \cdot x^{1/2} = x$ . Therefore,  $x^{1/2}$  means  $\sqrt{x}$

When we take the square root of an unknown, like  $\sqrt{x}$ , we have to keep two things in mind: The value under the square root sign must be positive to start with (unless we allow the result to be an imaginary number), and the result is going to be positive. For example, if you are asked to simplify  $\sqrt{x^2}$ , the first condition is taken care of because  $x^2$  is always positive. However, the resulting square root is positive, even if  $x$  was negative to start with. By definition,  $\sqrt{x}$  means the positive square root of  $x$ . [The negative square root of  $x$  is written as  $-\sqrt{x}$ .] You might be tempted to write  $\sqrt{x^2} = x$ , but that doesn't work so well if  $x$  was, say,  $-3$ . **To make sure that you actually get a positive number you should use absolute value signs:**  $\sqrt{x^2} = |x|$ . Think of these lines as safety bars that make sure that the square root is positive. Use the safety bars all the time as a habit, and then just remove them if you find they are not needed.  $\sqrt{4x^6} = |2x^3|$ , but 2 is already positive so it doesn't need safety bars. Write  $\sqrt{4x^6} =$

$|2x^3| = 2|x^3|$ . Many times the absolute value sign can be removed completely:  $\sqrt{x^8} = |x^4| = x^4$  since  $x^4$  can never be negative.

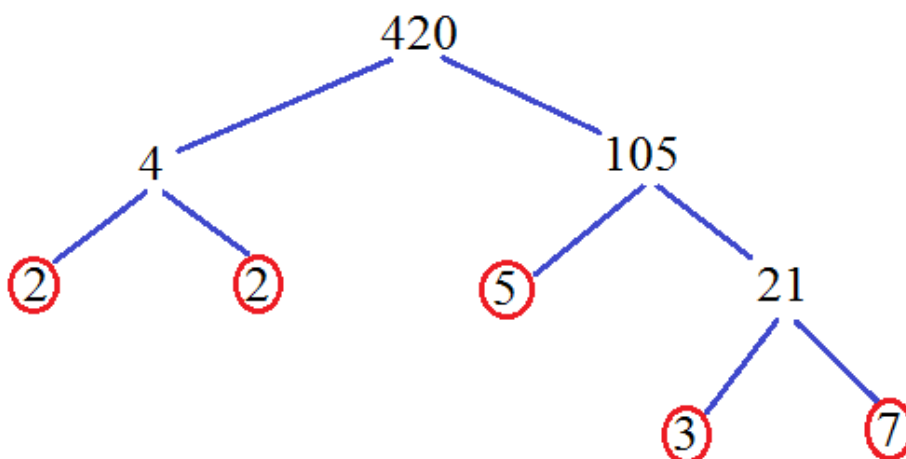
For most textbook problems the roots provided in the problem are considered to be real numbers rather than imaginary numbers. If you see something like  $\sqrt{x^3}$ , you can usually assume that  $x^3$  is positive or zero. If  $x^3$  isn't negative, then  $x$  cannot be a negative number.  $\sqrt{x^3} = \sqrt{x^2 \cdot x} = x\sqrt{x}$ . No absolute value sign is used because the question implies that  $x$  is positive.

## Simplifying Square Roots

Using the idea that  $\sqrt{ab} = \sqrt{a}\sqrt{b}$ , you can simplify an expression like  $\sqrt{50}$ . It goes like this:  
 $\sqrt{50} = \sqrt{25 \cdot 2} = \sqrt{25} \cdot \sqrt{2} = 5\sqrt{2}$ .

What if the question is a little harder, like: simplify  $\sqrt{420}$ ?

The tree method breaks down a number into its prime factors. The numbers spread out like branches of a tree. You can just divide without worrying about prime numbers. Here I started by dividing 420 by 4, but I could have divided by 2 first, or by 7, or by 10, or anything that works. The end result is always the same - just try it yourself to see that you get the same answer.



The prime factors are like the leaves at the end of the branches. Circle them when you are done and put them in order:

$$420 = 2 \cdot 2 \cdot 3 \cdot 5 \cdot 7$$

$$\sqrt{420} = \sqrt{2 \cdot 2 \cdot 3 \cdot 5 \cdot 7} = \sqrt{2^2 \cdot 3 \cdot 5 \cdot 7} = 2\sqrt{105}$$

Another example: Simplify  $\sqrt{1248}$

Break down the number 1248 into its prime factors:

$$1248 = 3 \cdot 416$$

$$416 = 2 \cdot 208$$

$$208 = 2 \cdot 104$$

$$104 = 2 \cdot 52$$

$$52 = 2 \cdot 26$$

$$26 = 2 \cdot 13$$

This shows that  $1248 = 3 \cdot 2 \cdot 2 \cdot 2 \cdot 2 \cdot 2 \cdot 13$ . That is the same as  $2^5 \cdot 3 \cdot 13$ .

So,  $\sqrt{1248} = \sqrt{2^5 \cdot 3 \cdot 13} = \sqrt{2^4 \cdot 2 \cdot 3 \cdot 13} = \sqrt{2^4} \cdot \sqrt{2 \cdot 3 \cdot 13} = 2^2 \cdot \sqrt{2 \cdot 3 \cdot 13}$ , which is  $4\sqrt{78}$ .

Note that the square root of  $2^4$  is  $2^2$ , because  $2^2 \cdot 2^2 = 2^4$ .

If you need  $\sqrt{2^6}$ , that would be  $2^3$ , since  $2^3 \cdot 2^3 = 2^6$ . Just cut any even exponent in half to get the square root. If the exponent is odd, split things up:  $\sqrt{2^9} = \sqrt{2^8 \cdot 2} = \sqrt{2^8} \cdot \sqrt{2} = 2^4 \cdot \sqrt{2}$ .

While it may look fine to us to write something like  $\frac{\sqrt{5}}{\sqrt{3}}$ , math teachers get very upset if you leave a root in the denominator of a fraction. For this reason you should always remove such a root as soon as you notice it. Fortunately that can be done easily by multiplying both the top

and the bottom of the fraction by the offending square root, as follows:  $\frac{\sqrt{5}}{\sqrt{3}} \cdot \frac{\sqrt{3}}{\sqrt{3}} = \frac{\sqrt{5} \cdot \sqrt{3}}{\sqrt{3} \cdot \sqrt{3}}$ .

Since  $\sqrt{3} \cdot \sqrt{3}$  is just 3, this can be written as  $\frac{\sqrt{5} \sqrt{3}}{3}$ . Remembering that  $\sqrt{a} \sqrt{b} = \sqrt{ab}$ , we simplify that to  $\frac{\sqrt{5 \cdot 3}}{3}$ . Now our answer is  $\frac{\sqrt{15}}{3}$ .

Do not try to get around having a square root in the denominator by writing  $\frac{\sqrt{5}}{\sqrt{3}}$  as  $\sqrt{\frac{5}{3}}$ . Leaving a fraction under a square root sign is also not acceptable to math teachers.

It is possible to remove a square root from the denominator of a fraction if it appears along with a regular number. For the expression  $\frac{\sqrt{x}}{4 + \sqrt{3}}$ , you can multiply the top and bottom by  $4 - \sqrt{3}$  to take advantage of the difference of two squares [ $a^2 - b^2 = (a + b)(a - b)$ ]:

$$\frac{\sqrt{x}}{4 + \sqrt{3}} \cdot \frac{4 - \sqrt{3}}{4 - \sqrt{3}} = \frac{\sqrt{x}(4 - \sqrt{3})}{4^2 - \sqrt{3}^2} = \frac{4\sqrt{x} - \sqrt{x}\sqrt{3}}{16 - 3} = \frac{4\sqrt{x} - \sqrt{3x}}{13}$$


---

Although it is **NOT** true that  $\sqrt{a + b} = \sqrt{a} + \sqrt{b}$ , regular math does continue underneath the square root sign. This may allow you to factor something out of an expression and then move it outside the square root sign, like this:  $\sqrt{4x + 100} = \sqrt{4(x + 25)}$ . Now you have two things under the square root that are multiplied by each other, so you can write  $\sqrt{4} \cdot \sqrt{x + 25}$ , which is  $2\sqrt{x + 25}$ .

## Cube Roots

Cube roots are easier to manage because we don't have to worry about positives and negatives. The number under the cube root sign can be either positive or negative, and the answer already has the correct sign automatically. The cube root of a positive number is always positive, and the cube root of a negative number is always negative. No absolute value signs are needed when you are taking the cube root of an unknown. You can manipulate cube roots using equations similar to those we used above for square roots:

$$\sqrt[3]{ab} = \sqrt[3]{a}\sqrt[3]{b} \quad \text{and} \quad \sqrt[3]{\frac{a}{b}} = \frac{\sqrt[3]{a}}{\sqrt[3]{b}} \quad \text{for all positive and negative values of } a \text{ and } b.$$

Cube roots can also be indicated by a fractional exponent. Because  $x^{1/3} \cdot x^{1/3} \cdot x^{1/3} = x$ , we say that  $x^{1/3} = \sqrt[3]{x}$ . You can rewrite the rules above using fractional exponents:

$$(ab)^{1/3} = a^{1/3} b^{1/3}, \text{ and } \left(\frac{a}{b}\right)^{\frac{1}{3}} = \frac{a^{\frac{1}{3}}}{b^{\frac{1}{3}}}$$

Cube roots can be simplified in the same way as square roots. For example,  $\sqrt[3]{250}$  can be written as  $\sqrt[3]{5 \cdot 5 \cdot 5 \cdot 2}$  or  $\sqrt[3]{5^3 \cdot 2}$ . We can split that up into  $\sqrt[3]{5^3} \cdot \sqrt[3]{2}$ , which simplifies to  $5\sqrt[3]{2}$ .

Math teachers don't like cube roots as denominators either. Remove them just like for square roots. For something like  $\frac{5}{\sqrt[3]{3}}$ , you can multiply the top and bottom by  $\sqrt[3]{3^2}$ . This has the advantage of turning the denominator into something with a nice cube root:

$$\frac{5}{\sqrt[3]{3}} \cdot \frac{\sqrt[3]{3^2}}{\sqrt[3]{3^2}} = \frac{5\sqrt[3]{3^2}}{\sqrt[3]{3^3}} = \frac{5\sqrt[3]{3^2}}{3}$$

## More Roots

Once you understand square roots and cube roots really well, you should not have problems with other roots. For all even values of  $n$ ,  $\sqrt[n]{x}$  works like square roots do. Make sure the value under the root sign is positive, and remember that the answer is always positive.  $\sqrt[4]{x^4} = |x|$ . Use an absolute value sign first, and then see if it is needed or not. For odd powers of  $n$ , you do not need to worry about signs. You can take the root of a negative number, and the absolute value sign is not needed in the answer because it will just be positive or negative as required.  $\sqrt[5]{x^5} = x$ .

As we saw earlier, roots can be represented by fractional exponents.  $\sqrt{x} = x^{1/2}$  and  $\sqrt[3]{x} = x^{1/3}$ . In general,  $\sqrt[n]{x} = x^{1/n}$ .

You can also convert radical expressions into fractional exponents. For example,  $\sqrt[3]{16} = \sqrt[3]{2^4} = 2^{4/3}$ . Sometimes this form is easier to work with. To simplify  $\sqrt[10]{3^4} \cdot \sqrt[5]{3^3}$ , use fractional exponents:  $\sqrt[10]{3^4} \cdot \sqrt[5]{3^3} = 3^{4/10} \cdot 3^{3/5} = 3^{2/5} \cdot 3^{3/5} = 3^{5/5} = 3$ .

When you have a 4<sup>th</sup> root of a number squared, you can convert that to a simple square root:  $\sqrt[4]{5^2} = (5^2)^{1/4} = 5^{2/4} = 5^{1/2} = \sqrt{5}$ . Just be careful when you do that with an unknown:  $\sqrt[4]{x^2} = (x^2)^{1/4} = \sqrt[4]{x^2}$ , since the original  $x$  could have been a negative number.

## Imaginary Numbers

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



Once you start using the quadratic formula to solve quadratic equations, square roots of negative numbers appear naturally. For a very long time people refused to even consider these roots, probably due to a general dislike of negative numbers. Finally  $\sqrt{-1}$  was defined as the **imaginary number  $i$** . This means that  $i \cdot i = -1$ .  $2i$  is  $2\sqrt{-1}$  and  $3i$  is  $3\sqrt{-1}$  and so on. These are valid numbers in their own right, and we can put them on a number line:

...,  $-4i$ ,  $-3i$ ,  $-2i$ ,  $-i$ ,  $0$ ,  $i$ ,  $2i$ ,  $3i$ ,  $4i$ , ...

When you need to take the square root of a negative number it helps to split the number up.

$$\sqrt{-4} = \sqrt{4 \cdot -1} = \sqrt{4} \cdot \sqrt{-1} = 2i$$

$$2i \cdot 2i = 2 \cdot 2 \cdot i \cdot i = 4 \cdot -1 = -4$$

The most common error students make is to say that  $\sqrt{-2}$  is  $2i$ . If you split the number up it is easier to see that this isn't true:

$$\sqrt{-2} = \sqrt{2 \cdot -1} = \sqrt{2} \cdot \sqrt{-1} = \sqrt{2} \cdot i$$

That last expression can be written as  $i\sqrt{2}$  or  $(\sqrt{2})i$  so you don't accidentally slip  $i$  underneath the square root sign.

Note that  $\sqrt{a}\sqrt{b} = \sqrt{ab}$  doesn't hold when both  $a$  and  $b$  are negative. Because we define the square root sign as representing the positive square root of a number only, we can't say that  $\sqrt{-4} \cdot \sqrt{-4}$  is equal to  $\sqrt{16}$ , because  $\sqrt{-4}$  is  $2i$ , and  $2i \cdot 2i$  is  $4i^2$ , which is  $-4$ . That gives us the negative square root of 16 rather than the positive one as indicated by the square root sign.

### Example

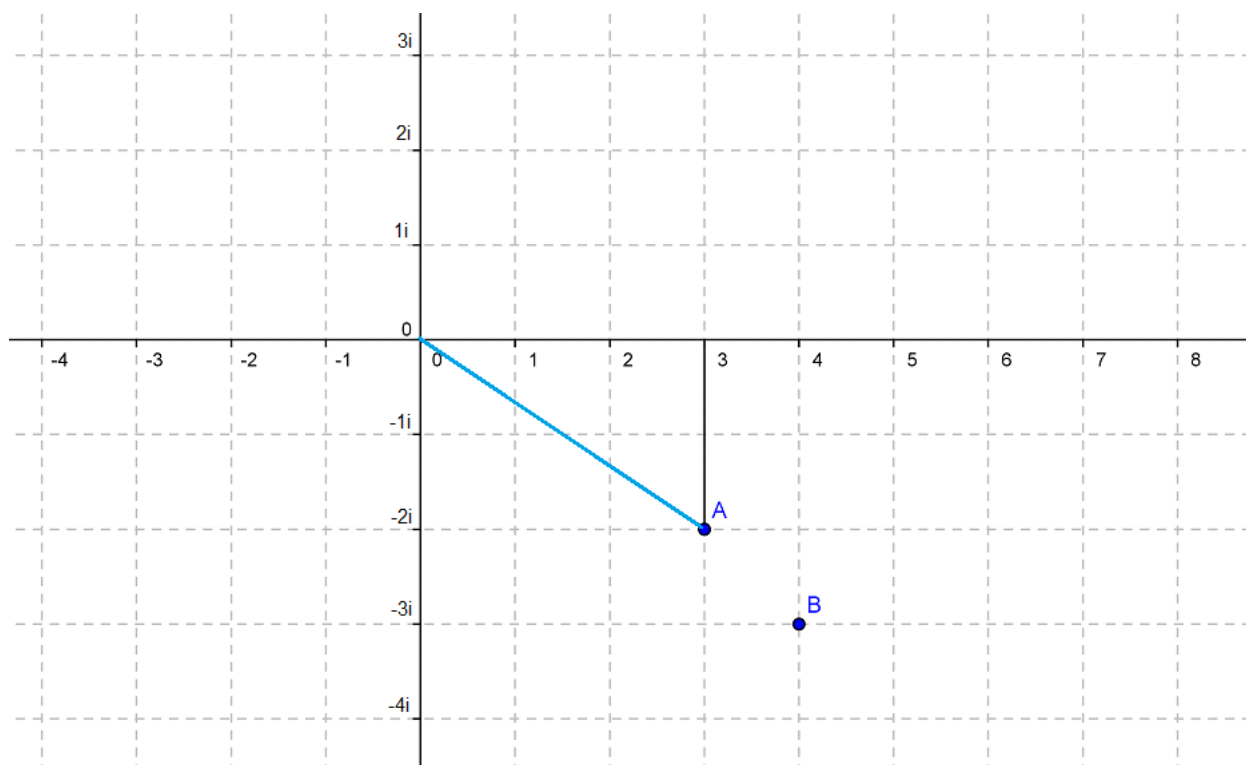
Simplify  $i^{14}$

By definition,  $i \cdot i = -1$ .  $i^4$  is  $i^2 \cdot i^2$ , so that would be  $-1 \cdot -1$ , which is just 1. Once you know that  $i^4 = 1$ , you'll want to use your knowledge of exponents to split up  $i^{14}$  into  $i^4 \cdot i^4 \cdot i^4 \cdot i^2$ . Now we can see that  $i^{14} = 1 \cdot 1 \cdot 1 \cdot -1 = -1$ .

---

“Imaginary” numbers turned out to have some very practical uses, especially in physics and electrical engineering. It is also fun to leave the boring regular number line and go off in a different direction. The most logical, and most useful, direction for that is at right angles to our plain number line. Having two different number lines at right angles allows us to create a number plane.

The quadratic formula often leaves us with numbers that look like  $3 + 2i$  and  $3 - 2i$ , or maybe like  $\frac{2}{5} + \frac{\sqrt{13}}{5}i$  and  $\frac{2}{5} - \frac{\sqrt{13}}{5}i$ . These numbers are called **complex numbers**. If you got them from the quadratic formula they always come in pairs due to the  $\pm$  sign in the formula. Complex numbers have a real part and an imaginary part, and we use those parts to place them on the number plane. Measure the real part along the real number axis, and the imaginary part along the imaginary number axis:



Here point A represents  $3 - 2i$ . Point B is the location of  $4 - 3i$

Just like for numbers on a number line, the “size” or **modulus** of these complex numbers is determined by the distance of the number from the origin. Since distance is always positive we need to use the absolute value sign. (The absolute value sign changes a negative number into a positive quantity. If the number inside the sign is positive or zero, it remains unchanged.)

$|3 - 2i|$  refers to the “size” of this number, which is its distance from the origin of the number plane. It is represented by the blue line in the image above. We find this distance by taking the square root of the squares of the individual components according to the Pythagorean Theorem:

$$|3 - 2i| = \sqrt{3^2 + |-2|^2} = \sqrt{13}$$

$$|4 - 3i| = \sqrt{4^2 + |-3|^2} = \sqrt{25} = 5$$

This shows that  $4 - 3i$  has a larger absolute value than  $3 - 2i$ , which makes sense because it is located further from the origin.

## Equations Containing Radicals

Because we use the expression  $\sqrt{x}$  to represent only the positive square root of  $x$ , we may occasionally find that we have *extraneous* solutions to our equations. This happens whenever the negative square root tries to go along for the ride, usually when we do a squaring operation to get rid of a square root.

Take for example the equation  $x + 2\sqrt{x} - 8 = 0$ . We can easily eliminate the confusing square root here by isolating it on one side of the equation and then squaring. [Why isolate the square root first? Try squaring both sides first so you can see what happens.]

$$x + 2\sqrt{x} - 8 = 0$$

$$2\sqrt{x} = 8 - x$$

Now square both sides:

$$(2\sqrt{x})^2 = (8 - x)^2$$

$$4x = 64 - 16x + x^2$$

Rearranging, we get  $x^2 - 20x + 64 = 0$ .

This factors to  $(x - 4)(x - 16) = 0$ , giving solutions of  $x = 4$  or  $x = 16$ . When you plug these solutions back into the original equation, you see that 16 doesn't work. If we allowed  $\sqrt{x}$  to

also represent the negative square root we could actually make it fit:  $16 + 2(-4) - 8 = 0$ , but that isn't how we have defined things. Mathematicians don't like surprises, so  $\sqrt{x}$  is always predictably positive.

The equation  $x + 2\sqrt{x} - 8 = 0$  is actually easier to solve by substitution because this method alerts us to the presence of the negative root. We substitute  $x = (\sqrt{x})^2$ . Now the equation looks like this:

$$(\sqrt{x})^2 + 2\sqrt{x} - 8 = 0$$

and we can solve it as a quadratic equation. This equation factors into  $(\sqrt{x} + 4)(\sqrt{x} - 2) = 0$ , which tells us that  $\sqrt{x} = -4$  or  $\sqrt{x} = 2$ . Since  $\sqrt{x}$  is never negative, we can immediately eliminate the first solution, so we know that  $\sqrt{x} = 2$  is the only answer. We can square both sides to get  $x = 4$ .

When your equation contains square roots, it is usually relatively easy to solve by squaring, although you may have to do that twice for the more complex ones. If there is a cube root, you can raise things to the 3<sup>rd</sup> power. You may also encounter equations with potentially confusing fractional exponents, but if you take the proper precautions you shouldn't find them too hard.

### Examples

$\sqrt{x} = 5$ , find the value of  $x$ .

$\sqrt{x} = 5$  means  $x^{\frac{1}{2}} = 5$ . You want  $x$ , which is really  $x^1$ , so  $\frac{1}{2}$  should be multiplied by 2:

$$\left(x^{\frac{1}{2}}\right)^2 = 5^2$$

$$x = 25$$

Now try this one:  $x^{\frac{2}{3}} = 4$ .

$x^{\frac{2}{3}} = 4$  means that if you take the cube root of  $x$ , and then square the result, you would get 4. It also means that you could square  $x$  first and then take the cube root, and the result would still be 4. (See "Fractional Exponents" in the Exponents chapter.) We can remove  $\frac{2}{3}$  by multiplying it by the reciprocal  $\frac{3}{2}$ . However, when we multiply by a fractional exponent with an even denominator, we're really taking an even root in order to solve the problem. In this case we are taking a square root, which you can see if we do it in two stages:

$$\left(x^{\frac{2}{3}}\right)^3 = 4^3$$

$$x^2 = 4^3$$

$$(x^2)^{\frac{1}{2}} = (4^3)^{\frac{1}{2}}$$

Whenever you are taking a square root on both sides, those square roots could be either positive or negative. Account for that by putting  $\pm$  on the right side:

$$x = \pm\sqrt{4^3}$$

Remember that you can also write  $\sqrt{4^3}$  as  $\sqrt[3]{4^3}$ . Either way the answer is  $x = \pm 8$ . **Check your work by plugging those answer back into the original equation.** The cube root of 8 is 2, and you can square that to get 4. You could also square 8 first, and then take the cube root of 64, which is 4 again. You can do the same with -8. The cube root of -8 is -2, which squares to give 4. Or, -8 squared is 64, and the cube root of that is 4. **To make sure you didn't miss any solutions, try graphing the equation as two functions:  $y = x^{2/3}$  and  $y = 4$ .**

Here is another example:

$$x^{\frac{4}{3}} = 81$$

$$\left(x^{\frac{4}{3}}\right)^{\frac{3}{4}} = \pm 81^{\frac{3}{4}}$$

Always remember to use a  $\pm$  sign when taking an even root of a number! To find the value of  $81^{\frac{3}{4}}$ , take the fourth root of 81, and then raise that to the 3<sup>rd</sup> power. You could do it the other way too, but I would rather not raise 81 to the third power first and then try to find the fourth root.  $\sqrt[4]{81} = 3$ , so  $\pm 81^{\frac{1}{4}} = \pm 3$ .

$$x^1 = \pm 81^{\frac{3}{4}}$$

$$x = \pm 3^3$$

$$x = \pm 27$$

# Factoring

## Factoring Summary

### First: Take out a common factor

Always check first to see if you can remove a common factor:

$$25x^2 + 10x = 5x(5x + 2).$$

$$x^3 + 8x^2 + 16x = x(x^2 + 8x + 16)$$

$$-x^2 + 4x - 4 = -1(x^2 - 4x + 4)$$

### Factor by Grouping

Specially designed polynomials can be factored by grouping. There are usually 4 or 6 terms:

$$3x^3 - 12x^2 - 4x + 16 = (3x^3 - 12x^2) - (4x - 16) = 3x^2(x - 4) - 4(x - 4) = (3x^2 - 4)(x - 4)$$

### The Difference of Two Squares

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

$$x^2 - 1 = (x + 1)(x - 1)$$

$$16x^2 - 1 = (4x + 1)(4x - 1)$$

$$x^4 - 16 = (x^2)^2 - 4^2 = (x^2 + 4)(x^2 - 4) = (x^2 + 4)((x + 2)(x - 2))$$

### Simple Factoring

There is nothing in front of  $x^2$ :

$$x^2 + 5x + 6 = (x + 3)(x + 2)$$

$$x^4 + 9x^2 - 8 = (x^2)^2 + 9(x^2) + 8 = (x^2 + 8)(x^2 - 1) = (x^2 + 8)(x + 1)(x - 1)$$

### ac Method

If there is a number in front of  $x^2$ , find two numbers that multiply to  $ac$  and add to  $b$ . Use those numbers to split up the middle term:

$$6x^2 - 7x - 5 = 6x^2 + 3x - 10x - 5 = (6x^2 + 3x) - (10x + 5) = 3x(2x + 1) - 5(2x + 1) = (3x - 5)(2x + 1)$$

### Sum or Difference of Two Cubes

$$a^3 + b^3 = (a + b)(a^2 - ab + b^2) \text{ or } a^3 - b^3 = (a - b)(a^2 + ab + b^2)$$

$$8x^3 + 27 = (2x + 3)(4x^2 + 6x + 9)$$

## Quadratic Equations

Rectangles have always been important to people. We build rectangular buildings and farm rectangular plots of land. Practical considerations prompted ancient mathematicians to start working with rectangles, and then they just naturally got carried away and discovered polynomial equations. How did this happen, and more importantly, why didn't they just leave things alone?

What matters to us about our rectangular buildings and fields is their usable area. If both the width and the length of a rectangle are unknown, we could call them  $x$  and  $y$  and say the area is  $xy$ . However, especially for buildings, we usually have some kind of relationship between the sides in mind before we start constructing. Imagine you are a mathematician working for the ancient king Mahotep I. The king wants a new audience hall with an area four times as big as the current one. The architect chooses to have the length of the hall twice as long as the width, and the required area is 5200 square cubits (a cubit is approximately the distance from your elbow to the tip of your middle finger). You call the width  $x$ , and the length  $2x$ . The area will be  $2x^2 = 5200$ , so  $x = \sqrt{2600}$ . You look up the square root of 2600 in your table of squares and think you are done. Then the architect informs you that he has to add an extra 2 cubits to the width to accommodate five large statues of the king along the western wall, but the required area will stay the same. Your new calculations are:  $(x + 2) \cdot 2x = 5200$ , or  $2x^2 + 4x = 5200$ . Eek, you have just discovered the quadratic equation, and now you are stuck solving it too! You divide by 2 to get  $x^2 + 2x = 2600$ , but now what? How can we find  $x$ ? Actually people were able

to do this a long time ago, and their methods are still used today.

## Simple Factoring Review

Factoring can help you find solutions to quadratic equations containing only one unknown value. These equations are named after their highest power term,  $x^2$  (x squared). The root "quadr" means 4 and "quadrus" means square, since a square has four sides. Originally these equations came from problems involving squares and rectangles. The general form of a quadratic equation is:

$$ax^2 + bx + c = 0 \quad \text{where } a, b \text{ and } c \text{ are constants.}$$

Now if you try to actually find the value of  $x$  for any quadratic equation, you can see that it is a bit of a challenge. Just try  $3x^2 + 5 - 2x^2 = -6x - 3$ . Quick, what is  $x$ ? Using factoring, we will see that there are actually 2 different answers that are both correct.

The first step in factoring a quadratic equation is to get it in a form you can recognize and deal with easily. If you have an equation like  $3x^2 + 5 - 2x^2 = -6x - 3$ , you'll never be able to factor it until you clean it up a little. It's sort of like cleaning up your room to help you find something you've lost. The equation rearranges to:  $x^2 + 6x + 8 = 0$ .

Next, we have to realize that many quadratic equations are the product of two simple factors, like  $(x + p)(x + q)$ , where  $p$  and  $q$  can be either positive or negative numbers. To see how this works, do the following multiplications:

$$(x + 3)(x + 1)$$

$$(x + 2)(x + 4)$$

$$(x - 5)(x - 6)$$

$$(x + 3)(x - 2)$$

In each case, the result is a quadratic expression of the form  $x^2 + \dots x + \dots$ . Usually mathematicians write these general quadratics as  $x^2 + bx + c$ . If you look closely at your answers, you'll see that  $b$  is always the sum of whatever numbers we picked for  $p$  and  $q$ , while  $c$  is always the product of  $p$  and  $q$ .

Watch what happens when we multiply just using  $p$  and  $q$ :

$$(x + p)(x + q) = x(x + q) + p(x + q) = x^2 + qx + px + pq$$



This can be written as  $x^2 + (p + q)x + pq$ . That's always how things work out. The middle term ends up as the sum of  $p$  and  $q$  [times  $x$ ], and the last term is the product  $pq$ .

For the equation  $x^2 + 6x + 8 = 0$  we need to look for  $p$  and  $q$  such that  $p + q$  is 6, and  $pq$  is 8. The only two numbers that fit this description are 2 and 4. It doesn't matter which one we call  $p$  and which one we call  $q$ , because  $(x + 2)(x + 4)$  is the same as  $(x + 4)(x + 2)$ .

So now we have changed  $x^2 + 6x + 8 = 0$  into  $(x + 2)(x + 4) = 0$ . At first you may not think that helps, but 0 has a remarkable property. If you take two numbers and multiply them to get 0, it means that at least one of those two numbers has to be 0. So either  $(x + 2) = 0$ , or  $(x + 4) = 0$ , or both of them are 0. This gives us two answers for the value of  $x$ :  $x = -2$ , or  $x = -4$ . **Plug these values into the original equation to see for yourself that they are both correct.**

**There is an interesting way to actually see the answers.** We can create the equation  $y = x^2 + 6x + 8$ . This gives many points  $(x, y)$  for which the equation is true. At the point where  $y = 0$  we should find our answers  $x = -2$  and  $x = -4$ . Use graphing software like [MathGV](#) or a graphing calculator to confirm this.

At this point you should stop to realize that while it is very helpful to rewrite  $x^2 + 6x + 8 = 0$  as  $(x + 2)(x + 4) = 0$ , it would not at all be useful to rearrange an equation like  $x^2 + 6x + 8 = 3$  into  $(x + 2)(x + 4) = 3$ . Here you cannot say that either  $x + 2$  must be 3, or  $x + 4 = 3$ , or they are both 3. **That trick only works with zero!** To solve  $x^2 + 6x + 8 = 3$ , first rewrite it as  $x^2 + 6x + 5 = 0$ , and then factor it.

Going back to the problem posed in the previous section, we would like to solve  $x^2 + 2x = 2600$ . That means that all we need to do is to rewrite the equation to say  $x^2 + 2x - 2600 = 0$ , and then find two numbers that multiply to -2600 and add to 2. Hmm...

## Finding Two Numbers that Multiply to P and Add to S

Most students have to spend a lot of time finding two numbers that multiply to some product  $P$  and add to some sum  $S$ . That isn't as hard as it seems at first, and you'll find it less frustrating if you follow a few simple rules.

1. Decide if you are looking for two positive numbers, two negative numbers, or one positive and one negative number.

For example, if two numbers multiply to 15 and add to -8, they would have to both be negative to make that work. If they multiply to 100 and add to 25, they must both be positive. If the two numbers multiply to -2600, there has to be one positive and one negative, regardless of what they add up to.

2. If one number is positive and the other is negative, decide which one should be “larger”.

If you want a positive sum, like 2, the positive number must be a bit larger than the negative one. If the sum should be negative, pick the larger number and give it a minus sign.

3. Work systematically.

Divide the product  $P$  by all possible factors in turn. **Always start with 1, so you don't overlook this somewhat obvious possibility!** If it is even, divide it by 2. Next, remember that a number is only divisible by 3 if the sum of its digits is divisible by 3. If you have divided the product by 2 and the result is still even, then you can divide it by 4. If you divide it by 3 and the result is even, it is also divisible by 6, and so on. Eventually you will reach a halfway point, after which the numbers repeat in reverse order, so you can stop. If you don't have the solution by then, you have either overlooked it, or it doesn't exist.

### Example

Factor  $x^2 - 20x + 84$ .

You need two numbers that multiply to 84 and add to -20. This means that both numbers have to be negative:

$$\begin{aligned} 84 = & -1 \cdot -84 \\ & -2 \cdot -42 \\ & -3 \cdot -28 \\ & -4 \cdot -21 \\ & -6 \cdot -14 \end{aligned}$$

Only the last two numbers add to -20. The quadratic expression  $x^2 - 20x + 84$  factors to  $(x - 6)(x - 14)$ .

### Example

Factor and solve  $x^2 + 2x - 2600 = 0$ .

Here we are looking for a positive and a negative number. To get a positive sum we have to make the positive number larger. The two numbers are close together in absolute value, so the answer should be at or near the end of the possible combinations. It may seem like there are a lot of numbers to try, but there are really a lot you can rule out. Since the digits of 2600 add to 8, the number is not divisible by 3. That means it is also not divisible by 6 since that is  $2 \cdot 3$ , or

any other multiple of 3. After we establish that it is not divisible by 7, any multiple of 7 is out too. 11 doesn't work, so neither will 22, 33 and 44, and so on. As shown below, a division by 8 results in an odd number, so that rules out 16 and 32. When you find that 17 doesn't work, you don't need to try 34, etc.

-2600 = -1 · 2600	= -13 · 200
-2 · 1300	-20 · 130
-4 · 650	-25 · 104
-5 · 520	-26 · 100
-8 · 325	-40 · 65
-10 · 260	-50 · 52

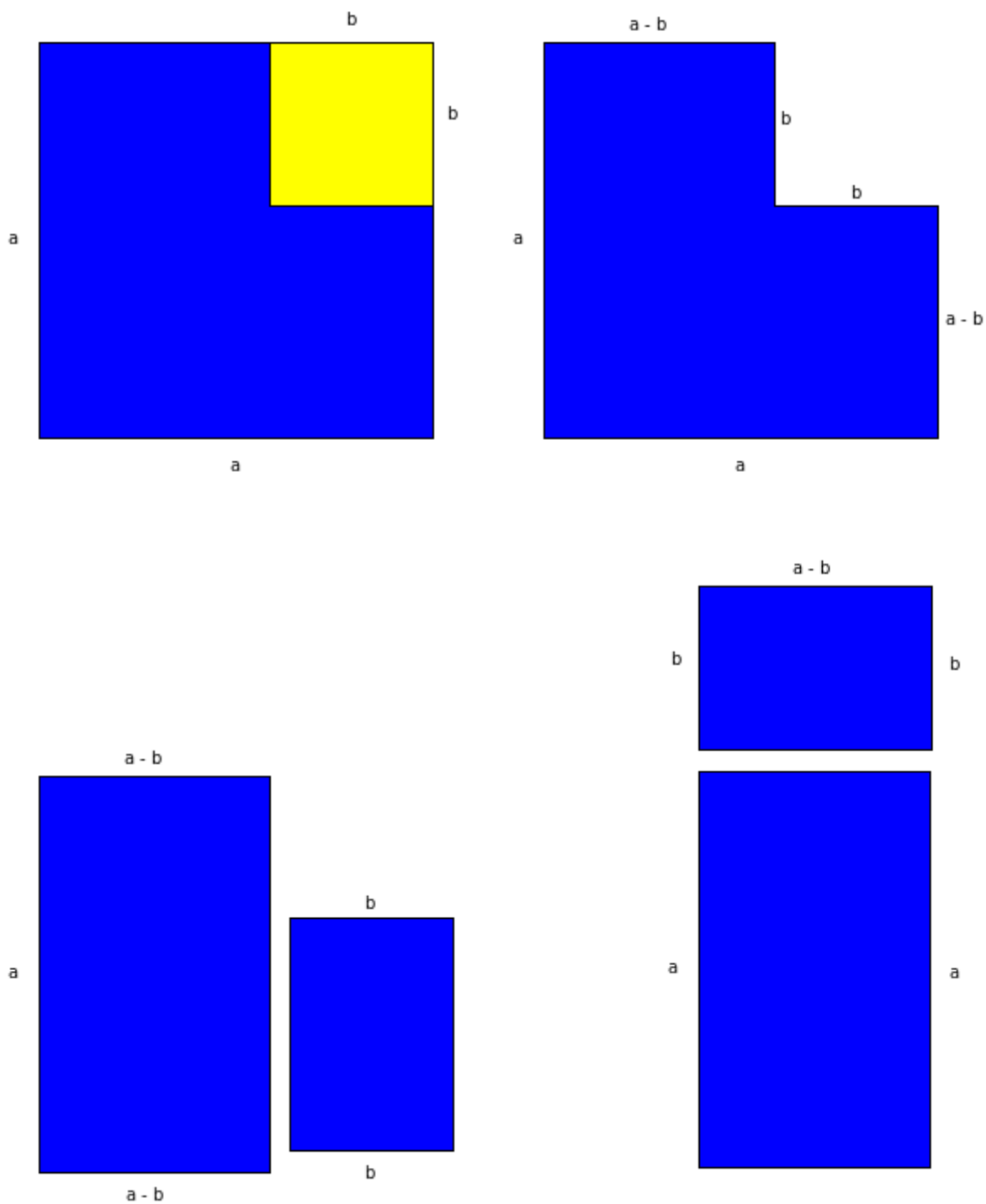
The very last combination is the one that works:  $x^2 + 2x - 2600 = 0$  factors into  $(x - 50)(x + 52) = 0$ , so  $x$  must be either 50 or -52. Because we got this problem from trying to construct a real building,  $x$  can only be 50 in this case.

When you already know that your two numbers must be very close in value, you can take a shortcut by considering what would happen if they were equal. In that case you would look at  $\sqrt{2600}$ , which is about 51. The most likely numbers will then be -50 and 52.

## The Difference of Two Squares

Sometimes your problem doesn't look like the standard quadratic equation, such as  $x^2 - 4 = 0$ . This equation can be solved by writing it as  $x^2 = 4$ , which means that  $x$  is 2 or -2. It can also be solved by factoring. We do this by using the difference of two squares, which you hopefully remember from Algebra 1. If not, here is a review:

We can represent the difference of two squares with a paper square. Draw a square with sides  $a$ . On one corner of this square cut out a smaller square with sides  $b$ , which is the yellow area in the picture. Get rid of the small square. That is it already:  $a^2 - b^2$ . To see that this is the same as  $(a + b)(a - b)$ , cut off one of the rectangular parts that is sticking out, and rotate it  $1/4$  turn. Now lay it against the remaining part to make a rectangle. The two sides of your rectangle are  $a + b$  and  $a - b$ , so its area is  $(a + b)(a - b)$ .



You can see that the area of the rightmost figure is  $(a + b)(a - b)$

This tells you that  $x^2 - 4 = (x + 2)(x - 2)$ . If  $(x + 2)(x - 2) = 0$ , your answers are also  $x = -2$  and  $x = 2$ . It is a good idea to create another graph [ $y = x^2 - 4$ ] to check these answers.

Another way to look at  $x^2 - 4$  is to write it as  $x^2 - 0x - 4$ . Now you could factor it by looking for two numbers that multiply to -4 and add to 0. Those numbers are 2 and -2, so the answer would still be  $(x + 2)(x - 2)$ . Even if you find this approach easier, you should still know the difference of two squares for more complex situations.

When you get to more difficult exams, the difference of squares may be disguised better, like  $1 - (x + 3)^2$ . Don't get confused; this is still a simple difference of two squares. It factors into  $(1 + (x + 3))(1 - (x + 3))$ . Pay careful attention to - signs as you change this to  $(x + 4)(-x - 2)$ .

There is also another way to solve  $1 - (x + 3)^2 = 0$ . You can square  $(x + 3)$  first, so the equation turns into  $1 - (x^2 + 6x + 9) = 0$ . Be careful now; everything in the parentheses has to be subtracted from 1. Doing this carefully, we get  $1 - x^2 - 6x - 9 = 0$ , which rearranges to  $-x^2 - 6x - 8 = 0$ . To get rid of the confusing - sign in front of  $x^2$ , we divide both sides by -1. Again be very careful; every term has to be divided by -1. We end up with  $x^2 + 6x + 8 = 0$ . Entirely by coincidence, this is the same equation we factored earlier in this chapter. Look back a few paragraphs to see that the answers are -4 and -2, which are the same answers we get by factoring this as a difference of two squares.

An even more clever way to disguise the difference of two squares is by squaring a square:

$$a^4 - 2^4$$

If we can recognize this as the difference of two squares we can factor it.  $a^4$  is really  $(a^2)^2$ , and  $2^4$  is  $(2^2)^2$ . Therefore, the equation factors to

$$a^4 - 2^4 = (a^2 + 2^2)(a^2 - 2^2) = (a^2 + 4)(a^2 - 4)$$

$$a^2 - 4 \text{ can be further factored into } (a + 2)(a - 2)$$

If you don't recognize the problem as either a difference of two squares, or as a standard quadratic equation, you can often still factor it.

Take the equation  $15x^2 + 5x = 0$  for example. The largest factor that  $15x^2$  and  $5x$  have in common is  $5x$ , so you can write the equation as  $5x(3x + 1) = 0$ . This means that either  $5x$  or  $3x + 1$  or both have to be zero. The answers are  $x = 0$  and  $x = -1/3$ .

## The ac Method and Grouping

Sometimes the quadratic equation will be presented to you with the middle term split up, like this:  $x^2 + 4x - 5x - 20 = 0$ . This is a subtle hint that you're expected to factor the equation by a method known as grouping. Grouping works by taking the four terms on the right side of the equation and putting them in two groups:  $x^2 + 4x - 5x - 20 = (x^2 + 4x) - (5x + 20)$ . **Be careful when you are placing the second set of parentheses to account for possible sign changes.**

Notice that one of the - signs had to be changed to a + sign. Next, factor each group separately to get  $x(x + 4) - 5(x + 4)$ . Because you have  $x$  times  $x + 4$  *and*  $-5$  times  $x + 4$ , you can change that to  $(x - 5)(x + 4)$  [verify that by multiplying  $(x - 5)(x + 4)$  back out]. Therefore the equation  $x^2 + 4x - 5x - 20 = 0$  factors to  $(x - 5)(x + 4) = 0$ , and  $x$  has to be 5 or -4. You'd get these same answers if you had added the middle terms up first.  $x^2 + 4x - 5x - 20 = 0$  is the same as  $x^2 - 1x - 20 = 0$ . Using the standard method you'd look for two numbers that add up to -1 and multiply to -20, and again you get  $(x - 5)(x + 4) = 0$ . Correct values for  $x$  are 5 and -4.

Some people prefer a more visual method to handle grouping. Starting with  $x^2 + 4x - 5x - 20$ , draw a square and place the four terms inside, making sure that the last term is diagonally opposite the first:

$x^2$	$-5x$
$4x$	$-20$

Next, take out the greatest common factors both vertically and horizontally, in the same way that we did before. Place the factors above and beside the square like this:

	$x$	$-5$
$x$	$x^2$	$-5x$
$4$	$4x$	$-20$

Notice that each little square is now the product of the terms on top and to the right:  $x^2$  is the product of  $x$  and  $x$ ,  $-5x$  is the product of  $x$  and  $-5$ ,  $4x$  is the product of  $4$  and  $x$ , and  $-20$  is the product of  $-5$  and  $4$ . Add up the parts outside the square to get  $(x + 4)(x - 5)$ .

Try this method with the other problems shown in this section, so you can see if you like it better.

Grouping may also work on cubic expressions or equations, so try it if there are 4 terms and nothing else to do. For example, factor  $2x^3 + 3x^2 - 18x - 27$ . First place your parentheses carefully:  $(2x^3 + 3x^2) - (18x + 27)$ . Next, take out the largest common factors to get  $x^2(2x + 3) - 9(2x + 3)$ , which equals  $(x^2 - 9)(2x + 3)$ . Note that not all cubic polynomials can be factored this way.

Rarely, you may see an expression with 6 terms that can be factored by grouping.

There may be a number in front of  $x^2$ , as in  $2x^2 + 10x + 8 = 0$ . The first thing to do is to see if you can get rid of it. In this particular case you can divide the entire equation by 2 to get  $x^2 + 5x + 4 = 0$ . This factors as  $(x + 4)(x + 1) = 0$ , and the answers are  $-4$  and  $-1$ . If there is no equals sign, and you are just asked to factor  $2x^2 + 10x + 8$ , put the 2 in front:  $2x^2 + 10x + 8 = 2(x + 4)(x + 1)$ .

Potentially confusing is the presence of a minus sign in front of the  $x$ , as in  $-x^2 + x + 2$ . Fortunately this is easy to remove; just factor out  $-1$ :

$$-x^2 + x + 2$$

$$-1(x^2 - x - 2)$$

$$-(x + 1)(x - 2)$$

Usually you cannot remove the number in front of  $x^2$ . An equation like  $6x^2 + 17x + 5 = 0$  would be very hard to solve by guessing. We need a more systematic approach.

Let's start by seeing how such a complex equation is created. The real factors of  $6x^2 + 17x + 5$  are  $(2x + 5)$  and  $(3x + 1)$ , so we'll multiply those out to see how they generate the left side of the equation:

$$(2x + 5)(3x + 1)$$

$$2x(3x + 1) + 5(3x + 1)$$

$$6x^2 + 2x + 15x + 5$$

$$6x^2 + 17x + 5$$

Notice that if only we could take  $6x^2 + 17x + 5$  and go back to the step just before that,  $6x^2 + 2x + 15x + 5$ , we could factor by grouping like this:

$$6x^2 + 2x + 15x + 5$$

$$(6x^2 + 2x) + (15x + 5)$$

$$2x(3x + 1) + 5(3x + 1)$$

$$(2x + 5)(3x + 1)$$

Aaggh, so close! But how could we know that the middle term of  $6x^2 + 17x + 5$  was created by adding  $2x$  and  $15x$ ? Maybe it was  $6x^2 + 16x + x + 5$  instead. Does it matter? Well, let me try to factor  $6x^2 + 16x + x + 5$  by grouping:

$$6x^2 + 16x + x + 5$$

$$(6x^2 + 16x) + (x + 5)$$

$$2x(3x + 8) + 1(x + 5)$$

Now the two things in the parentheses are different. It is not possible to combine them into two factors. If you don't split the middle term up in just the right way things just don't work out. To see if we can recover those two numbers that should make up the coefficient of the middle term, we'll use the general factors  $(mx + p)$  and  $(nx + q)$  [where  $m$  and  $n$  are letters that stand for random numbers just like  $p$  and  $q$ ]

Multiplying these factors gives:



$$(mx + p)(nx + q)$$

$$mx(nx + q) + p(nx + q)$$

$$mnx^2 + mnx + npn + pq.$$

Since  $mq$  and  $np$  are just numbers, that would be written as:

$$mnx^2 + (mq + np)x + pq$$

Comparing this with the actual quadratic,  $6x^2 + 17x + 5$ , we see that  $(mq + np) = 17$ . The numbers that we need to split up the middle term correctly are  $mq$  and  $np$ , but how to find them?

Every quadratic expression can be written in the form  $ax^2 + bx + c$ , where  $a$ ,  $b$  and  $c$  are positive or negative numbers, or  $b$  and/or  $c$  could be 0. For  $mnx^2 + (mq + np)x + pq$ , the number  $a$  is  $mn$ , and the number  $c$  is  $pq$ . When we multiply those two numbers, we get  $mnpq$ . However, that is also what you get when you multiply  $mq$  and  $np$ , the two middle numbers we want to find. This tells you that if you multiply the first number and the last number in your quadratic, you get a number that the two numbers you want will multiply to. For  $6x^2 + 17x + 5$ , multiplying the first and the last number gives you  $6 \cdot 5 = 30$ . So, to find  $mq$  and  $np$ , look for two numbers that add to 17 and multiply to 30. After a bit of trial and error you would determine that these numbers are 2 and 15. Now we use these two numbers to work backwards and split up the middle term. Write  $6x^2 + 17x + 5$  as  $6x^2 + 2x + 15x + 5$ , or  $6x^2 + 15x + 2x + 5$ .

When we write an equation as  $mnx^2 + mnx + npn + pq$ , we can factor it by grouping in two ways:

$$(mnx^2 + mnx) + (npn + pq) = mx(nx + q) + p(nx + q) = (mx + p)(nx + q) \text{ or}$$

$$(mnx^2 + npn) + (mnx + pq) = nx(mx + p) + q(mx + p) = (nx + q)(mx + p)$$

We can do the same thing with  $6x^2 + 17x + 5$  in a less abstract way, using the numbers 2 and 15 that we found by trial and error:

$$6x^2 + 2x + 15x + 5 = 2x(3x + 1) + 5(3x + 1) = (2x + 5)(3x + 1)$$

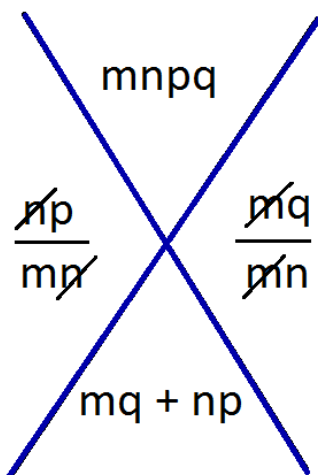
$$6x^2 + 15x + 2x + 5 = 3x(2x + 5) + 1(2x + 5) = (3x + 1)(2x + 5)$$

For a quadratic that has a number in front of the  $x^2$ , like  $6x^2 + 17x + 5$ , the best strategy is to find two numbers that add up to the middle number, 17, and multiply to  $6 \cdot 5$ , the product of the first and the last number. This method is called the **ac method** of factoring because we

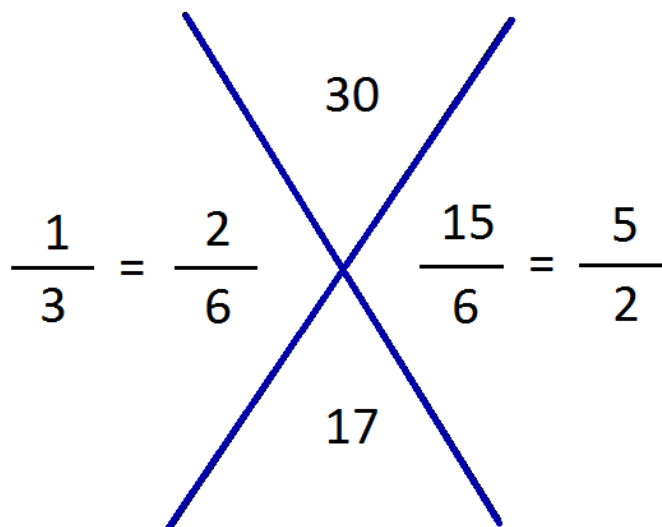
multiply a and c in the expression  $ax^2 + bx + c$ . Although you can factor simpler quadratics where  $a = 1$  this way too, that is more work than you need to do!

Now to solve the original equation. We have seen that  $6x^2 + 17x + 5 = 0$  is the same as  $(3x + 1)(2x + 5) = 0$ . This means that  $3x + 1 = 0$  or  $2x + 5 = 0$ , giving answers of  $x = -1/3$  or  $x = -5/2$ .

There is a shortcut used by some teachers, although it is hard to remember long term. Learn it if you have to, otherwise don't bother.



You multiply the first and last numbers to get  $mnpq$  and put that in the top part of the X. Put the middle number, which represents  $mq + np$  in the bottom part of the X. Find your two numbers  $np$  and  $mq$ , and place them on the sides as shown, above the first number  $mn$ . Then simplify the fractions and you will have  $p$ ,  $m$ ,  $q$  and  $n$ . Here is how it works:



$$m = 3, p = 1, n = 2 \text{ and } q = 5: (3x + 1)(2x + 5)$$

## The Quadratic Formula

Usually you first learn to solve quadratic equations by changing them to a form where there is a zero on one side, and then factoring them. This method takes advantage of the fact that if  $a \cdot b = 0$ , then either  $a = 0$  or  $b = 0$  or both  $a$  and  $b$  are equal to zero. For example, if  $5 \cdot b = 0$  there is no other choice;  $b$  must be 0. We write the quadratic as  $(\dots)(\dots) = 0$  and solve the parts separately. Because your textbook carefully selected suitable problems for you, you may not have noticed that only some quadratic equations can be solved this way.

A quadratic equation like  $x^2 - 6x = -4$  can be easily changed to  $x^2 - 6x + 4 = 0$ , but now you can't factor it! There are no whole numbers that add to -6 and multiply to 4. That doesn't mean that there are no solutions, but the solutions are not nice numbers.

Use a graphing app like Desmos, or a graphing calculator to draw a graph of  $y = x^2 - 6x + 4$ . You can see that this graph is a nice parabola, and it intersects the x-axis at two points. At those points,  $y$  is equal to zero, so those are the solutions. The exact value of  $x$  at these points is hard to see, but we can calculate it.

One way to do so is by realizing that the graph is very symmetrical. If you draw a vertical line through the middle you can see that the solutions are at equal distances from the lowest point of the graph (the vertex of the parabola). The x-coordinate of the vertex is right in the middle of the x-coordinates of the solutions.  $x^2 - 6x + 4 = 0$  can still be factored into  $(x + p)(x + q) = 0$ . This equation is true when  $x = -p$  and/or when  $x = -q$ . The x-coordinate of the vertex is a number that is the average of the roots,  $-p$  and  $-q$ .

$$(x + p)(x + q) = x^2 + qx + px + pq$$

$$(x + p)(x + q) = x^2 + (p + q)x + pq$$

$$x^2 + 6x + 4 = x^2 + (p + q)x + pq$$

$p + q = -6$  in this case, meaning that  $-p + -q = 6$ . To get the average of  $-p$  and  $-q$  we divide that sum by 2 to get 3. Look at the graph to see that the x-coordinate of the vertex is in fact 3. You can see that more clearly by adding the line  $x = 3$  to your graph. Our solutions lie at equal distances from  $x = 3$ , but we still don't know what this distance is. Let's just call it  $d$ . The solutions will be  $3 + d$  and  $3 - d$ . Our clue to solving this is that the last number in the

quadratic, 4, is the product of p and q. Since two minus signs cancel in multiplication, 4 is also the product of -p and -q. We can just multiply the two solutions to get 4:

$$(3 + d)(3 - d) = 4$$

$$3^2 - 3d + 3d - d^2 = 4$$

$$9 - d^2 = 4$$

$$-d^2 = 4 - 9$$

$$d^2 = 5$$

$$d = \sqrt{5}$$

The solutions are  $x = 3 + \sqrt{5}$  and  $x = 3 - \sqrt{5}$

In general, we can solve a quadratic of the form  $x^2 + bx + c = 0$  in the same way. Again, it will factor into  $(x + p)(x + q) = 0$ . The average of the solutions is the sum of -p and -q divided by 2.

Because  $b = p + q$ , we use -b and divide by 2. The solutions are  $\frac{-b}{2} + d$  and  $\frac{-b}{2} - d$ .

$$\left(\frac{-b}{2} + d\right)\left(\frac{-b}{2} - d\right) = c$$

$$\left(\frac{-b}{2}\right)^2 - d^2 = c$$

$$-d^2 = -\left(\frac{-b}{2}\right)^2 + c$$

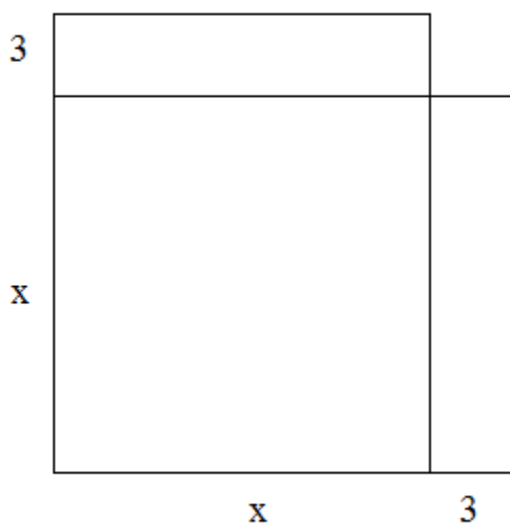
$$d^2 = \left(\frac{-b}{2}\right)^2 - c$$

$$d = \sqrt{\left(\frac{-b}{2}\right)^2 - c}$$

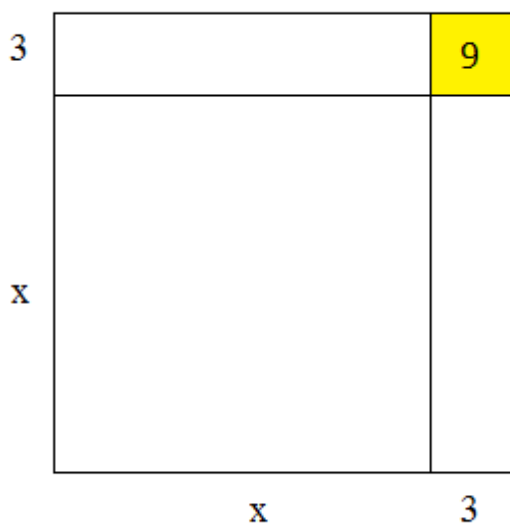
The solutions are  $\frac{-b}{2} + \sqrt{\left(\frac{-b}{2}\right)^2 - c}$  and  $\frac{-b}{2} - \sqrt{\left(\frac{-b}{2}\right)^2 - c}$

The quadratic formula can also be derived by completing the square. Many thousands of years ago the ancient Babylonians discovered this method by actually drawing squares to represent the squaring of numbers. The purpose of doing so was to find the area of the square, which would tell you the length of the sides, and therefore the value of the unknown  $x$ . It looked like this:

Suppose that you have an equation like  $x^2 + 6x = 91$ . You draw a corresponding picture that has a square with area  $x^2$  and two rectangles with an area of  $3x$  each:



From the original problem, you know that you have just drawn a figure with a total area of 91. To complete the square, you have to add a little piece with an area of 3 times 3, or 9. Once you have done that, your figure will be a perfect square with a total area of  $91 + 9$ , or 100.



Now that we know the area of the completed square, the solution is obvious just by looking at the picture. The only value for  $x$  that would give a completed square with a total area of 100 is  $x = 7$ . For cases where it is not immediately obvious, write out what you did:

$$x^2 + 6x = 91$$

$$x^2 + 3x + 3x = 91$$

$$x^2 + 3x + 3x + 9 = 91 + 9$$

$$x^2 + 6x + 9 = 100$$

The sides of the completed square are  $x + 3$  long, so its area is  $(x + 3)^2$

$$(x + 3)^2 = 100$$

$$(x + 3) = 10$$

$$x = 7$$

Because we are looking at a real square, we cannot consider the negative square root, -10, which would lead us to a value for  $x$  of -13. Thousands of years ago people either ignored these negative values or considered them evil. Today we know we can work with them, so the old simple method of completing an actual square has fallen out of favor, but the idea is still useful.

For a simple quadratic equation like  $x^2 + 8x - 20 = 0$ , we can find  $x$  by factoring:

$$(x + 10)(x - 2) = 0, \text{ so } x = -10 \text{ or } x = 2.$$

Completing the square can also be done easily by rewriting the equation as  $x^2 + 8x = 20$ . Draw a square with sides  $x$  to represent  $x^2$ . Then split the number in front of the  $x$  term, 8, in half to

add two pieces with area  $4x$  to your square. Now you can see that a square piece with an area of 16 (sides of length 4) is needed to complete the square. Add 16 to both sides to get

$$x^2 + 8x + 16 = 20 + 16.$$

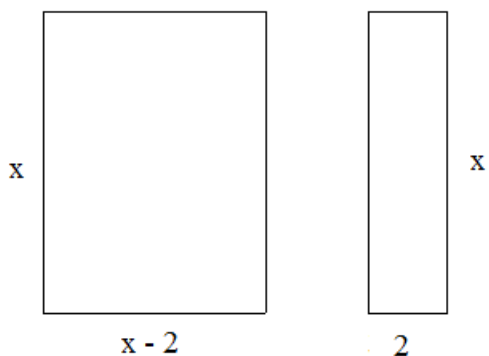
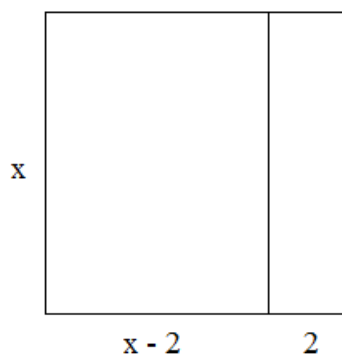
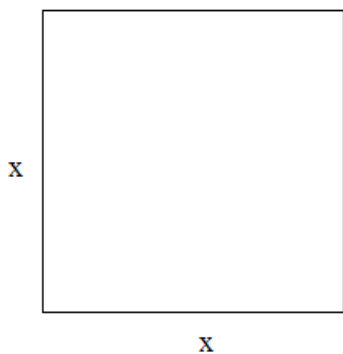
$$x^2 + 8x + 16 = 36$$

$$(x + 4)^2 = 36$$

$$x + 4 = \pm\sqrt{36}$$

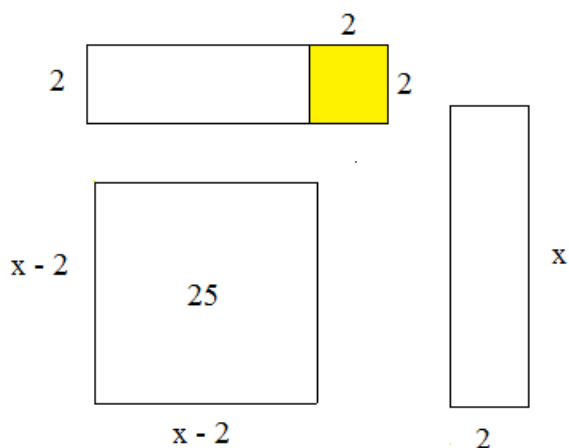
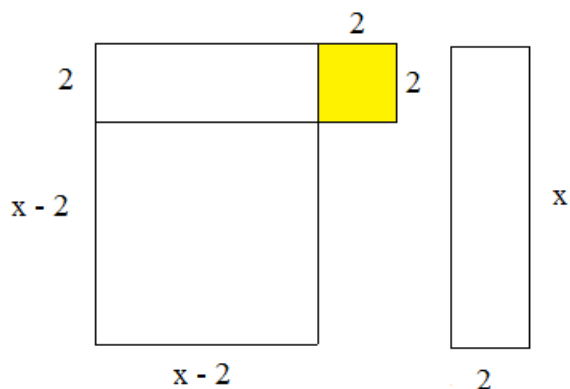
Your picture should show a square with sides  $x + 4$  and an area of 36. For that to work out  $x$  must be 2. The negative solution gives  $x + 4 = -6$  so  $x = -10$ .

Surprisingly, we can still use the ancient method of literally completing the square when a quadratic equation has a negative number in front of the  $x$  term. Let's tackle  $x^2 - 4x = 21$ . First, we'll draw a square with sides  $x$ . Then we will attempt to subtract  $4x$  by removing  $2x$  from both sides:



Now there is a problem, since there is not another piece with area  $2x$  to remove. Let's fix that by adding a little piece. Just remember to add it on both sides of the equation:

$$x^2 - 4x + 4 = 21 + 4.$$



Since  $x^2 - 4x + 4 = 25$ , the area that remains after we have added the little piece and taken away two strips of  $2x$  is 25. Each side of the remaining square is  $x - 2$  long, which has to equal 5. That means  $x = 7$ , and this solution fits the original equation  $x^2 - 4x = 21$ . The negative square root of 25 is -5, which provides the other solution:  $x - 2 = -5$ , so  $x = -3$ .

If you look at the pictures carefully, you will see that completing the square doesn't work so well when there is a number in front of the  $x^2$  part of the equation, like  $3x^2 + x - 18$ . Make sure to get rid of the 3 before completing your square. This is easy to do if you have an equation with a zero on one side, since you can divide both sides by 3:



$$3x^2 + x - 18 = 0$$

$$x^2 + \frac{1}{3}x - 6 = 0$$

Now you need to split  $\frac{1}{3}x$  in half. To quickly divide  $\frac{1}{3}$  by 2, just shove a 2 underneath the fraction line:  $\frac{1}{2 \cdot 3}$ . This makes the fraction 2 times smaller [dividing by 2 is the same as multiplying by  $\frac{1}{2}$ ]:  $\frac{1}{2 \cdot 3} = \frac{1}{6}$ . Add  $\left(\frac{1}{6}\right)^2$  to both sides of the equation.

$$x^2 + \frac{1}{3}x + \left(\frac{1}{6}\right)^2 - 6 = 0 + \left(\frac{1}{6}\right)^2$$

$$\left(x + \frac{1}{6}\right)^2 - 6 = \frac{1}{36}$$

$$\left(x + \frac{1}{6}\right)^2 = 6 - \frac{1}{36} = \frac{216}{36} - \frac{1}{36}$$

$$x + \frac{1}{6} = \pm \sqrt{\frac{215}{36}}$$

$$x = -\frac{1}{6} \pm \frac{\sqrt{215}}{6} = \frac{-1 \pm \sqrt{215}}{6}$$

If you are asked to complete the square when there is no equals sign, as in  $3x^2 + 4x - 18$ , you can put the 3 outside parentheses, and work inside those parentheses to complete the square:

$$3\left(x^2 + \frac{4}{3}x - 6\right) \quad \frac{4}{3} \text{ divided by 2 is } \frac{2}{3}:$$

$$3\left(x^2 + \frac{4}{3}x + \left(\frac{2}{3}\right)^2 - \left(\frac{2}{3}\right)^2 - 6\right) \quad \text{Notice the clever trick of both adding and subtracting } \left(\frac{2}{3}\right)^2.$$

$$3\left(\left(x + \frac{2}{3}\right)^2 - \left(\frac{2}{3}\right)^2 - 6\right)$$

$$3\left(\left(x + \frac{2}{3}\right)^2 - \frac{58}{9}\right)$$

$$3\left(x + \frac{2}{3}\right)^2 - \frac{58}{3}$$

**Be careful to multiply everything by 3, including the last term!**

Note that this was not an equation, so there is no actual solution (no values for x).

The method of completing the square has been generalized as the quadratic formula. The following explanation shows where the quadratic formula comes from. Instead of solving a specific quadratic equation like the ones above, we solve the general quadratic  $ax^2 + bx + c = 0$ . By doing this we end up with a way to solve all quadratic equations, without having to go through all those hard steps each time.

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

$a$ ,  $b$  and  $c$  are numbers in a quadratic equation

$$x^2 + \frac{b}{a}x + \frac{c}{a} = 0$$

Divide both sides by the number  $a$

$$x^2 + \frac{b}{a}x + \left(\frac{b}{2a}\right)^2 + \frac{c}{a} = \left(\frac{b}{2a}\right)^2$$

Complete the square

$$\left(x + \frac{b}{2a}\right)^2 + \frac{c}{a} = \left(\frac{b}{2a}\right)^2$$

$$\left(x + \frac{b}{2a}\right)^2 = \left(\frac{b}{2a}\right)^2 - \frac{c}{a}$$

Subtract  $c/a$  from both sides

$$\left(x + \frac{b}{2a}\right)^2 = \frac{b^2}{4a^2} - \frac{c}{a}$$

Square the first fraction

$$\left(x + \frac{b}{2a}\right)^2 = \frac{b^2}{4a^2} - \frac{4ac}{4a^2}$$

Multiply the top and bottom of the second fraction by  $4a$

$$\left(x + \frac{b}{2a}\right)^2 = \frac{b^2 - 4ac}{4a^2}$$

Combine the two fractions

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

Take the square root on both sides

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

Subtract  $\frac{b}{2a}$  from both sides

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

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

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

Take the square root of  $4a^2$

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

Combine the two fractions. Notice that the fraction line extends all the way underneath  $-b$ .

---

In Algebra 2, you will often encounter equations that need to have a common factor taken out. Always look for this first. An example would be  $4x^3 + 16x^2 - 8x = 0$ . Here there are only three terms, but it is not a quadratic equation. While it is possible to solve cubic equations in general, the method is complicated and tedious, and it is not commonly taught anymore. After all, we have graphing calculators and computers now. So, if you see a cubic equation with three terms chances are it has a common factor that can be taken out.  $4x^3 + 16x^2 - 8x = 0$  can be written as  $4x(x^2 + 4x - 2) = 0$ . Since  $4x$  could be 0, this shows that  $x = 0$  is one solution. The part in the parentheses can then be solved by using the quadratic formula. Your answers from the quadratic formula would be  $x = -2 \pm \sqrt{6}$ . As for all cubic equations, there are three answers altogether.

---

There may also be scary-looking 4<sup>th</sup> degree, or quartic, equations or expressions to factor. If there are no common factors, these can usually be handled by a sneaky substitution method.  $x^4 - 6x^2 + 8$  would be so easy to factor if it looked like  $x^2 - 6x + 8$ . We can make that happen by substituting something like  $u = x^2$ . Now we get  $u^2 - 6u + 8$ , which factors to  $(u - 4)(u - 2)$ . Since  $u = x^2$ , that gives us  $(x^2 - 4)(x^2 - 2)$ . Often we skip the  $u$  part and substitute directly, creating  $(x^2)^2 - 6(x^2) + 8$ , so we can factor straight to  $(x^2 - 4)(x^2 - 2)$ .

Many times at least one of the resulting factors is the difference of two squares, so it should be factored further.  $(x^2 - 4) = (x + 2)(x - 2)$ .

If you have an equation rather than an expression, as in  $x^4 - 6x^2 + 8 = 0$ , you can write  $(x^2 - 4)(x^2 - 2) = 0$ . Then you can say that  $(x^2 - 4) = 0$  so  $x^2 = 4$ , and  $(x^2 - 2) = 0$  so  $x^2 = 2$ . The solutions are  $x = \pm 2$  and  $x = \pm \sqrt{2}$ .

Any 4<sup>th</sup> degree expressions with only two terms are usually the difference of two squares:

$x^4 - 16$  can be written as  $(x^2)^2 - 4^2$ , which factors to  $(x^2 + 4)(x^2 - 4)$ . Note that one of these resulting terms is also the difference of two squares, so we get  $(x^2 + 4)(x + 2)(x - 2)$ . It is not uncommon for test questions to disguise the difference of two squares by multiplying both terms by some other factor. For example,  $3x^4 - 48y^4$  should be rewritten as  $3(x^4 - 16y^4)$ . Now you can see the difference of two squares:  $3((x^2)^2 - (4y^2)^2)$ .

## The Quadratic Formula and Factors

Factor  $21x^2 - 29x + 10$ .

To factor this you need two numbers that multiply to 210 and add to -29. It takes a fair bit of work to find these numbers, which are -14 and -15. If you have a calculator handy you may want to use the quadratic formula instead. All quadratic equations that have real number solutions can be solved this way. When you do that the answers you get are  $\frac{2}{3}$  and  $\frac{5}{7}$ , but now what? Remember that the quadratic formula gives you the values for  $x$  that make  $21x^2 - 29x + 10$  equal to zero. In this case, it tells you that  $x = \frac{2}{3}$  or  $x = \frac{5}{7}$ . Do not be tempted to write  $(x - \frac{2}{3})(x - \frac{5}{7}) = 0$ ! Multiplying this out would give you  $x^2 - \frac{29}{21}x + \frac{10}{21} = 0$ , which is a polynomial equation that has the same zeroes as  $21x^2 - 29x + 10$ . [You can see that if you divide both sides of  $21x^2 - 29x + 10 = 0$  by 21.]

Instead, take the equations for  $x$  and rearrange them to get the actual factors:

$$x = \frac{2}{3}, \text{ so } 3x = 2 \text{ and } 3x - 2 = 0.$$

$$x = \frac{5}{7}, \text{ so } 7x = 5 \text{ and } 7x - 5 = 0.$$

Now we can write  $(3x - 2)(7x - 5) = 0$ , and the factored form of the expression is  $(3x - 2)(7x - 5)$ .

Note that this is really also what you should do when you get whole number values for  $x$ , like for example  $x = 4$  and  $x = -5$ . Rearrange to get  $x - 4 = 0$  or  $x + 5 = 0$ , and write the factored form as  $(x - 4)(x + 5)$ .

For  $x^2 + 4x - 2 = 0$ , your answers from the quadratic formula would be  $x = -2 \pm \sqrt{6}$ . If you are looking for factors rather than answers, you could write:

$$x = -2 + \sqrt{6}, \text{ so } x + 2 - \sqrt{6} = 0$$

$$x = -2 - \sqrt{6}, \text{ so } x + 2 + \sqrt{6} = 0$$

That means that  $x^2 + 4x - 2$  could potentially be written like this:  $(x + 2 - \sqrt{6})(x + 2 + \sqrt{6})$ . You should multiply this out to see that it is correct. Fortunately teachers rarely ask for factors that contain radicals. Something like  $x^2 - 2$  could also be factored further as  $(x + \sqrt{2})(x - \sqrt{2})$ , but again that is usually not required.

If you find values for  $x$ , substitute them into the original equation to see if they are correct. Be careful, since you could be missing one of the solutions and not know it.

When you don't have an equation and you are just factoring an expression, multiply your factors back out to see if you get the original expression back.

If you have solved a quadratic equation by factoring, also use the quadratic formula to check your answers.

For any polynomial equation, use a graphing calculator or graphing software to check the zeros. The function crosses or touches the x-axis at these points. The “CALC” function on your graphing calculator will help you find the exact points. If you are using math GV you have to zoom in on the point and use the mouse marker values to estimate the point’s location. (If you zoom in a lot you can get accuracy up to two places after the decimal point.)

## The Sum or Difference of Two Cubes

Another change from Algebra 1 is that you may encounter the sum or difference of two cubes:

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

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

As an example, let’s find x if  $8x^3 + 27 = 0$ . To use the formula above, just rewrite this as the sum of two cubes:

$$(2x)^3 + 3^3 = (2x + 3)((2x)^2 - 6x + 3^2) = 0$$

$$(2x + 3)(4x^2 - 6x + 9) = 0$$

$$\text{Either } (2x + 3) = 0 \text{ or } 4x^2 - 6x + 9 = 0$$

That last equation can be solved using the quadratic formula:  $x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$

$$x = \frac{-(-6) \pm \sqrt{(-6)^2 - 4(4)(9)}}{2(4)} = \frac{6 \pm \sqrt{36 - 126}}{8} = \frac{6 \pm \sqrt{-90}}{8} = \frac{6 \pm \sqrt{-1 \cdot 9 \cdot 10}}{8} = \frac{6 \pm 3i\sqrt{10}}{8}$$

The difference of two cubes can be found by actually taking a cube with sides a and removing a little cube with sides b from it at one of the top corners. When you do that, the remaining volume can be found as follows:

On the bottom there is a volume of the height, (a – b), times the bottom,  $a^2$ .

The part above that has height  $b$ . Looking at this from the top, you can divide that into two sections: a smaller section with top area  $b$  times  $(a - b)$ , and a larger section with top area  $a$  times  $(a - b)$ .

The total volume is found by adding up these three sections:

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

Once you have found the difference of two cubes by simple geometry, you can derive the formula for the sum of two cubes by substituting  $-b$  for  $b$  in the equation  $a^3 - b^3 = (a - b)(a^2 + ab + b^2)$ :

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

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

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

If you have already learned about imaginary numbers, you can now find two complex cube roots for every number that has a nice cube root. For example, we know that  $\sqrt[3]{8} = 2$ , but let's set it equal to  $x$  just for fun:  $\sqrt[3]{8} = x$ . Then raise both sides to the third power:  $(8^{1/3})^3 = x^3$ , which means that  $8 - x^3 = 0$ . Hey, we can write that as the difference of two cubes:  $2^3 - x^3 = 0$ , and this factors as  $(2 - x)(x^2 + 2x + 2^2) = 0$ . One of the answers is  $x = 2$  as we would expect. The other two answers come from using the quadratic formula to solve  $(x^2 + 2x + 2^2) = 0$ . We get  $x = \frac{-2 \pm \sqrt{4 - 16}}{2}$ , which works out to  $\frac{-2 \pm \sqrt{-12}}{2} = \frac{-2 \pm \sqrt{-3 \cdot 4}}{2} = \frac{-2 \pm 2\sqrt{-3}}{2}$ , so the answers are  $x = -1 + i\sqrt{3}$  and  $-1 - i\sqrt{3}$ .

$$2^3 = 8, (-1 + i\sqrt{3})^3 = 8, \text{ and } (-1 - i\sqrt{3})^3 = 8.$$

The idea of the difference of two cubes may have played an important role in a historical mystery. The ancient Egyptians knew a lot about pyramids. They were apparently able to determine that the volume of a pyramid is  $1/3$  that of the rectangular box that would contain it. They did many calculations involving pyramids, and surprised today's mathematicians by finding a formula for the volume of a square pyramid that has its top cut off. No records survive to show how they accomplished this, but it seems amazing that they would come up with a formula like  $1/3 H(S^2 + Ss + s^2)$  when they didn't know algebra. In this formula,  $H$  is the height of the cut-off pyramid,  $S$  is the bottom side, and  $s$  is the top side. Today we can use algebra to find this formula.

The top part of a pyramid may be smaller than the whole, but it has the same proportions as the original pyramid. So, if we have a cut-off pyramid with a base of  $S^2$  and a top surface of  $s^2$ , the original pyramid would have had a height of  $kS$  (where  $k$  is some constant) and the small pyramid that was cut off the top would have a height of  $ks$ . To solve this in a general way we can start with blocks. Take a rectangular block with base  $S^2$  and height  $kS$ . Then remove a smaller block with base  $s^2$  and height  $ks$ . The result can be expressed like this:

$$kS \cdot S^2 - ks \cdot s^2$$

That is equal to  $k(S^3 - s^3)$ , which we know factors as  $k(S - s)(S^2 + Ss + s^2)$ . Change that to  $(kS - ks)(S^2 + Ss + s^2)$ , because  $kS - ks$  represents the height of the bottom part of the pyramid. Now take  $1/3$  of this volume to account for the fact that we are dealing with pyramids rather than blocks. The formula is  $1/3 H(S^2 + Ss + s^2)$ , where  $H$  is the height of the cut-off pyramid.

But how did the ancient Egyptians figure that out? They didn't know algebra! Were they contacted by aliens who supplied them with the correct formula? No one really knows what happened, but the only example that we found solves for the volume of the bottom half of a pyramid with base 4 and height 12. Here is a hypothetical story that might explain how the Egyptians found the solution.

*I know when I cut a square pyramid out of a block of clay, the volume of that pyramid is 1/3 of the block. More of that volume is in the bottom half. I wonder: what could be the volume of the bottom half of the pyramid? Let's start with a fresh block of clay. Hmm, I could cut out a pyramid and then cut it in half, but that's a lot of work and it would be hard to find the volume of the bottom piece when I'm done. Why don't I just leave the block and remove from it a block that would contain the top piece of the pyramid. How big would that smaller block have to be? Well, if I'm cutting the pyramid exactly in half, the top piece would still have the same proportions as the original. Let's say the original pyramid is 4 units by 4 units on the bottom, and 12 units high. Then the top half would be 2 units by 2 units on the bottom and 6 units high. I can just remove a 2 by 2 by 6 block that would contain the top part. Ugh, that would be hard to cut out of the middle of the block. I'll take it out of the top corner instead. There, I took a block that contained the original pyramid, and removed a block that contained the top half. The top half of the pyramid was 1/3 of the block I removed. 1/3 of the remaining shape is the bottom half of the pyramid. Now, what is the volume of the remaining clay? The bottom area is the square of the base, so 4 times 4. On the top surface we have 2 squared and 2 times 4. The height of each piece is 6, so the total is 6 times (16 + 4 + 8). Now take a third part of that, which is 2 times (16 + 4 + 8). That's 56. I can see that it is right, because the volume of the original pyramid was 1/3 times 12 times 4<sup>2</sup>, and I removed the top part with volume 1/3 times 6 times 2<sup>2</sup>. That is 64 - 8 = 56. The general formula must be 1/3 of H times (S<sup>2</sup> + s<sup>2</sup> + sS).*

## Factoring with Fractional Exponents

Fractional exponents are really no different than regular exponents, but because they look more complicated you may not find it as easy to factor them out.

$x^5 + x^2$  factors as  $x^2(x^3 + 1)$ . What you are doing here is dividing both terms by  $x^2$ . An expression like  $x^{3/2} + x^{1/2}$  factors in the same way: just divide both terms by  $x^{1/2}$ . When you divide you subtract the exponents, so  $x^{3/2} + x^{1/2}$  factors as  $x^{1/2}(x + 1)$ , which is the same as  $\sqrt{x}(x + 1)$ .

## Calculator Programs

Factoring can be a real pain. You often have to find two numbers that add up to something and multiply to something else, which is hard if those numbers are large. If you own a TI-84 or similar calculator you may want to program it to do this for you. The idea behind this is that we can use algebra to find the numbers we want.

Suppose that I am factoring  $6x^2 + 25x + 26$ . Now I need two numbers that add up to 25 and multiply to 6 times 26, or 156. I can call these numbers  $x$  and  $y$ , and write two equations:

$$x + y = 25$$

$$xy = 156$$

This system of equations is easy to solve by substitution. Let's eliminate  $y$ , by saying that  $y = 25 - x$ . Now substitute that into the second equation to get

$$x(25 - x) = 156$$

$$25x - x^2 = 156$$

$$-x^2 + 25x - 156 = 0$$

$$x^2 - 25x + 156 = 0$$

Hmm, that looks like an equation that can be solved by using the quadratic formula, which is easily programmed into a calculator (see later in this section). It can tell you that there are two possible values for  $x$  which are 12 and 13. Those are the numbers you want.

Now try solving this in a general way. Let  $x + y =$  the sum, and  $xy =$  the product:

$$x + y = S$$



$$xy = P$$

We can eliminate either x or y, with the same result. Let  $y = S - x$ , and substitute that into  $xy = P$ . You get

$$x(S - x) = P$$

$$Sx - x^2 = P$$

$$-x^2 + Sx - P = 0$$

$$x^2 - Sx + P = 0.$$

Remember that S and P are just numbers. Plug them into the quadratic equation:

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

Here is a sample program for the TI-84:

PRGM → NEW → ENTER

Use the ALPHA function of your calculator to enter a name for the program, up to eight characters long. This same function is used to enter text or variables in your program.

Commands are actions that the program must perform. They are entered by pressing the PRGRM key. Each separate line in your program starts with “:”, which the calculator puts in for you when you hit the enter key. When you tell the program to display a particular variable it will display the numerical value stored in that variable. This numerical value can be entered by the user through the Prompt command, or through calculations performed by the program.

```
: Prompt S,P           [The Prompt command is in PRGM → I/O which stand for Input/Output]
: S2 - 4P STO→A       [Use the STO key. This stores the value of S2- 4P in the variable A]
: If A<0               [The If command is in PRGM → CTL, which is already open as soon as you
                        hit PRGM. The Then and Else commands are found in the same place. The < symbol is in TEST,
                        which is located just above the MATH key]
: Then
: Disp “No Real Solutions” [Disp is in PRGM → I/O. Note that the text is in quotation marks.]
: Else
: (S + √(S2 - 4P))/2 STO→B
: (S - √(S2 - 4P))/2 STO→C
: Disp B,C
: End                  [In PRGM → CTL.]
```

Leave this screen by entering QUIT [2<sup>ND</sup> → MODE]

To run your program, press PRGM → ENTER, and ENTER again. The display should show S=? Enter the sum of the two numbers and press ENTER. The display will then show P=? Enter the product of the two numbers. Press ENTER to see the two numbers.

If your program doesn't work, you'll have to edit it. Go to PRGM → EDIT, and check carefully for any errors.

To use the TI-84 with the general quadratic formula, try the following program:

```
:Prompt A, B, C
: (B2 - 4AC) STO→D
:DISP "      SQRT"      [Leave some spaces by pressing ALPHA → 0 for the space symbol]
:DISP D
:If D<0
:Then
: Goto 10                [Goto is in PRGRM. Bypasses the next calculations if D is negative]
:Else
:-B - (√B2 - 4AC) STO→E
:2A STO→F
:-B + (√B2 - 4AC) STO→G
:DISP E/F ⌵Frac          [⌵Frac is located in MATH, and displays the answer as a fraction]
:DISP G/F ⌵Frac
:DISP E/F
:DISP G/F
:Lbl 10                  [Lbl is in PRGRM]
:END
```

First, the program will display the value under the square root sign, so you can check on your manual calculations. If this value is negative no answers are displayed. If the answers are integers, the program will display them twice. Otherwise, the answers will be shown as fractions first, and then as decimal numbers.

## Functions

You can replace  $f(x)$  with  $y$  to graph a function, or to make it easier to work with.

A function should not have more than one  $y$ -value for any  $x$ -value. A vertical line should not touch the graph in more than one spot.

A function may have two or more  $x$ -values that produce the same  $y$ -value.

The  $x$ -values are the domain (input), and the  $y$ -values are the range (output).

The  $x$ -intercept occurs where  $y = 0$ , and the  $y$ -intercept occurs where  $x = 0$ .

A one-to-one function has a unique  $y$  value for every  $x$ , so if you reverse it (switch the  $x$  and  $y$  values) you still get a function. Use the horizontal line test.

An onto function uses all of the possible  $y$ 's.

A function can be both one-to-one and onto, which makes it easy to reverse without having to worry about the domain of the new function.

The word “function” appears in this e-book nearly 500 times, which probably means that is a fairly important concept that you should be familiar with. The topic of functions is also included in college admission tests. What you need to know is that a function is a special relationship between two quantities. Each separate “input” has a unique “output”. A simple function is  $y = x^2$ . For every  $x$  (the input) you select, you get one, and only one, value of  $y$  (the output). If you pick  $x = 3$ ,  $y$  will be 9, which is the same value you get if you select  $-3$  for  $x$ . That last part isn't a problem, because the output value is always predictable. However, a relation like  $x^2 + y^2 = 100$  is not considered to be a function because almost every value of  $x$  results in two different outputs for  $y$ . There is not a single predictable output when you insert a value for  $x$ . For example, if  $x$  is 6, then  $6^2 + y^2 = 100$  and  $y^2 = 64$ . Now there are two values for  $y$ : 8 and  $-8$ . This relation graphs as a circle with a radius of 10. It can be rearranged as follows and split into two functions:

$$x^2 + y^2 = 100$$

$$y^2 = 100 - x^2$$

$$y = \sqrt{100 - x^2} \text{ or } y = -\sqrt{100 - x^2}$$

The input value for a function is usually called  $x$ , but you may see other letters. When time is one of the quantities, people usually use  $t$  as the input variable. The output value may be called  $y$ , especially when the function needs to be graphed. However, there are some disadvantages to that. Suppose that you have two different functions, such as  $y = x^2 - 10$  and  $y = 2x + 1$ . If you are working with both functions at the same time, you might call the output of one function  $y_1$ , and the other one  $y_2$ . Subscripts like that are a bit of a pain to type, which makes them less useful for computer programmers who need functions in their code. There is also another problem. Let's pick a value of 5 for  $x$  in the function  $y_1 = x^2 - 10$ . When you are done calculating, your output says  $y_1 = 15$ , with no indication of what value of  $x$  you put in. If a function is really complicated it would be hard to guess the input just by looking at the output. You might want to add something, like:  $y_1$ , when  $x$  is 5, = 15. That works, but it looks awkward.

Function notation solves the problems associated with the  $x$ - $y$  method of expressing functions. Instead of  $y$  we use the letter  $f$ , for function. When we have a second function, we can use  $g$ , and even  $h$  for the third. Just change  $y_1 = x^2 - 10$  and  $y_2 = 2x + 1$  to  $f = x^2 - 10$  and  $g = 2x + 1$ . There, that takes care of those subscripts. Now, instead of writing:  $f$ , when  $x$  is 5, = 15, we're just going to use some parentheses:  $f(5) = 15$ . This still means the same thing: when  $x$  is 5, the output of the function is 15. To shorten what we are saying when we read that out, we say that "f of 5 equals 15", but you're welcome to read it the long way as "f, when  $x$  is 5, equals 15". Notice that  $f(6)$  equals 26, and  $f(x) = x^2 - 10$ . Students sometimes find it confusing that there are multiple  $x$ 's in this last expression, but just remember that once you pick a value for  $x$ , it is the same everywhere. So, if  $x = 3$ , then put that into the function expression everywhere for  $x$ :  $f(3) = 3^2 - 10$ .

Whenever you feel confused by scary-looking function notation, or if you need to graph a function, just change it back to regular  $x$ - $y$  notation:

$$f(x) = 15x^3 - 2x + 3 \quad \text{is the same as} \quad y = 15x^3 - 2x + 3$$

## Domain, Range and Intercepts

You also need to know that functions have a domain and a range. The word domain means an area that is controlled by someone or something. The **domain** of a function consists of all of

the input values that the function is able to use and produce an output for. The domain is all of the x-values. The **range** of a function indicates all of the possible output values. The range is all of the y-values. **The x-intercept** is the point where the function crosses the x-axis. That always happens when y is zero, so the coordinate of the x-intercept looks like (x, 0). The **y-intercept** is the point where the function crosses the y-axis, so x is zero. The coordinate of the y-intercept looks like (0, y).

### Example

Find the domain, range, and the intercepts for  $y = \sqrt{x + 4}$ .

Because the square root of a negative number is an imaginary (i) number, we don't want that in our functions. So, we need to make sure that there will be no negative number under the square root sign. Zero is fine, because the square root of zero is just zero:  $0 \cdot 0 = 0$ , so  $\sqrt{0} = 0$ . To get a positive number or zero, x needs to be -4 or greater. The domain is all numbers greater than or equal to -4. Depending on what your teacher prefers, you can write this as  $x \geq -4$ , or as the interval -4 to infinity:  $[-4, \infty)$ . Here the square bracket, [ , indicates that -4 is included in the interval. Infinity is not actually a number, so it is never included. This is indicated by an open bracket, ). The y-intercept occurs when x is 0:  $y = \sqrt{0 + 4}$ . They-intercept point is (0, 2). The x-intercept occurs when y is zero:  $0 = \sqrt{x + 4}$ , so x should be -4. The x-intercept point is (-4, 0).

## Finding the Domain and Range

In most cases it is not very difficult to determine the domain and then find the range of possible output values of a function. Some functions have a restriction on their domain, for example  $f(x) = \sqrt{x}$ . Because the functions we deal with involve only real numbers, no negative value may appear under the square root sign. The domain is the interval  $[0, \infty)$ , and as output it will return values that are anywhere from zero to positive infinity for all possible values of x in the domain. If you are lucky your course may only ask you to determine the domain and range of fairly simple functions. Examples of increasing complexity are provided here for you to use as needed.

### Example

Find the domain and range of the function  $f(x) = \sqrt{x^2 - 9}$ .

Because of the square root in this function, there is a restriction on the domain.  $x^2 - 9$  must not be negative. Start by saying that  $x^2 - 9 \geq 0$ . Therefore,  $x^2 \geq 9$ . Be careful when you take the square root on both sides because the inequality sign complicates things a bit. If  $x^2 = 9$ , we just consider that there is both a positive and a negative square root of 9, and write that  $x = 3$  or  $-3$ . The negative possibility will flip the inequality around, so that  $x \geq 3$  or  $x \leq -3$ . Now that you know how to actually take the square root of  $x^2$ , you can write that like this:

$$\sqrt{x^2} \geq \sqrt{9}$$

$$|x| \geq 3$$

$$x \geq 3 \text{ or } -x \geq 3$$

$$x \geq 3 \text{ or } x \leq -3$$

The domain is  $(-\infty, -3] \cup [3, \infty)$ . The two intervals are joined by a “union” sign.

Since the smallest value for  $x^2 - 9$  is 0, the range is the interval  $[0, \infty)$ . In your own work use your teacher’s preferred notation to indicate the domain and range.

### Example

Find the domain and range of the function  $f(x) = \sqrt{x^2 + x}$ .

To make sure that the part under the square root sign is not less than zero, we have to find where  $x^2 + x$  would be less than zero.

$$x^2 + x < 0$$

$$x^2 < -x$$

**Before you divide by an unknown like  $x$ , think carefully!** First,  $x$  must not be zero when you divide by it. Here we can rule out a zero value for  $x$  because  $0^2$  is not less than  $-0$ . Now, if  $x$  is positive we can divide by  $x$  without changing the  $>$  sign:

For  $x > 0$ , a division by  $x$  gives the result  $x < -1$ . Notice there are no solutions to this, since no positive number is less than  $-1$ .

If  $x$  is negative, the  $>$  sign changes to  $<$  when we divide by  $x$ :

For  $x < 0$ ,  $x > -1$ . The solutions for this are  $-1 < x < 0$ . These are the values that have to be excluded from the domain.

The domain is  $(-\infty, -1) \cup [0, \infty)$ . The range can't be negative because this is a square root function. The smallest possible output is 0, when  $x^2 + x = 0$ . There is no limit on the largest possible output. The range is  $[0, \infty)$ .

### Example

Find the domain and range of  $f(x) = -\frac{3}{4} \sqrt[3]{5x^2 - 1} + 2$ .

It is more difficult to imagine the range of a complex function, so we would like to determine it in a systematic way. The nice thing about this function is that there are no restrictions on the domain (the possible input values), because there is a cube root which is fine to use with negative values. The domain is the interval  $(-\infty, \infty)$ . We can just imagine putting large negative numbers, 0, or large positive numbers into the function. It is easiest to decompose this function into several parts. Isolating the cube root first is not so helpful, because its output can be anywhere from  $-\infty$  to  $+\infty$ . We have to start by considering the range of what we are taking the cube root of. First let's look at the range of  $5x^2$ , which is of course 0 to  $+\infty$ . Next look at the range of  $5x^2 - 1$ , which is  $-1$  to  $+\infty$  (subtracting 1 from infinity doesn't really make a dent in it).

Next we can take the cube root of all of these numbers. The cube root of  $-1$  is  $-1$ , which is the smallest possible output here. Taking the cube root of  $+\infty$  (so to speak) still leaves you with  $+\infty$ . The range is still  $-1$  to  $+\infty$ .

After this, we have to multiply by  $-\frac{3}{4}$ . The result of this operation ranges from  $\frac{3}{4}$  to  $-\infty$ , which is more properly ordered as  $-\infty$  to  $\frac{3}{4}$ . The last thing that needs to be done is to add 2, which brings the range to:  $-\infty$  to (and including)  $2\frac{3}{4}$ , or  $(-\infty, 2\frac{3}{4}]$  which is our final answer.

### Example

Find the domain and range of  $f(x) = -\frac{1}{2} \sqrt{3x^2 - 3} + 4$ .

This function has a restriction on its domain due to the square root. No negative number may appear under the square root sign, so  $3x^2 - 3$  must be positive or zero:

$$3x^2 - 3 \geq 0$$

$$3x^2 \geq 3$$

$$x^2 \geq 1$$

$$\sqrt{x^2} \geq \sqrt{1}$$

$$|x| \geq 1$$

$$x \geq 1 \text{ or } -x \geq 1$$

$x \geq 1$  or  $x \leq -1$ . The domain therefore excludes the interval  $(-1,1)$ , but that doesn't keep the range from going all the way from 0 to + infinity for the part with the square root. The restriction on the domain doesn't affect the range. Multiply this range by  $-\frac{1}{2}$  to get 0 to - infinity. (Cutting infinity in half doesn't make a dent in it either!) Turn that around to get  $-\infty$  to zero, and add 4 which leaves you with  $-\infty$  to 4, or  $(-\infty, 4]$ .

For additional information, see "Range of a Rational Function".

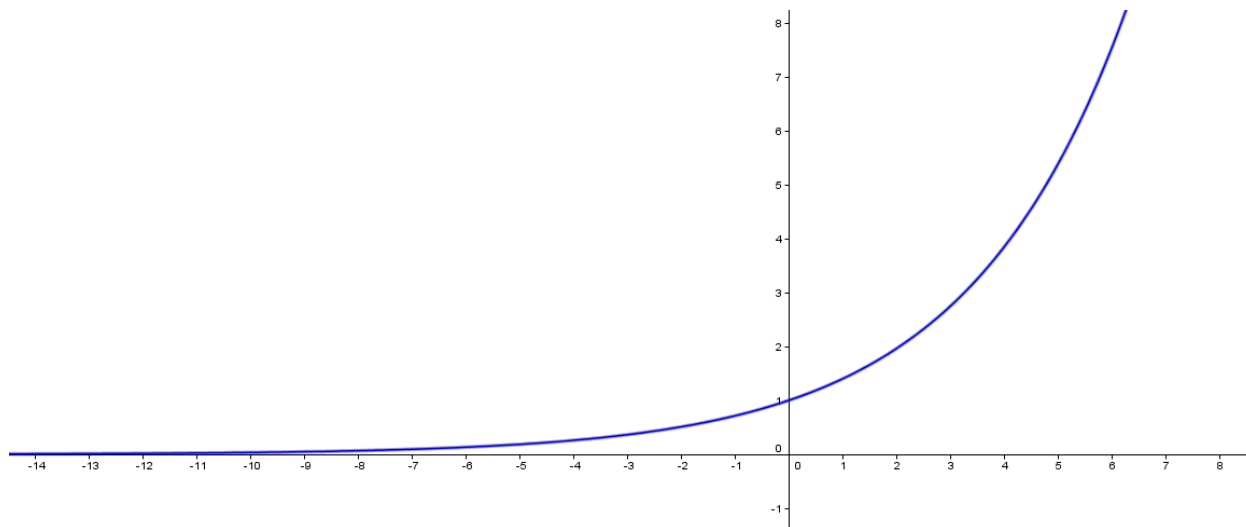
## One-to-One Functions

A relation can't be a function if a given  $x$  value can produce more than one  $y$ -value, but the other way around isn't a problem. For  $y = x^2$ , two different  $x$  values can have the same output.  $f(3) = 9$  and  $f(-3) = 9$ . That is fine so long as you don't have to turn the function around. Later on we will look at inverse functions which flip the original around. In the case of  $y = x^2$ , the inverse would be  $x = y^2$ , which can also be written as  $y = \pm \sqrt{x}$ . This is not a function, as we could have predicted from the fact that the original function has the same output for different  $x$ -values. If you want a function to be reversible then it has to be **one-to-one**, which means that every  $x$  has a unique  $y$  as its output. For example,  $y = x + 5$  produces a unique  $y$  for every  $x$ . We can flip that around to get  $x = y + 5$ , which rearranges to give  $y = x - 5$ . Notice that this inverse function does the opposite of what the original function does. Adding 5 turns into subtracting 5. The inverse relation  $y = x - 5$  is a function because the original function was one-to-one. To see if a function is one-to-one, use the horizontal line test. A horizontal line should only touch the graph in one spot if there are no two  $x$ 's that have the same  $y$ -value.

Some one-to-one functions can create a bit of a surprise when they are reversed. The graph in the picture below is a function because there is only one  $y$  for every  $x$ . It is also one-to-one because each  $x$  has a unique  $y$ . No two  $x$ 's share the same  $y$  value. It should be reversible. However, if you look at the left side of the graph, you can see that this function appears to have no intention of crossing the  $x$ -axis. In fact, the  $y$  values keep getting smaller and smaller but



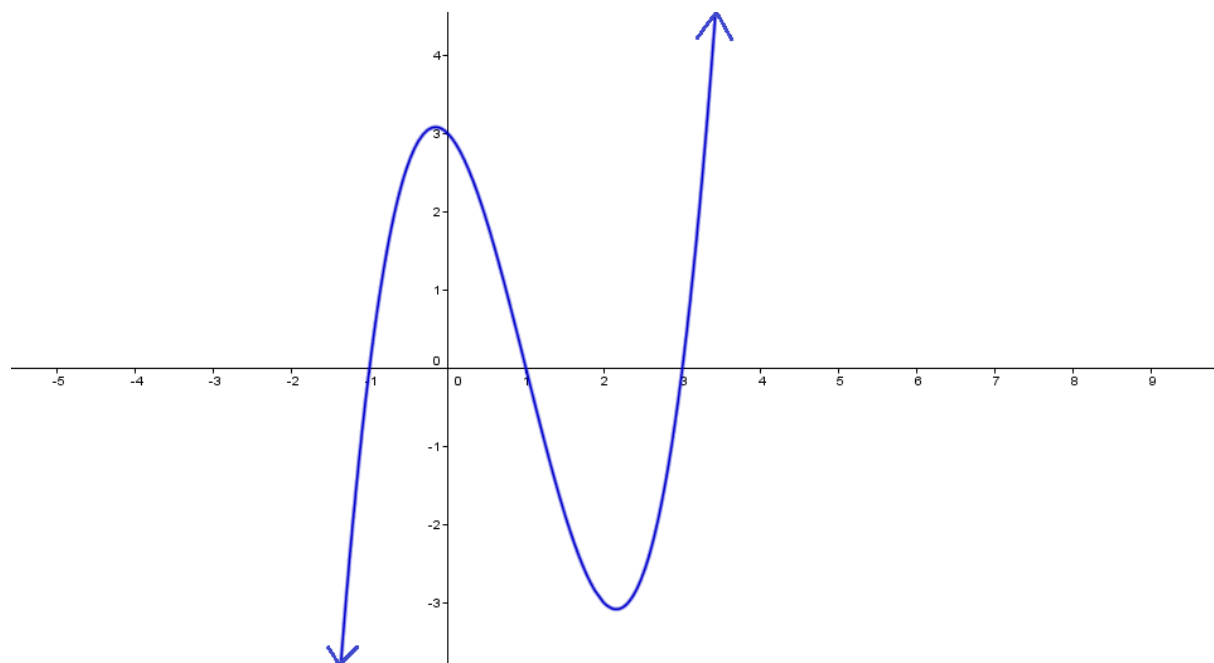
never reach zero. The function actually never produces a negative output. That means that its inverse function will not be able to take negative inputs. Obviously a function like this is lacking some property that was present in the previous function,  $y = x + 5$ .



## Onto Functions

What is missing in the function shown above is that it doesn't produce some of the  $y$ -values that other functions do. It takes both positive and negative  $x$ -values and maps them to only the positive  $y$ 's. A function that *does* use all the  $y$ 's is called an **onto function**. It maps a set of  $x$ 's onto a set of  $y$ 's so that all of the possible  $y$ 's are used. A rather lame trick for turning the above function into an onto function is to specify that the set of possible  $y$ 's should only include the positive real numbers. That isn't really necessary because there are plenty of functions that are clearly onto functions without using special definitions. Any simple linear function like  $y = x + 1$  will cover all possible  $y$ -values. So will functions that have an odd power of  $x$ , like  $y = x^3$  or  $y = x^5 + 7$ . Many functions that you might initially think of when you are looking for an onto function are also one-to-one functions. They have both a unique  $y$  for every  $x$ , and every possible  $y$  has a unique  $x$ . That is a good thing because these functions are easily reversible to create a good inverse function without having to worry about the domain.

Some onto functions are not one-to-one. Even if a function covers the entire range from negative infinity to positive infinity, it may have one or more bumps along the way that cause some of the  $y$ -values to be duplicated. The graph below shows a function that is onto, but not one-to-one. This particular function is  $y = x^3 - 3x^2 - x + 3$ :



## More About Functions

An even function is symmetrical around the y-axis.  $f(x) = f(-x)$ .

An odd function has a graph that is the same if it is rotated  $180^\circ$  around the origin.  $f(x) = -f(-x)$ .

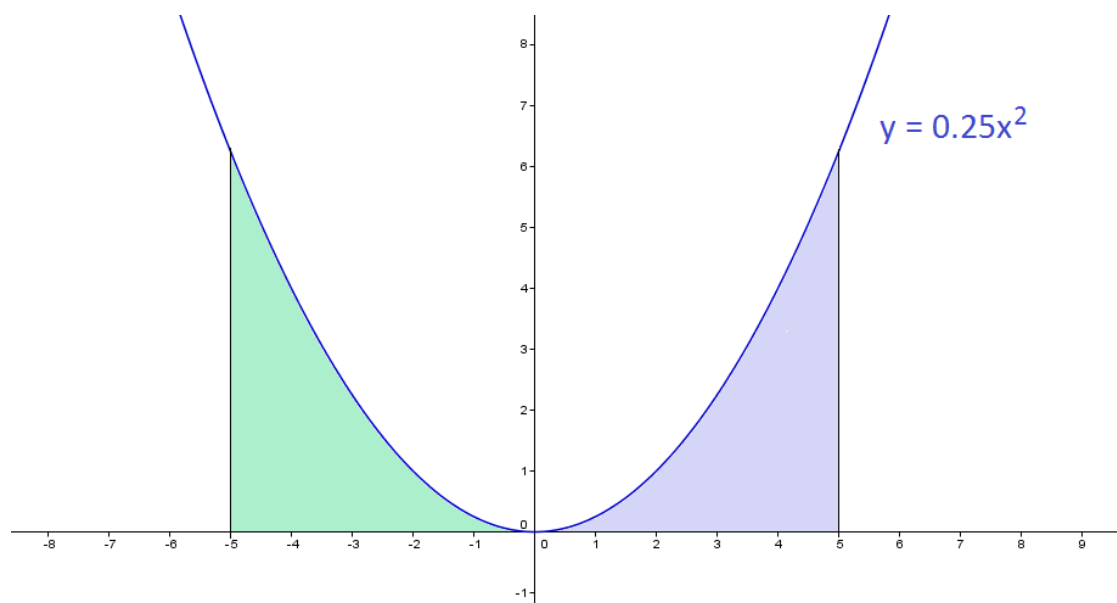
To find an inverse function, simply switch  $x$  and  $y$ , and then solve for  $y$ .

The range of the original function becomes the domain of the inverse function.

$f \circ g$  means “ $f$  composed with  $g$ , which is  $f(g(x))$ . Use the output of  $g$  as the input for  $f$ . Be careful about the domain of the new function.

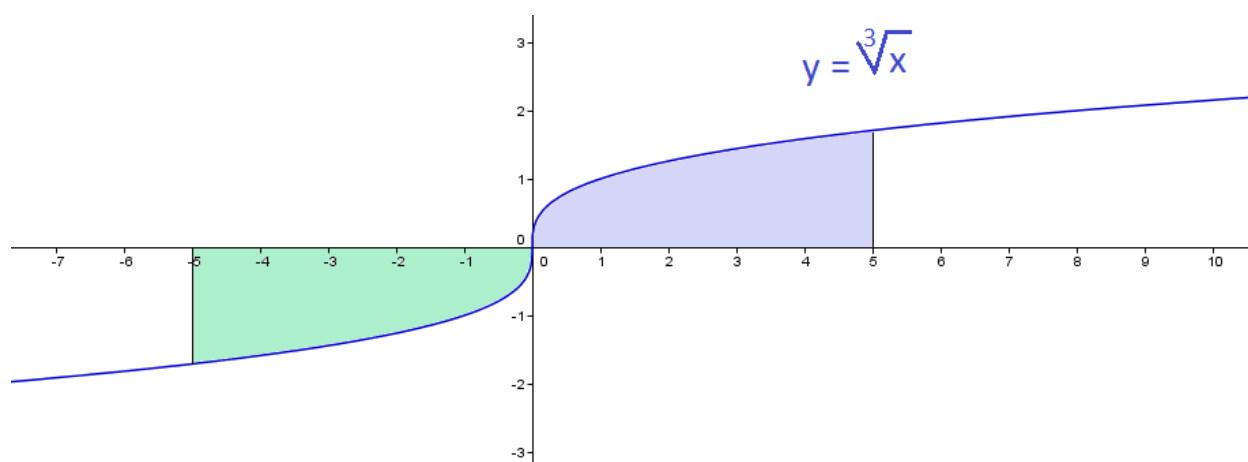
## Even and Odd Functions

When you get to calculus, you will be determining the area between the curve of a function and the  $x$ -axis. For the function  $f(x) = 0.25x^2$ , the area under the curve between  $-5$  and  $0$  is exactly the same as the area between  $0$  and  $5$ . When you look at the graph you can see why:  $y = x^2$  is symmetrical around the  $y$ -axis. Many functions are symmetrical like this so it is often faster to calculate half the area (starting at  $0$ ) and then double it. The area marked in green is equal to the area marked in blue:



You can recognize **even functions** fairly easily, because they all have the property that  $f(x) = f(-x)$ . Just put  $-x$  into the function in place of  $x$  and see if you get the same result. For example,  $f(x) = x^4 + 2x^2 + 11$  gives the same output for  $x = 2$  as for  $x = -2$ , due to the even powers on  $x$ .

Functions that have equal areas above and below the  $x$ -axis are called **odd functions**. If you rotate the graph around the origin 180 degrees, it looks exactly the same. Odd functions are also easy to spot because  $f(-x) = -f(x)$ . The output for  $x = 2$  is the opposite of the output for  $x = -2$ . A function with only odd powers of  $x$  is odd, like  $y = x^5 - x^3$ . For  $f(x) = x^3$ ,  $f(2)$  is 8 and  $f(-2)$  is -8. The graph below shows another odd function,  $y = \sqrt[3]{x}$ . The cube root of 8 is 2, because  $2 \cdot 2 \cdot 2 = 8$ . The cube root of -8 is -2 because  $-2 \cdot -2 \cdot -2$  is -8.



## Inverse Functions

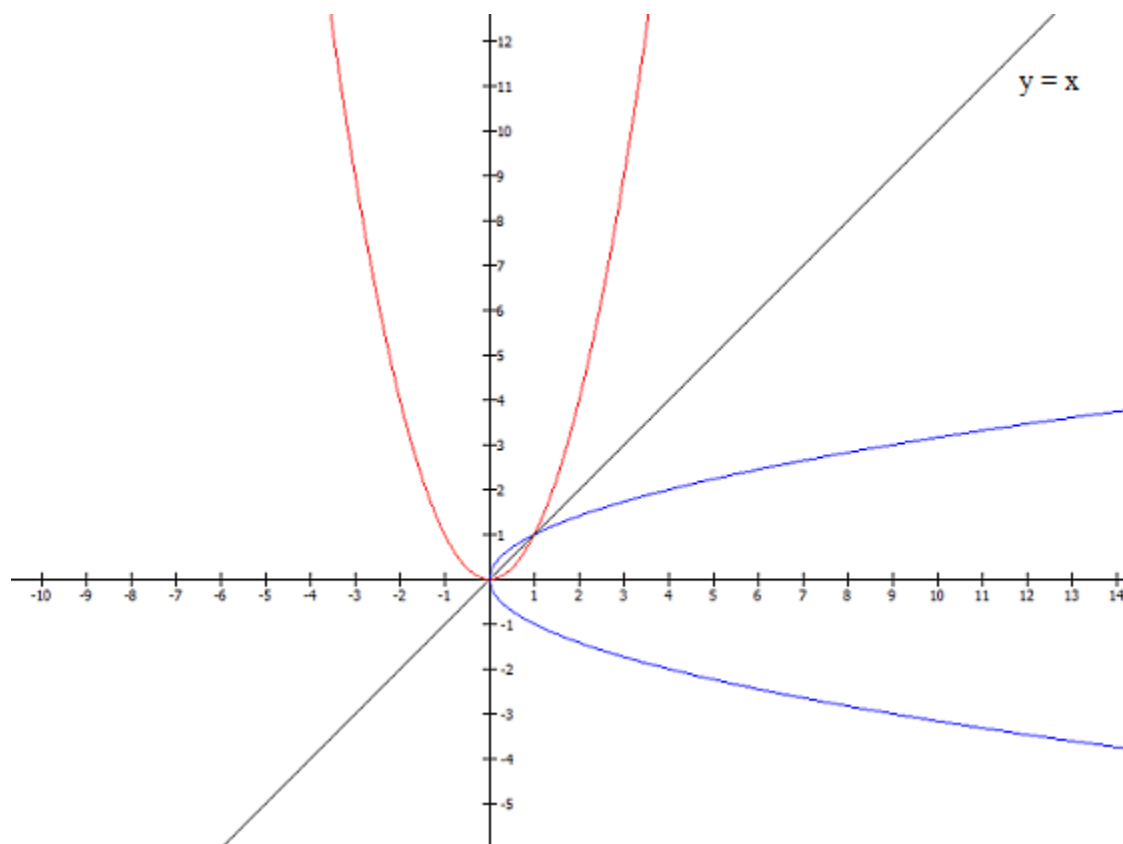
The inverse of a function is a process that “undoes” the operation of the function. If function  $f$  adds 2 to  $x$ , the inverse of  $f$  subtracts 2 from  $x$ . So, if  $f(x) = x + 2$ , and  $g(x) = x - 2$ , then these functions are inverses of each other. You can check this by picking a random value of  $x$ , say 5, and inserting it into the function  $f$ . The result is  $x + 2$ , or 7. If you then insert this result as the  $x$ -value for function  $g$ , the output is  $7 - 2 = 5$ , the value you started with. The inverse of a function  $f$  is often written as  $f^{-1}$ . Note that here -1 does NOT indicate an exponent.  $f^{-1}(x)$  is not equal to  $\frac{1}{f(x)}$  !

Finding inverse functions is not particularly difficult. First write the function in  $x,y$  notation, so write  $f(x) = x + 2$  as  $y = x + 2$ . Then we just switch  $x$  and  $y$  around:  $y = x + 2$  changes into  $x = y + 2$ . In this way the input becomes the output. After you do that however you should

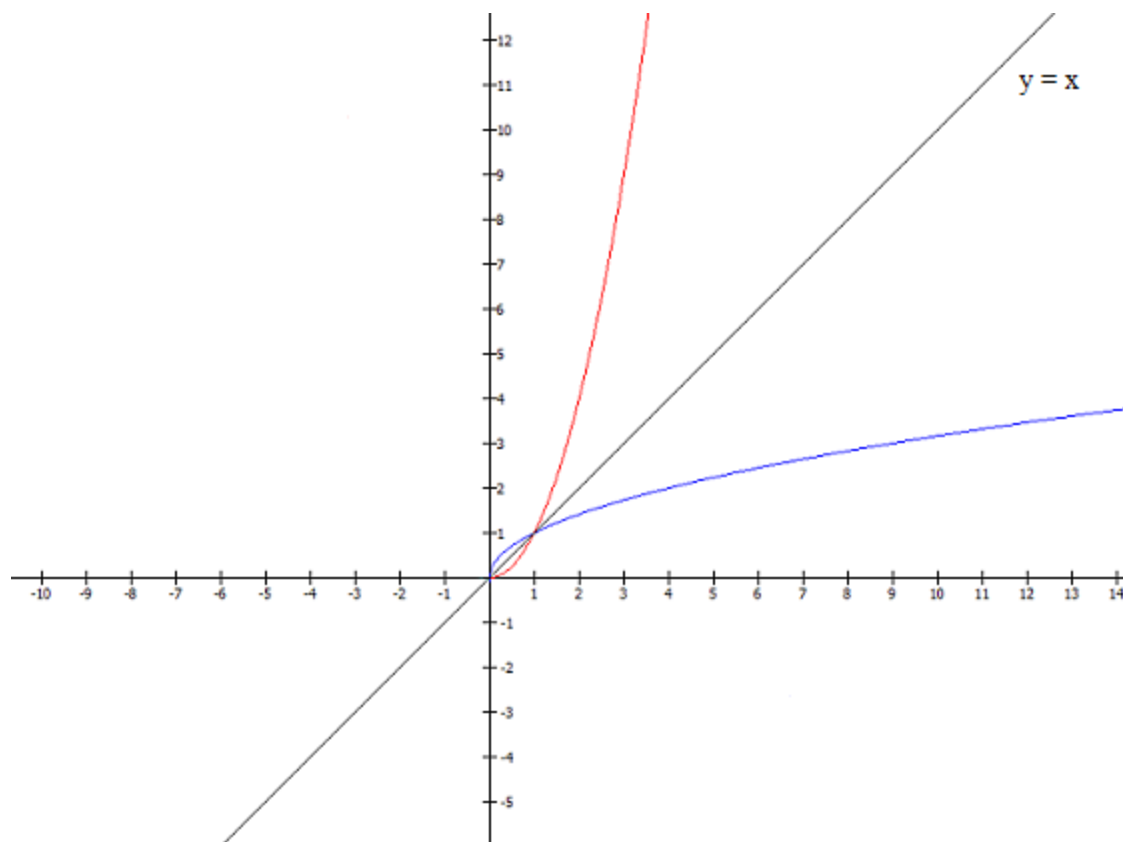
rearrange the function so that it again says  $y = \dots$ ; in this case  $y = x - 2$ . The only potentially difficult part here is solving for  $y$  which takes good algebra skills.

When you take a function and switch  $x$  and  $y$  around, you end up with a new relation between  $x$  and  $y$ , but this relation is not necessarily a function. The original function may have two different  $x$  values that result in the same output value. For example, look at  $f(x) = x^2$ . We can write this as  $y = x^2$ , so the inverse is  $x = y^2$ . Taking the square root on both sides we get  $\pm\sqrt{x} = y$ , or  $y = \pm\sqrt{x}$  (there is always both a positive and a negative square root). Here negative values for  $x$  are not permitted, but that is not a problem since the original  $y$  values were all positive to start with. **The range, or collection of output values for  $y$ , becomes the domain, or collection of input values for  $x$ , of the inverse function.**

The image below shows the original function in red, and the inverse in blue. Notice that the inverse is the original graph mirrored in the line  $y = x$ . The inverse graph is not a function because it fails the vertical line test. A vertical line drawn on the graph intersects the function more than once. Every valid input for  $x$  except 0 results in two different values for  $y$ .



Mathematicians are a little obsessive compulsive about that sort of thing, so they like to have that fixed. The domain of a function is the collection of valid input values for the function, and we can simply change the domain. In this case we will restrict the domain of  $f(x) = x^2$  to values greater than or equal to zero.



There, that takes care of things. The domain of the original function consists of the input values for  $x$ . The range of the original function consists of the output values for  $y$ . When we switch  $x$  and  $y$  around, the domain becomes the range and vice versa.

### Example

Find the inverse of the function  $y = \frac{1}{(x+1)^2}$ . Determine the domain and range of both the function and the inverse.

We find the inverse by simply switching  $x$  and  $y$  around:  $x = \frac{1}{(y+1)^2}$ . Take the square root on both sides, remembering that there is always a negative square root:  $\pm\sqrt{x} = \frac{1}{y+1}$ . Here we can use a shortcut: If  $2 = \frac{6}{3}$  then  $3 = \frac{6}{2}$ , so therefore  $y + 1 = \frac{1}{\pm\sqrt{x}}$ . Now we see that  $y = \pm\frac{1}{\sqrt{x}} - 1$ .

The domain of the original function is all real numbers except for  $x = -1$ . The range consists of positive numbers due to the square on the bottom. When  $x$  is very close to  $-1$  the number on the bottom is very small. That results in a really large output. When  $x$  is a large positive or negative number, the output is nearly 0, but it never actually gets there. The range is the interval  $(0, \infty)$ .

The inverse is not a function. The input  $x$  must be greater than 0 to avoid having a negative number under the square root sign, or a division by 0. The domain is the interval  $(0, \infty)$ . The range is all numbers except for  $-1$ . This is caused by the fact that  $\frac{1}{\sqrt{x}}$  can never quite reach 0, no matter how large a value for  $x$  we select. Notice that the domain of the inverse relation is the range of the original function, and vice versa.

## Composing Functions

Thankfully, composing functions is much easier than composing music. For most functions, we insert a value of  $x$ , and out comes a value for  $f(x)$ . The idea of composition is that we can insert the output of one function as the input of another. This usually involves a function  $f(x)$ , and a second function  $g(x)$ .

$f \circ g$  means “ $f$  composed with  $g$ ”. The function  $f$  will now take the output of the function  $g$ . The notation  $f(g(x))$  shows what is happening. For example, if  $f(x) = \sqrt{x}$ , and  $g(x) = x + 2$ , then  $f \circ g$  would be  $f(x+2)$ , which is  $\sqrt{x+2}$ . The new function that results from the composition is usually called  $h(x)$ , so we can say that  $h(x) = \sqrt{x+2}$  in this case.

To show this in a more practical way, let's look at a puddle. This particular puddle is created by a leaking pipe, and it is perfectly circular. The size of the puddle is increasing, and the length of the radius at time  $t$  is given by the function  $r(t) = 5t$ , where  $r$  is in millimeters, and  $t$  is in minutes. From this we can create a function that shows how the area of the puddle is related to the time  $t$ .  $A = \pi r^2$ , which can be written as a function:  $A(r) = \pi r^2$ . We can compose this area function with the radius function, by using the output of  $r$  as the input for  $A$ :  $A(5t) = \pi(5t)^2$ .

That simplifies to  $A(5t) = 25\pi t^2$ . This gives us a new function, which we will call  $P(t)$ , because it tells us what the area of the puddle is at time  $t$ :  $P(t) = 25\pi t^2$ .

If two functions  $f$  and  $g$  are inverses of each other, then  $f(g(x)) = x$  and  $g(f(x))$  is also  $x$ . That is what inverse functions do. One function changes  $x$  into something, and then the inverse function reverses that change.

Normally we have to be careful about the domain of a function, so that we do not put something into the function that wouldn't make sense. For  $f(x) = \frac{1}{x}$  we know not to put 0 for  $x$ . However, if we are creating a new function  $h(x)$  that is  $f(x)$  composed with  $g(x) = x - 2$ , we must make sure that  $g(x)$  is not zero, so that 0 will not be the input for  $f(x)$ . In this case we will exclude  $x = 2$  from the domain, so that we can use the output of  $g(x)$  in  $f(x)$  without creating a problem.

Sometimes both functions have restrictions. For example,  $f(x) = \sqrt{4 - x}$ , and  $g(x) = \sqrt{x}$ . We will be doing  $f \circ g$ , which gives us the new function  $h(x) = \sqrt{4 - \sqrt{x}}$ . We start by looking at  $g(x)$ . The restriction on  $x$  here is that  $x \geq 0$ , so that there will not be a negative number under the square root sign.  $f(x)$  on the other hand can take any number that is less than or equal to 4. A negative number would be fine, but 5 would not be a suitable input. Therefore we have to put additional restrictions on  $x$ . If we want  $\sqrt{x}$  to be less than or equal to 4,  $x$  must be less than or equal to 16. In addition we still have the original restriction that  $x$  cannot be negative. The domain of the new function will be the interval  $[0, 16]$ .

When you are dealing with a function that is the result of a composition, it is sometimes tempting to just look at the function itself to determine the domain. This is not suitable. For example, take the function  $h(x) = x - 1$ , which was created by composing  $f(x) = x^2$  with  $g(x) = \sqrt{x - 1}$ . We might think that there would be no restrictions on the domain of  $h(x)$  by just looking at it. However, no values of  $x$  smaller than 1 can be put into the function  $g(x)$ . This also restricts the domain of  $h(x)$  to values larger than or equal to 1. To see this more clearly, look at  $h(x)$  before it was simplified:  $h(x) = (\sqrt{x - 1})^2$ .

Just as we can compose functions, we can also decompose them. If we have a function  $h(x) = (x + 2)^2$ , we can look at it as the composition of two separate functions:  $f(x) = x^2$  and  $g(x) = x + 2$ .  $h(x)$  is  $f \circ g$ , or  $f(g(x))$ . The ability to view a single function as consisting of a composition of two (or more) separate functions can come in handy at times, as we will see when we try to determine the range of a fairly complex function.



## Absolute Value Equations

$|x| = 5$  means  $x = 5$  or  $-x = 5$

Watch out for extraneous solutions!

Use  $y = \text{abs}(x)$  to graph absolute value functions.

Modifying the graph of  $y = |x|$ :

$y = |cx|$  changes the slope of the lines

$y = -|x|$  flips graph upside down

$y = |x| + b$  moves graph up or down by  $b$  units

$y = |x + a|$  moves graph left by  $a$  units if  $a$  is positive, and right by  $a$  units if  $a$  is negative.

Absolute value functions can be written as piecewise functions. Use the two different possibilities created by the absolute value sign as the conditions that specify which function rule to use.

First, you should make sure that you feel comfortable solving basic equations that don't involve absolute value. If you need some review, see "Balancing Equations" in the [Pre-Algebra / Algebra 1](#) e-book. Sample questions of increasing difficulty are provided here so you can test your skills. Solutions are located at the bottom of this page.<sup>1</sup>

1.  $x + b - 20 = 13 + b$ , solve for  $x$

2.  $4 = \frac{3}{5}x - 2$

3.  $1.2x = 3x - 6$

4. Solve by cross-multiplying:  $\frac{4}{x} = \frac{7}{12}$

5.  $8 + bx = 12 + 2x$ , solve for  $x$

---

<sup>1</sup> Solutions: 1.  $x = 33$  2.  $x = 10$  3.  $x = \frac{30}{9}$  4.  $x = \frac{48}{7}$  5.  $x = \frac{4}{b-2}$  6.  $x = 0$  or  $x = \frac{1}{4}$

6.  $x = 4x^2$ , find both values of  $x$

---

The absolute value sign changes a negative number into a positive quantity. If the number inside the sign is positive or zero, it remains unchanged.

The absolute value sign is sometimes found in equations, where it creates two distinct possibilities.

### Example

First, let's solve a really simple equation:  $|x| = 5$ .

The absolute value sign only actually does something when  $x$  is negative. If  $x$  is positive or zero we can just write the equation as  $x = 5$ . On the other hand, if  $x$  is a negative number it is changed into a positive number. That is,  $x$  becomes the opposite, or  $-x$ . In this case the equation reads:  $-x = 5$ . Now we multiply both sides by  $-1$  to get  $x = -5$ . The two possibilities

are:  $x = 5$  or  $x = -5$ . That's easy since we could have guessed it in the first place. This is also the basis for a commonly used shortcut. People often solve simple equations of the type  $|\text{.....}| = 5$  by saying that  $\text{.....} = 5$  or  $\text{.....} = -5$ . For example,  $|3x + 2| = 8$  could be solved by saying:  $3x + 2 = 8$  or  $3x + 2 = -8$  which solves to  $x = 2$  or  $x = -3\frac{1}{3}$ .

**Caution: an equation like  $|4x| = -20$  will have no solution. The part with the absolute value sign can never be negative!**

If the equation is more complicated, the usual strategy is to isolate the absolute value expression and deal with it last, as shown in the next example.

### Example

Solve:  $4|6x - 7| - 8 = 12$

The absolute value sign is not mentioned in the order of operations phrase Please Excuse My Dear Aunt Sally, but that doesn't mean that you can just reach inside of it. Sometimes students will try to solve this kind of problem by combining like terms, like adding the  $-7$  and  $-8$  together first, or by multiplying out  $4(6x - 7)$ . Don't do that! Things inside an absolute value sign require special handling. Isolate the part with the absolute value sign:

$$4|6x - 7| = 20$$

$$|6x - 7| = 5$$

$$6x - 7 = 5 \quad \text{or} \quad 6x - 7 = -5$$

$$6x = 12 \qquad 6x = 2$$

$$x = 2 \qquad x = \frac{2}{6} = \frac{1}{3}$$

Check your answers by substituting them back into the original equation!

These steps work well as long as you have a simple equation with just one absolute value expression. However, as soon as you try solving equations with more than one absolute value sign, or inequalities containing the absolute value sign, things can get confusing. For this reason you may want to look beyond the steps to see what the absolute value sign is actually doing to the expression inside of it. That takes a bit longer, but it really helps you to understand this topic better.

### Example

Consider the equation  $|x + 1| = 5$ .

The absolute value sign always creates two possibilities. The first possibility is that  $x + 1$  was positive or zero, so the absolute value sign did nothing to it. It was unchanged, so we could just remove the absolute value bars. The second possibility is that the expression was negative and the absolute value sign made it positive. When the expression  $x + 1$  is changed by the absolute value sign, it turns into the opposite:  $-(x + 1)$

For  $x + 1 \geq 0$

$$x + 1 = 5$$

$$x = 4$$

For  $x + 1 < 0$

$$-(x + 1) = 5$$

$$x + 1 = -5 \quad [\text{multiplying both sides by } -1]$$

$$x = -6$$

The answers are:  $x = 4$  or  $x = -6$

### Example

We can still use the same strategy if we have an equation like this:  $|x - 7| + 3 = 5$ .

There is no need to isolate the absolute value part now, because we are dealing with it directly:

For  $x - 7 \geq 0$

$$x - 7 + 3 = 5$$

$$x - 4 = 5$$

$$x = 9$$

For  $x - 7 < 0$

$$-(x - 7) + 3 = 5$$

$$-x + 7 + 3 = 5$$

$$-x + 10 = 5$$

$$-x = -5$$

$$\text{or} \quad x = 5$$

Notice that these solutions match the conditions that we set for them. If  $x - 7 \geq 0$  then we can add 7 to both sides to get  $x \geq 7$ . The solution  $x = 9$  meets the condition  $x \geq 7$ . In the same way, if  $x - 7 < 0$  then  $x < 7$ , and the solution  $x = 5$  is reasonable. Both solutions make the equation  $|x - 7| + 3 = 5$  true.

To practice solving absolute value equations, you should create your own. Pick a value for  $x$ , make the problem, and then pretend you don't know what  $x$  is. If you don't get the original value for  $x$  when you solve the problem, troubleshoot carefully. Create a simpler problem if things don't work out. That may seem like a lot of work, but it's worth it.

If the unknown value  $x$  is present more than once in an equation containing absolute value signs, there could be *extraneous solutions*. Extraneous solutions are the results of your correct calculations that nevertheless do not fit back into the original equation. These extraneous solutions are caused by sign changes produced by the absolute value sign; they do not appear if the absolute value sign is removed from the equation before you start solving it.

### Example

Solve  $|x + 6| = 2x + 15$ .

If the expression " $x + 6$ " is positive or zero, the absolute value sign does nothing, so we try removing it. For  $x + 6 \geq 0$ :

$$x + 6 = 2x + 15$$

$$6 = x + 15$$

$$-9 = x$$

Already we know that this will not work, because if  $x = -9$  then the expression  $x + 6$  is not positive.  $x + 6 \geq 0$  means  $x \geq -6$ , and  $-9$  doesn't fit that. Checking the whole equation, we get

$$|-9 + 6| = 2(-9) + 15$$

$$|-3| = -18 + 15$$

$$3 = -3.$$

$x = -9$  is an extraneous solution.

If the expression  $x + 6$  is negative, the absolute value sign changed it to its opposite,  $-(x + 6)$ :

$$-(x + 6) = 2x + 15$$

$$-x - 6 = 2x + 15$$

$$-6 = 3x + 15$$

$$-21 = 3x$$

$$x = -7$$

This solution makes the expression  $x + 6$  negative as we had assumed (if  $x + 6 < 0$  then  $x < -6$ ).

Inserting  $x = -7$  into the original equation gives

$$|-7 + 6| = 2(-7) + 15$$

$$|-1| = -14 + 15$$

$$1 = 1.$$

$x = -7$  is the correct solution.

### Example

Let's look at an equation that contains not one but **two** absolute value signs:

$$|x - 4| = |2x - 2| - 2$$

This actually creates four distinct possibilities! We'll look at each one in turn.

- Both expressions inside the absolute value signs were positive or zero so they did not change.

$$x - 4 = 2x - 2 - 2$$

$$x - 4 = 2x - 4$$

$$x = 2x \quad \text{Subtract } 2x \text{ from each side:}$$

$$-x = 0$$

$x = 0$ . We can see that this solution will not work. If  $x - 4$  actually was a positive number or zero then  $x$  could never be zero; it has to be at least 4. This is an extraneous solution. If you substitute it into the original equation you can see that it won't work.

2. The first expression,  $x - 4$ , was positive or zero while the second expression,  $2x - 2$ , was negative. The first expression stayed the same, but  $2x - 2$  was changed by the absolute value sign:

$$x - 4 = -(2x - 2) - 2$$

$$x - 4 = -2x + 2 - 2$$

$$x - 4 = -2x$$

$$x = -2x + 4$$

$$3x = 4$$

$x = 4/3$ . Again we get an extraneous solution because this value is less than 4.

3. The first expression was negative and the second is positive or zero:

$$-(x - 4) = 2x - 2 - 2$$

$$-x + 4 = 2x - 4$$

$$-x = 2x - 8$$

$$0 = 3x - 8$$

$$-3x = -8$$

$x = 8/3$ . This value meets both requirements, making  $x - 4$  negative and  $2x - 2$  positive. It works as a solution when you substitute it back into the original equation.

4. Both expressions were negative:

$$-(x - 4) = -(2x - 2) - 2$$

$$-x + 4 = -2x + 2 - 2$$

$$-x + 4 = -2x$$

$$-x = -2x - 4$$

$$0 = -x - 4$$

$x = -4$ . This value makes both  $x - 4$  and  $2x - 2$  negative. It is a valid solution and it makes the original equation true.

Note that if the problem had contained only the absolute value expressions and nothing else, the solution would have been simpler. If we were solving  $|x - 4| = |2x - 2|$ , then we could see that  $-(x - 4) = -(2x - 2)$  is the same as  $(x - 4) = (2x - 2)$ , and  $-(x - 4) = (2x - 2)$  is the same as  $(x - 4) = -(2x - 2)$ . The only thing that matters here is if the expressions had the same sign before the absolute value operation, or if they had different signs. There are only two possibilities.

## Absolute Value Functions and Piecewise Functions

Of course, where there is an equation there can be a function. Recall that functions are relations that have a unique output for every input. First draw the straight line  $y = x$ . Then graph  $y = |x|$ . This function takes the part of the line  $y = x$  that is to the left of the y-axis and makes it positive. The result is a V-shaped graph. The point of the V is located at (0,0). To graph absolute value functions on a TI-84 graphing calculator, press the “Y=” button as usual. Then press the MATH button and use the right arrow key to get to the NUM menu. The first choice on this menu is “abs(“, and once you select it you will be able to enter  $\text{abs}(x)$ .

To graph absolute value functions using [MathGV](#), select File → New 2D Cartesian Graph. Then select Graph → New 2D Function. Type in  $\text{abs}(x)$ .

Turning the expressions on each side of an absolute value equation into functions and graphing them allows you to check your solutions. The x values of the intersect points are the solutions to the equation. For the example in the previous section, graph  $y = \text{abs}(x-4)$  and  $y = \text{abs}(2x-2) - 2$ .

Here are some examples of absolute value equations that have 2 solutions, one solution, and no solutions:

$$|x + 2| = 2$$

$$|x + 2| + 1 = -7|x + 2| + 1$$

$$|x + 2| + 1 = -|x - 3|$$

If you graph these equations as two separate functions you can see why.

The graphs that you just created are all V-shaped, and similar to  $y = |x|$ . To create a narrower “V”, we can take the line  $y = 2x$  and apply the absolute value sign:  $y = |2x|$ . Since the number 2 is always positive, we could also write that as  $y = 2|x|$ . Now try to see if you can flatten the V shape of  $y = |x|$  by modifying the function. There has to be something you can use to multiply by that will make the slope less steep.<sup>2</sup>

The V shapes created by these functions are easy to turn upside down. To do so all we need is to give every y value a minus sign:  $-y = |2x|$ . Rewrite that as  $y = -|2x|$ , or  $y = -2|x|$ , to get a standard function. To emphasize that it is a function, you could write it as  $f(x) = -|2x|$ . The advantage of this notation is that you can show what value of x you used to get a particular

---

<sup>2</sup> Multiply by a number less than one, as in  $y = 0.5|x|$

result. For  $y = -|2x|$ , when  $x = -5$ , we just say that  $y = -10$ . With function notation like  $f(x) = -|2x|$ , we can say that  $f(-5) = -10$ .

It is also not difficult to take a V-shaped function like  $y = |x|$  and shift it up or down. All we need to do is add something to the function or subtract something from it. If you graph  $y = |x| - 4$  you will see that the whole graph has shifted down by 4 units, and that the point of the V is now located at  $(0, -4)$ . Notice that  $y = |x| - 4$  is really the same as  $y + 4 = |x|$ . Adding 4 to the value of  $y$  caused the function to shift down.

In the same way, *adding* something to  $x$  will cause the graph to shift left, toward the negative  $x$ -direction.

### Example

Look at the graph of  $y = |x + 2|$  that you created earlier. You might think that adding two units to  $x$  would cause the graph to move 2 units to the right, but it shifts to the left instead. It may help you to think of it this way: by adding 2 units to  $x$  we have made it possible for a smaller  $x$  to do the same job that a larger  $x$  did before. The graph shifts toward the smaller  $x$  values.

### Example

Predict what will happen if you graph  $y = |x - 3| - 5$ .

Subtracting 3 from  $x$  will cause the original base function  $y = |x|$  to shift to the right, since a larger  $x$  is now needed to do the same job that  $x$  did before. The function moves 3 units in a positive horizontal direction. Subtracting 5 from the total function value (which is the same as adding 5 to  $y$ ) will cause the function to shift down by 5 units.

If you graph this function by hand, you can see that, like all absolute value functions, it has two parts:  $y = x - 3 - 5$  for  $x - 3 \geq 0$  (which means that  $x \geq 3$ ), and  $y = -(x - 3) - 5$  for  $x - 3 < 0$  (which means that  $x < 3$ ). We can simplify this and then write it in **piecewise notation**:

$$\begin{aligned} y &= x - 8 && \text{for } x \geq 3 \\ y &= -x - 2 && \text{for } x < 3 \end{aligned}$$

Or more properly as a **piecewise function**:

$$f(x) = \begin{cases} x - 8 & \text{for } x \geq 3 \\ -x - 2 & \text{for } x < 3 \end{cases}$$



Note the use of a bracket,  $\{$ , to indicate that this is a single function that follows two different rules depending on the input value of  $x$ . Piecewise functions may have more than two parts, and are used for situations other than those involving absolute value. For example, you may get a better price on some items if you buy larger quantities.

**Practice**

1. Solve for  $x$ :  $|x - 4| = |2x - 2|$ . Use graphing software to check your work.
2. Write  $y = -5|x + 6| + 10$  as a piecewise function. Use graphing software to check your work.

## Inequalities

When you multiply or divide by a negative number, turn the sign around!

$|x| > 3$  means  $x > 3$  or  $-x > 3$ . That works out to  $x > 3$  or  $x < -3$ .

$|x| < 3$  means  $-3 < x < 3$

If you need  $x$  to be within a certain range, find the midpoint of that range:

$|x - \text{midpoint}| \leq \frac{1}{2} \text{ the range.}$

To solve a non-linear inequality it is usually best to rearrange to get a zero on one side.

Look where the equality points are, then use test values to see where the inequality holds.

Factoring can help you find the equality points.

## Graphing Inequalities

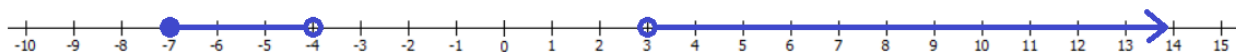
Inequalities usually describe a range of numbers. For example,  $x \leq -3$  means all values of  $x$  less than or equal to  $-3$ . That expression describes the infinitely many solutions for this inequality quite well, but there are several other ways to write the same thing. **Interval notation** shows the solutions by marking their boundaries:  $(-\infty, -3]$ . The interval extends from negative infinity to  $-3$ . The square bracket at the end indicates that  $-3$  is included in the solution. If you wanted to show  $x < -3$ , which means  $x$  is smaller than  $-3$ , you would write  $(-\infty, -3)$ . Because infinity is not a number, it can never be “included” as if the interval ended there. Always use a curved parenthesis before or after the infinity symbol. Another way to write  $x \leq -3$  is by using **set-builder notation**. This shows the set of all numbers  $x$  such that  $x$  is less than or equal to  $-3$ :  $\{x \mid x \leq -3\}$ . Your teacher may also want you to graph solutions like this on a number line. Just remember the conventions for that: a closed circle means that the indicated number is included in the solution, while an open circle means that the number is not part of the solution. In this case you need a closed circle at  $-3$ :



The arrow indicates that the line continues to the left infinitely far.

Multiple inequalities are often used to indicate a specific range of numbers.  $4 < x \leq 10$  means all real numbers that are larger than 4 and smaller than or equal to 10. In interval notation that would look like  $(4, 10]$ . “Between” generally means not including the numbers on either side, but when you only use words you should be specific about whether those numbers are included or not. If you want to indicate all numbers between 1 and 10, not including 1 and 10, and you need to exclude the number 7, you can split the interval into two parts. To join those parts we use “or” between the two inequalities:  $1 < x < 7$  or  $7 < x < 10$ . This says that we want all the  $x$  values that meet either the first condition or the second condition. To indicate the same thing in interval notation, we use the  $\cup$  symbol (union symbol) to join the two intervals:  $(1, 7) \cup (7, 10)$ . Graph this as a line between 1 and 10, using open circles at 1, 7 and 10.

The image below shows  $-7 \leq x < -4$  or  $x > 3$ , which may also be written as  $[-7, -4) \cup (3, \infty)$ .



Occasionally we may want to indicate numbers that are in the overlapping region of two inequalities. Here we use the word “and” to indicate that we want numbers that can meet both conditions. For example,  $-3 < x < 10$  and  $-2 < x \leq 0$  means only those numbers that are greater than negative 2 and less than or equal to zero. For interval notation we use the intersect symbol:  $(-3, 10) \cap (-2, 0]$ . This says find the numbers in the intersection of the two sets. Graph only this solution, not both inequalities!



## Solving Inequalities

Consider the inequality  $4 > 3$ . Now multiply both sides by  $-1$ . You get  $-4 < -3$ .

Solving inequalities works like solving equations, except that **the inequality sign turns around when you multiply or divide by a negative number**. That is, the reality represented by the sign turns around, but unfortunately the sign itself needs to be adjusted manually ☹. **Don't forget to do this!!**

While equations usually have no more than a few solutions, inequalities tend to have infinitely many. A simple inequality like  $3 - 2x \geq 9$  solves as  $-2x \geq 6$ , so  $x \leq -3$ . The division by  $-2$  meant that the sign had to be adjusted.

Whether you are solving equations or inequalities, remember that dividing by zero is a big no-no. So, if you are dividing both sides by  $x$ , the first thing you must do is to make sure that  $x$  is not zero. Consider this possibility separately, before you do the division. If you must divide by  $x$  when solving an inequality rather than an equation, there is an additional factor to consider, since  $x$  could be either positive or negative. So what happens to the  $<$  or  $>$  sign, does it turn around or not? Well, that depends, so you have to consider each possibility separately.

### Example

Solve  $x^2 - 2x > 0$ .

We can add  $x$  to both sides without creating any problems:

$$x^2 > 2x$$

Next, we would like to divide both sides by  $x$ . First, consider whether  $x$  could be zero. No, that wouldn't work because  $0^2$  is the same as  $2 \cdot 0$ . Now we can go on and do a division. First, consider what happens if  $x$  is positive:

For  $x > 0$ :

$$\frac{x^2}{x} > \frac{2x}{x}$$

$$x > 2$$

So, if  $x$  is a positive number, then it must be greater than 2. Next consider the case where  $x$  is a negative number:

For  $x < 0$ :

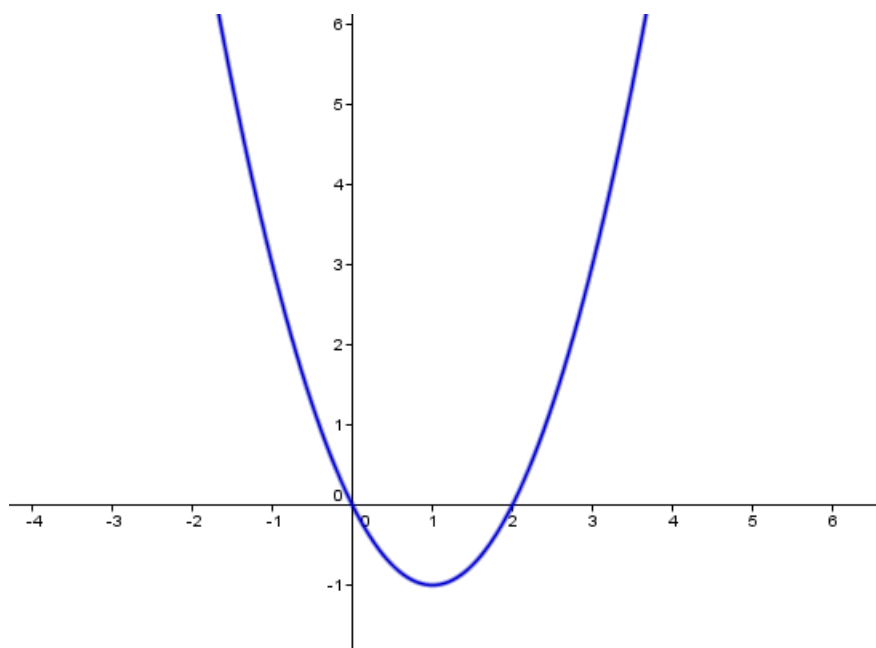
$$\frac{x^2}{x} > \frac{2x}{x}$$

$$x < 2$$

Notice that this doesn't just say that  $x$  is less than 2; it says that if  $x$  is negative it must be less than 2. This excludes numbers like 1 and 0, but includes all negative numbers.

Our final solution is that  $x < 0$  or  $x > 2$ .

The image below shows the function  $y = x^2 - 2x$ . You can see that the value of the function is greater than 0 everywhere except between 0 and 2:



Practice solving some basic inequalities by creating your own, like maybe  $5x + 3 > 10x - 22$ .

Check your solutions by picking an  $x$ -value near the equal point. For example, if your solution is  $x > 5$ , use  $x = 6$  to check your work.

## Inequalities Involving Absolute Value

If you follow the method outlined in the section on absolute value, these inequalities should not be too difficult to manage.

### Example

In our first example, the part with the absolute value is greater than something:

$$|5x + 2| > 12$$

There are two possibilities. First,  $5x + 2$  was a positive quantity or zero and the absolute value sign did not change it.  $5x + 2 \geq 0$ , so  $5x \geq -2$  and  $x \geq -\frac{2}{5}$

For  $x \geq -\frac{2}{5}$ :

$$5x + 2 > 12$$

$$5x > 10$$

$$x > 2.$$

This solution is compatible with  $x$  being greater than or equal to  $-\frac{2}{5}$ , so  $x > 2$  is valid. Using interval notation we would write that as  $(2, \infty)$ , which says that  $x$  is in the interval from 2 to infinity with the number 2 itself not included. If we wanted to include the number 2 we would use the notation  $[2, \infty)$ . Notice that since “infinity” is not a number,  $\infty$  itself is never marked as included in the interval.

The other possibility is that  $5x + 2$  was a negative quantity that was made positive by the absolute value sign:

$$-(5x + 2) > 12$$

$$(5x + 2) < -12$$

Remember to turn the inequality sign around when you divide by  $-1$ . Notice that now you no longer need the parentheses:

$$5x + 2 < -12$$

$$5x < -14$$

$$x < -\frac{14}{5}.$$

There is no problem with the solution  $x < -\frac{14}{5}$ ; no values of  $x$  less than  $-\frac{14}{5}$  will interfere with the expression  $5x + 2$  being negative since  $5x + 2 < 0$  means  $x < -\frac{2}{5}$ . Using interval notation we would write  $(-\infty, -\frac{14}{5})$ .

Our final solution is:  $x < -\frac{14}{5}$  or  $x > 2$ .

### Example

Now let's look at a situation where the part with the absolute value is less than something:

$$|x| < 3.$$

If you stop and think carefully, you will realize that the actual value of  $x$  must be between  $-3$  and  $3$ , so we could write  $-3 < x < 3$ . Personally I tend to be too lazy to think carefully every time I do this stuff, so let me first do a problem the same way we did before.

$$|2x + 4| < 10$$

Consider the possibility that  $2x + 4$  was positive or zero, so it is unchanged:

$$2x + 4 < 10.$$

$$2x < 6, \text{ so } x < 3.$$

You may think that this solution implies that  $x$  can be any value so long as it is less than  $3$ . However, the condition for obtaining this answer is that  $2x + 4$  is positive or  $0$ . Only some values for  $x$  will make that true:

$$2x + 4 \geq 0$$

$$2x \geq -4$$

$$x \geq -2$$

Provided that  $x$  is bigger than or equal to  $-2$ , the solution of the inequality is  $x < 3$ . Therefore the permissible values of  $x$  are between  $-2$  and  $3$ :  $-2 \leq x < 3$ .

Now think about the possibility that  $2x + 4$  was negative:

$$-(2x + 4) < 10$$

$$2x + 4 > -10$$

$$2x > -14$$

$$x > -7$$

Notice that I marked a minus sign in red in this solution too, because it is easy to forget it when you are concentrating on turning your inequality sign around. In order for  $2x + 4$  to be negative, there is another restriction on  $x$ :

$$2x + 4 < 0$$

$$2x < -4$$

$$x < -2$$

Provided that  $x < -2$ , the solution to the inequality is  $x > -7$ . This part of the solution gives us  $-7 < x < -2$

Combining our two solutions we get  $-7 < x < 3$

Here is the more sophisticated, and much faster, way to do the same problem:

$|2x + 4| < 10$  really means  $-10 < 2x + 4 < 10$ . Do not be intimidated by the two inequality signs. Just follow the logical steps and subtract 4 from all parts of the inequality so it is still true:  $-14 < 2x < 6$ . Next, divide by 2 all the way along:  $-7 < x < 3$ . Notice that this is a lot easier, so it does help to put some thought into it up front.

To graph the solution for this example you need two open circles because the numbers -7 and 3 are not included:



Your graphing calculator or graphing software can help you check your answers. Turn the expressions on each side of the inequality into functions and graph them. If you graph  $y = \text{abs}(2x+4)$  and  $y = 10$ , you will see that the inequality is true when  $-7 < x < 3$ . The  $x$ -values of the intersect points are -7 and 3.

If you have an expression like  $4|2x + 5| - 3 < 9$ , first change it so that the absolute value sign is on one side by itself:  $4|2x + 5| < 12$ , and  $|2x + 5| < 3$ . Now we know that  $2x + 5$  must be between -3 and 3:  $-3 < 2x + 5 < 3$ . Subtract 5 from both sides:  $-8 < 2x < -2$ , so  $-4 < x < -1$ . Alternatively, you can do it the long way around and get the same answer.

Practice solving some basic linear inequalities (no exponents on  $x$ ) by creating your own. Use graphing software to check your solutions.



## Tolerance Levels

Sometimes you may be asked to create your own absolute value expressions. Let's look at a real-life example. The other day I bought a pack of six 500ml soft drink bottles. When I got home I noticed that one of these sealed bottles contained noticeably less liquid than the others. "What a rip off!" and "Is that safe to drink?" were the reactions from my family. Manufacturers normally try to avoid situations like this by establishing tolerance levels for the filling process. If all 500ml bottles contain between 499 and 501 milliliters consumers are unlikely to notice the difference, because a milliliter is a small amount of liquid that fits in a cube with sides of 1cm. Another way of writing this tolerance level is to say that bottles should contain  $500\text{ml} \pm 1\text{ml}$ . The difference can be positive or negative, so long as it is not more than 1 ml. This is an ideal situation for the use of an absolute value sign:

$$|\text{amount of liquid in bottle} - 500\text{ml}| \leq 1\text{ml}$$

Your word problem might say: create an absolute value equation that shows that bottles of a certain brand of soft drink should be filled to contain between 499 and 501 ml. To create an inequality like this from a word problem, find the midpoint of the acceptable range of values. In this case the acceptable range is from 499 to 501, so the midpoint is 500 and the range is  $501 - 499 = 2$ . The difference between the actual amount and the midpoint should not be more than half the range:

$$|\text{actual value} - \text{midpoint}| \leq \frac{1}{2} \text{ the range.}$$

### Example

The safe storage temperature for a vaccine is between 1 and 4 degrees Celsius. Find an absolute value expression for the storage temperature.

There is a range of 3 degrees for the safe temperature, and the midpoint of that range is 2.5 degrees Celsius.

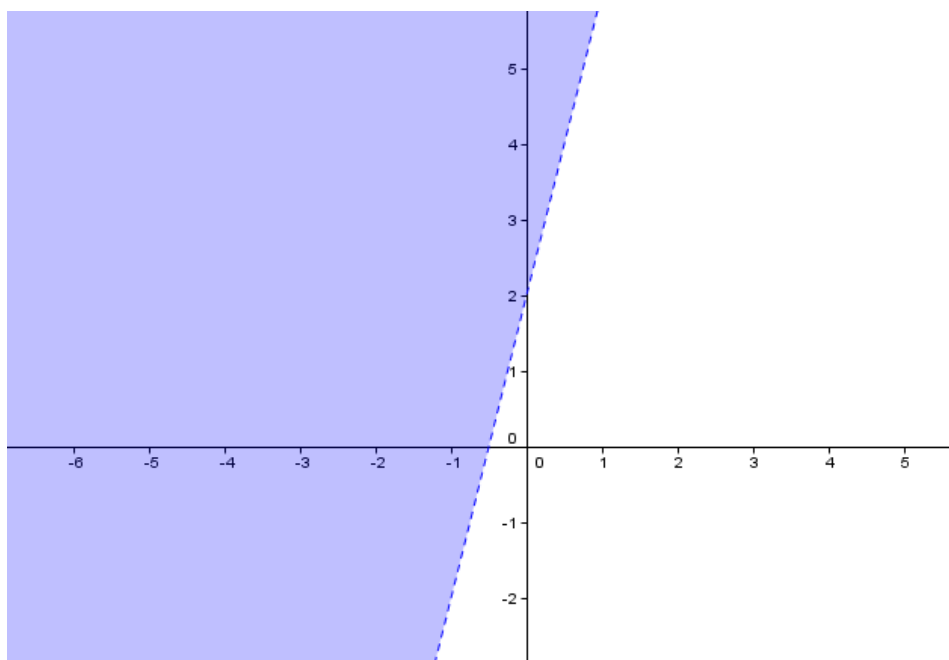
$$|\text{actual temperature} - 2.5 \text{ degrees C}| \leq 1.5 \text{ degrees C.}$$

This says that we would like the storage temperature to be 2.5 degrees C, with no more than 1.5 degrees of variation on either side.

## Linear Inequalities

Linear equations are equations like  $y = 4x + 2$  that graph as a line. The general form of a linear equation is  $y = mx + b$ , where  $m$  is the slope, and  $b$  is the y-intercept. If you need some review on working with linear equations, check out “Caution! Steep Slope Ahead” and “Linear Equations” in the [Pre-Algebra / Algebra 1](#) e-book.

Linear inequalities are very similar to linear equations. To graph them, first graph the line of the corresponding linear equation. Then consider if you want  $y$  to be bigger than that or smaller. If  $y$  is larger than the value indicated by the line, shade the area above the line to show the solution of the inequality. Otherwise shade the area below the line. If the values on the line are not included in the inequality, as in  $y > 4x + 2$ , draw a dashed line to represent  $y = 4x + 2$ . Then shade the area above the line:



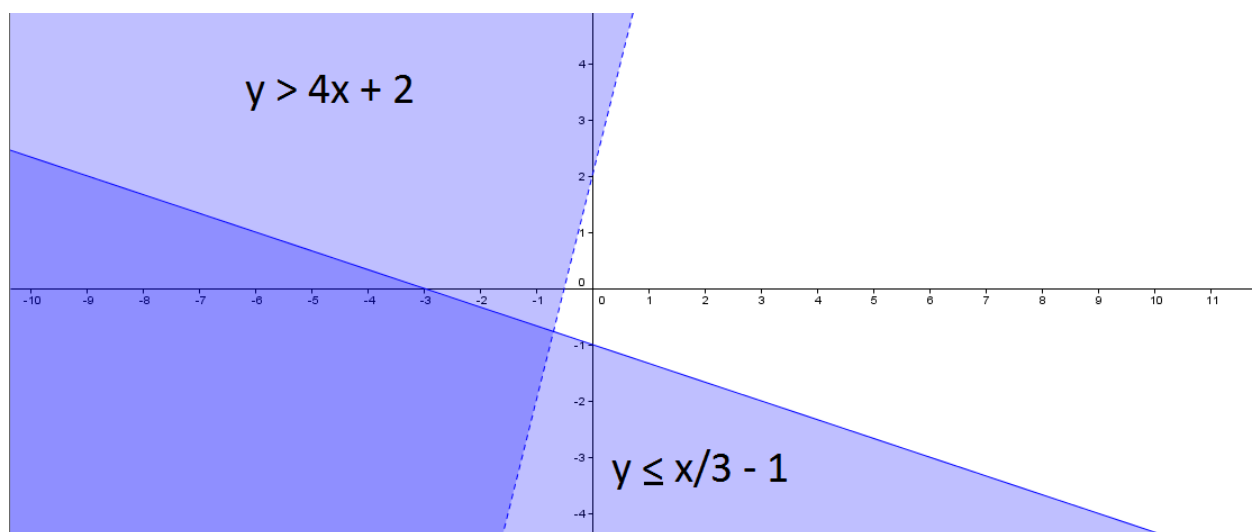
If the values on the line are to be included, use a solid line instead.

When you solve a system of two equations, the solution is usually a single point, and it is found where the two lines intersect. You can solve a system of two inequalities, like

$$y > 4x + 2$$

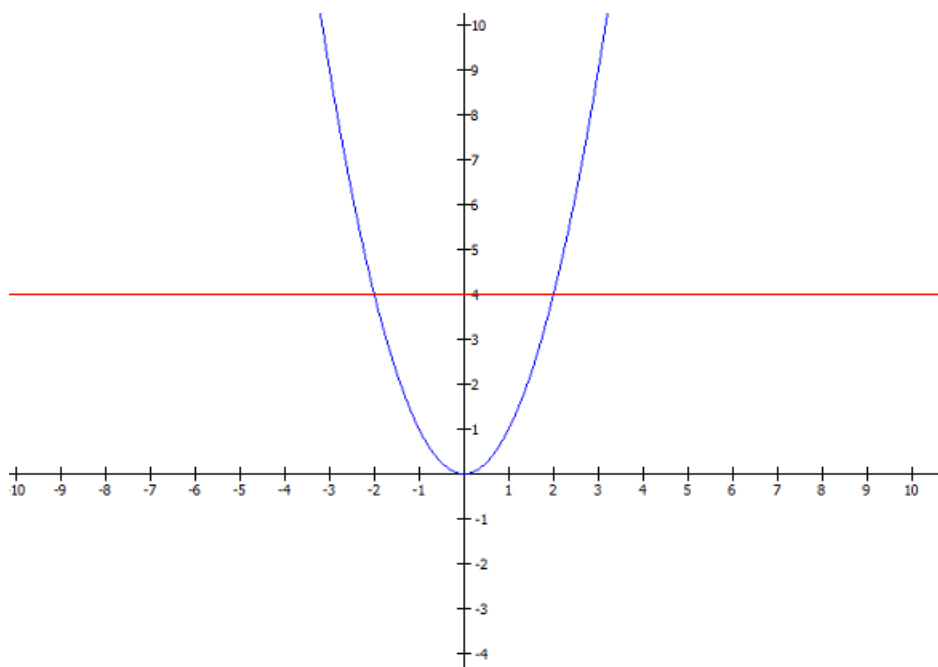
$$y \leq -\frac{1}{3}x - 1$$

If there is a solution, the result will be a region of overlap:



## Non-Linear Inequalities

For those inequalities that involve higher powers of  $x$ , it really helps to think of the problem as a comparison between two functions. Let's look at something we can draw a simple picture for:  $x^2 < 4$ . We can think of that this way:

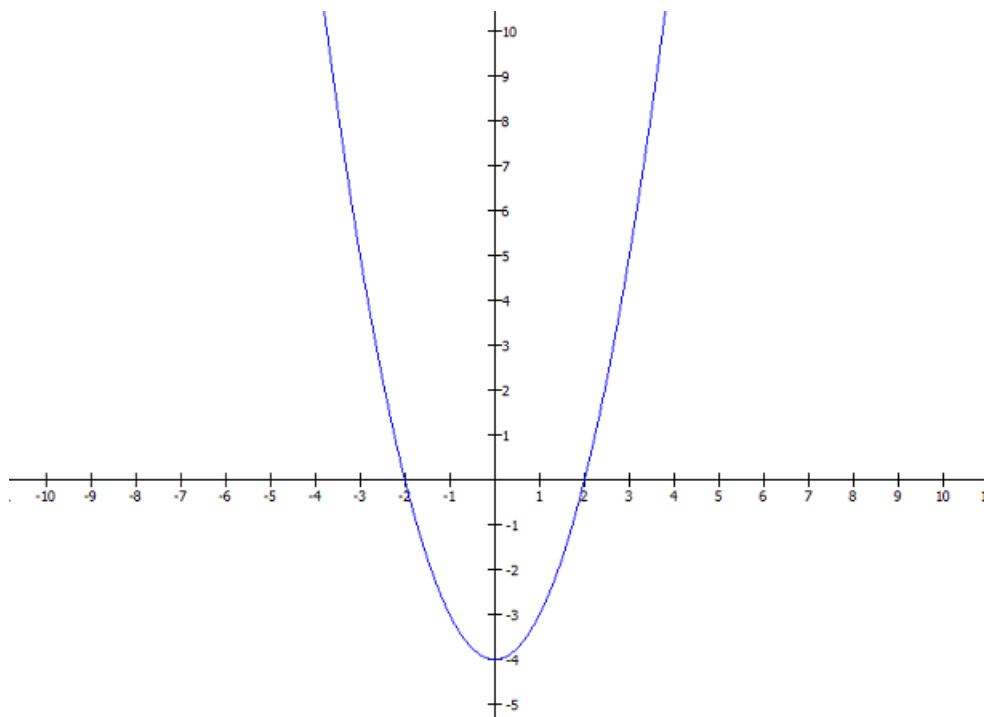


Here we are visualizing the function  $y = x^2$ , and asking where the function value is less than 4. Notice that the red line represents the function  $y = 4$ . We can see that  $x$  should be between -2 and 2 by looking at the picture. It is also easy to solve  $x^2 < 4$  without a picture. If you don't see the answer right away, you can first solve  $x^2 = 4$ , which gives  $x = 2$  or  $x = -2$ . Where would  $x^2$  be less than 4? Well, that should either be in the interval  $(-2, 2)$ , or outside of it. Here you can use test values. Pick a value in the interval, like  $x = 0$ . Now the inequality works. Pick one below or above the interval, like -3 or 3, and you get  $9 < 4$  which is not true.

This is a good general way to solve an inequality. First look where the equality points are, and then use some test values to see where the inequality holds.

Because we are doing non-linear inequalities here, it is often to our advantage to rearrange them to get an inequality that is less than or greater than zero. This makes the equality points much easier to find.  $(x - 2)(x + 3) = 0$  is much easier to solve than  $(x - 2)(x + 3) = 4$ . For the first equation we can say that  $x - 2$  is zero or  $x + 3$  is zero, or both are zero. However, for the second equation we cannot say that  $x - 2 = 4$  or  $x + 3 = 4$ !

To illustrate, I have changed the inequality  $x^2 < 4$  to  $x^2 - 4 < 0$ , and adjusted the picture:



We can think of this as taking the function  $y = x^2 - 4$ , and asking where the function values are negative. First look at where the function is equal to zero.  $x^2 - 4 = 0$  can be factored as

$(x + 2)(x - 2) = 0$ . This shows that the equality points are still -2 and 2. The function touches the x-axis at  $x = 2$  and  $x = -2$ , so if it is going to change from positive to negative or negative to positive that would happen at these points. We can quickly see where the function is less than zero by using test values or a graphing calculator. Test values require that you test in between the equality points and on both sides of them. In this case you might use -3, 0 and 3.

When you use a test value, you can insert it into the original inequality, into the adjusted form that has 0 on one side, or into the factored form. For example, a test value of 3 can be put into  $x^2 < 4$ , into  $x^2 - 4 < 0$ , or  $(x + 2)(x - 2) < 0$ , all with the same result. However, if you are working with more complex inequalities and you don't have a calculator handy, it may be to your advantage to put your test values into the factored form. That way you do not have to actually calculate specific values. Instead, you can just look for positives and negatives, like this:

$$(-3 + 2)(-3 - 2) < 0$$

The first part,  $(-3 + 2)$ , is negative, and the second part,  $(-3 - 2)$ , is also negative. Multiplying two negative quantities will always produce a positive number, so we can tell right away that -3 doesn't work. A test value of 0 gives us  $(0 + 2)(0 - 2)$ , which produces the negative value we want. A positive value results when we test  $x = 3$ . Report your answer as the interval  $(-2, 2)$ . Note that the endpoints of the interval are not included, because the inequality has a  $<$  sign and not a  $\leq$  sign. Including -2 or 2 would make  $(x + 2)(x - 2)$  equal to zero.

Sometimes inequalities don't factor quite so nicely, and the equality points end up being icky fractions, like this:  $(3x - 2)(7x - 5) > 0$ . When you solve  $(3x - 2)(7x - 5) = 0$ , you get  $x = \frac{2}{3}$  or  $x = \frac{5}{7}$ . Now it is not so easy to get a test value that is in between  $\frac{2}{3}$  and  $\frac{5}{7}$ . Here we can cheat a little and work with the equality points themselves.

First test a value that is just a little tiny bit less than  $\frac{2}{3}$ .  $(3x - 2)$  is 0 at  $x = \frac{2}{3}$ , but if you use your imagination you can see that it would be negative if you put in something for  $x$  that is just a little smaller than  $x = \frac{2}{3}$ . When  $x$  is  $\frac{2}{3}$ ,  $(7x - 5)$  works out to  $\frac{14}{3} - 5$ . We would need  $\frac{15}{3} - 5$  to get 0, so  $\frac{14}{3} - 5$  is negative, and it is still negative slightly below  $\frac{2}{3}$ . Now we have a negative quantity times a negative quantity, which will give a positive and make the inequality true.

Next, check just a very, very tiny bit above  $\frac{2}{3}$ .  $(3x - 2)$  turns positive as soon as you get even the minutest fraction above  $\frac{2}{3}$ .  $(7x - 5)$  will still be negative at that point, because it requires

an  $x$  value of  $5/7$  or above in order to change its sign.  $(3x - 2)(7x - 5)$  gives a positive times a negative, which is a negative value. The inequality is false when  $x$  is between  $2/3$  and  $5/7$ .

Now we will check just a little above  $5/7$ :  $(3x - 2)$  is still positive, and so is  $(7x - 5)$ . Because we are multiplying two positive values, the inequality is true for  $x > 5/7$ . In fact, the only place where the inequality is not true is between  $2/3$  and  $5/7$ , an interval so small that you can't see the negative part of the function on your graphing calculator unless you zoom in really close.

## Function Graphs

**For  $y = x^2$  and a positive constant ( $c > 0$ ):**

$y = x^2 + c$	Shifts graph up
$y = x^2 - c$	Shifts graph down
$y = (x - c)^2$	Shifts graph right
$y = (x + c)^2$	Shifts graph left
$y = -x^2$	Reflects graph over the x-axis
$y = (-x)^2$	Reflects graph over the y-axis

**For  $y = f(x)$  and a constant  $c > 1$ :**

$y = c x^2$	Stretches graph vertically
$y = \frac{1}{c} x^2$	Compresses graph vertically
$y = \left(\frac{x}{c}\right)^2$	Stretches graph horizontally
$y = (cx)^2$	Compresses graph horizontally

Follow order of operations when you draw the graph or describe the transformations. Do vertical shifts *after* any vertical stretches or compressions. If  $x$  has a minus sign, first consider the entire graph without the minus sign, and then replace  $x$  with  $-x$  to see the reflection over the y-axis.

## Function Transformations

[To graph functions using [MathGV](#), select File → New 2D Cartesian Graph. Then select Graph → New 2D Function. To graph functions using a TI-84 calculator, press “Y=” and enter your function. To insert an  $x$ , press the button labeled “X,T,Ø,n”. Press “GRAPH” to see the graph. If

necessary, adjust the viewing rectangle by pressing “WINDOW”. Alternatively, use a graphing app for your phone or tablet.]

Graph the following functions:

$$y = x$$

$$y = x^2$$

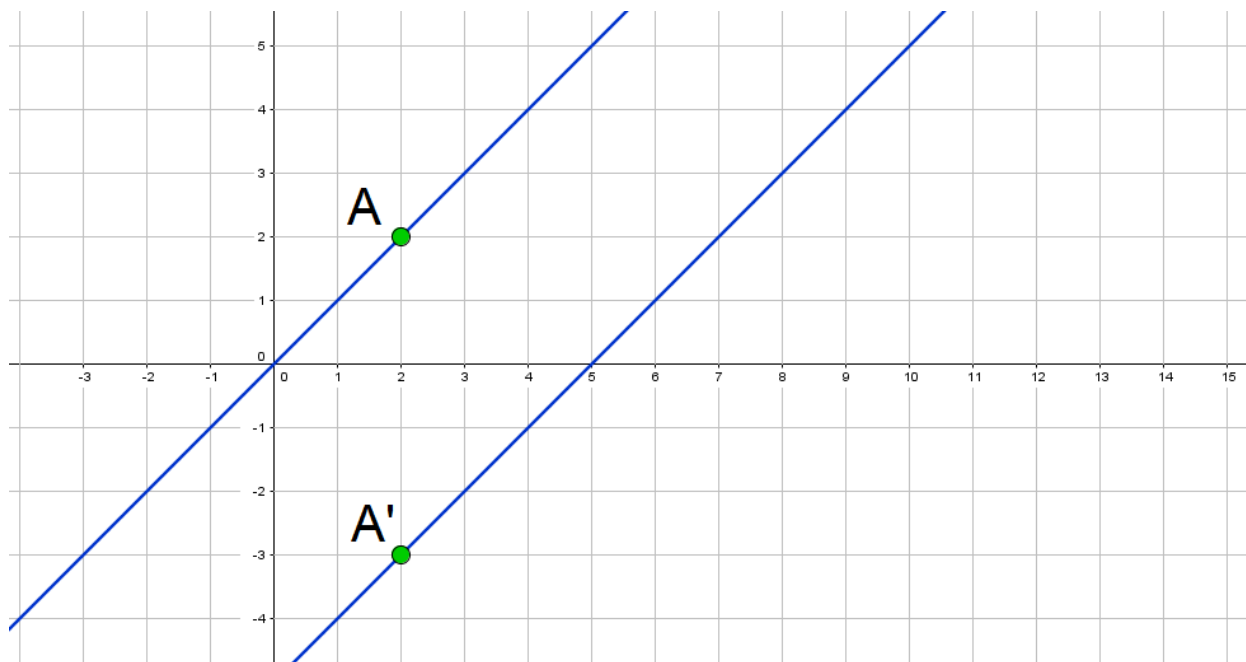
$$y - 2 = x$$

$$y = x^2 + 2$$

$$y = x - 5$$

$$y = x^2 - 5$$

Notice that you have to rearrange some of these functions in order to graph them. When you subtract 2 from  $y$ , this adds 2 to the value of the function, which moves the graph 2 units higher up on the  $y$ -axis. It is more intuitive that  $y = x + 2$  would move the graph up by 2 units. When you add 5 to  $y$ , this subtracts 5 from the value of the function so that the graph  $y = x - 5$  appears 5 units lower down than the original  $y = x$ . To illustrate this change, the point A below has been changed to A' on the lower graph. The  $y$ -coordinate has been decreased by 5 units.



This works the same way for any function  $f(x)$ . If we create a new function  $y = f(x) - 5$ , the graph will move down by 5 units. You may see this written as  $g(x) = f(x) - 5$ , which looks more complicated but still means the same thing.



Graph the following functions:

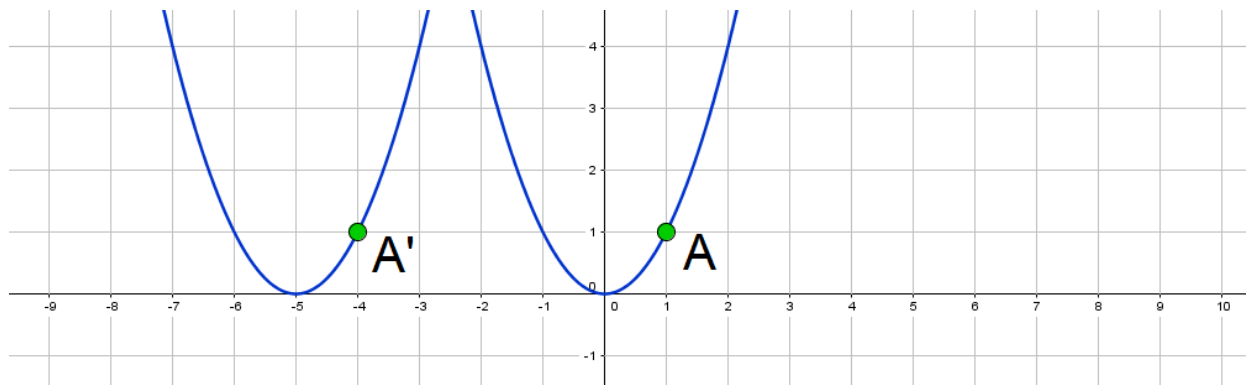
$$y = x^2$$

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

$$y = (x + 5)^2$$

Subtracting 2 from  $x$  shifts the graph 2 units to the right. It now takes a larger  $x$  to do the same job that  $x$  did before, which causes the graph to appear further to the right.

Adding 5 to  $x$  causes the graph to shift 5 units to the left. A smaller  $x$  can now do the same job that  $x$  did before, causing the graph to appear further to the left, as shown below.



The point  $A'$  is 5 units further to the left than the equivalent point  $A$  on the original graph. Again this works for any function  $f(x)$ . If we create a new function  $y = f(x + 5)$  the graph will shift 5 units to the left. You can also write the new function as  $g(x) = f(x + 5)$ .

Try this out with some other functions, like  $y = x^3$ ,  $y = \sqrt{x}$  or  $y = |x|$  [To graph absolute value functions on a TI-84 graphing calculator, press the “Y=” button as usual. Then press the MATH button and use the right arrow key to get to the NUM menu. The first choice on this menu is “abs(“, and once you select it you will be able to enter  $\text{abs}(x)$ . If you are using MathGV, type in  $\text{abs}(x)$ .]

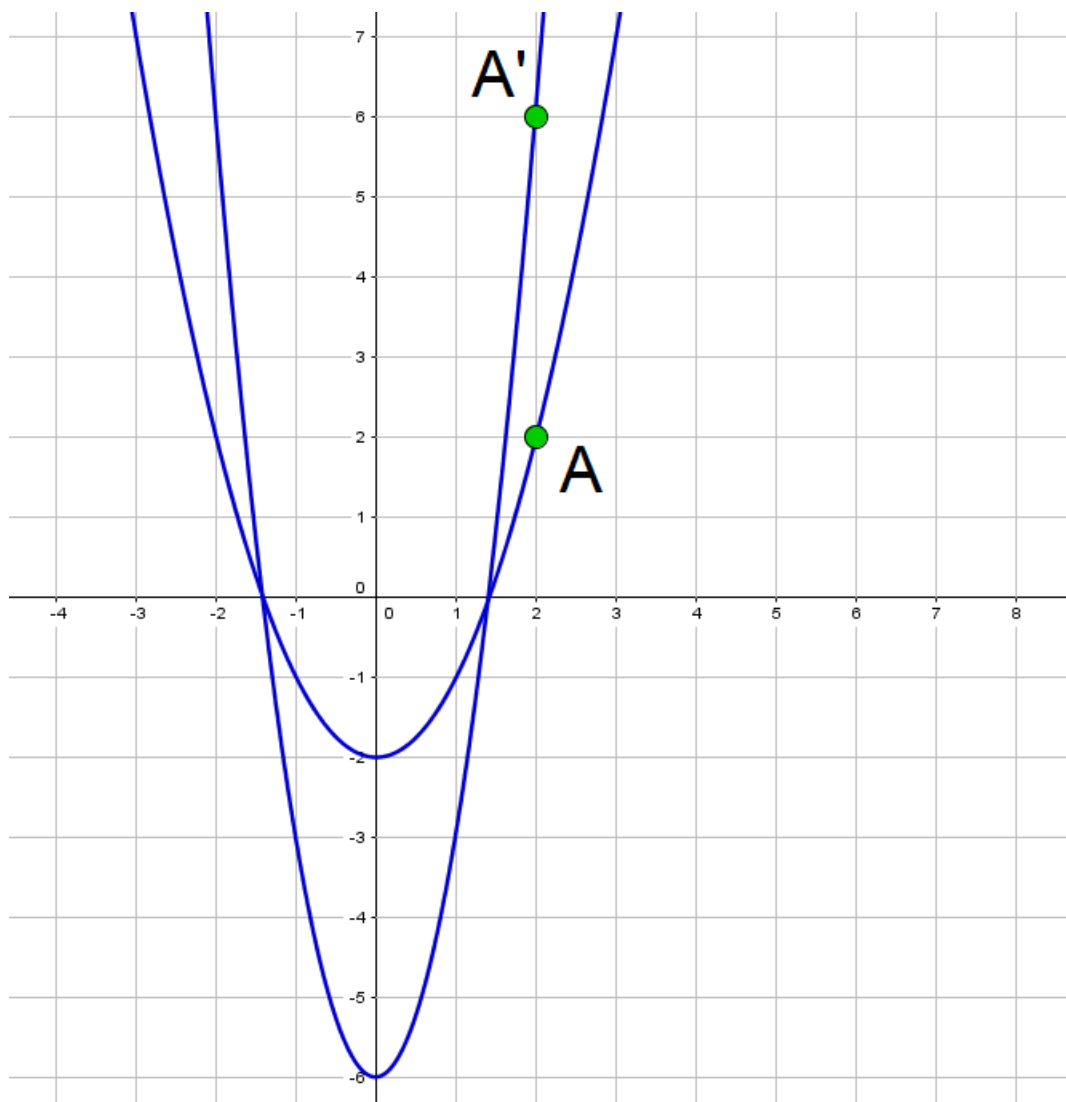
Next try graphing these functions:

$$y = x^2 - 2$$

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

$$y = \frac{x^2 - 2}{3}, \text{ which is the same as } y = \frac{1}{3}(x^2 - 2)$$

It probably seems sensible that multiplying the function by 3 would cause it to stretch vertically by a factor of 3.  $y = 3(x^2 - 2)$  is 3 times larger than  $y = x^2 - 2$ . Notice however that you can also see this as  $y$  being divided by 3:  $\frac{y}{3} = x^2 - 2$ . Now it doesn't seem quite as obvious that this manipulation would cause the function to get bigger in the  $y$ -direction, but of course it does. The image below shows that the  $y$ -coordinate of  $A'$  is three times larger than that of point  $A$ .



In function notation that would look like  $y = 3f(x)$ , or  $g(x) = 3f(x)$ .

Now graph the following functions:

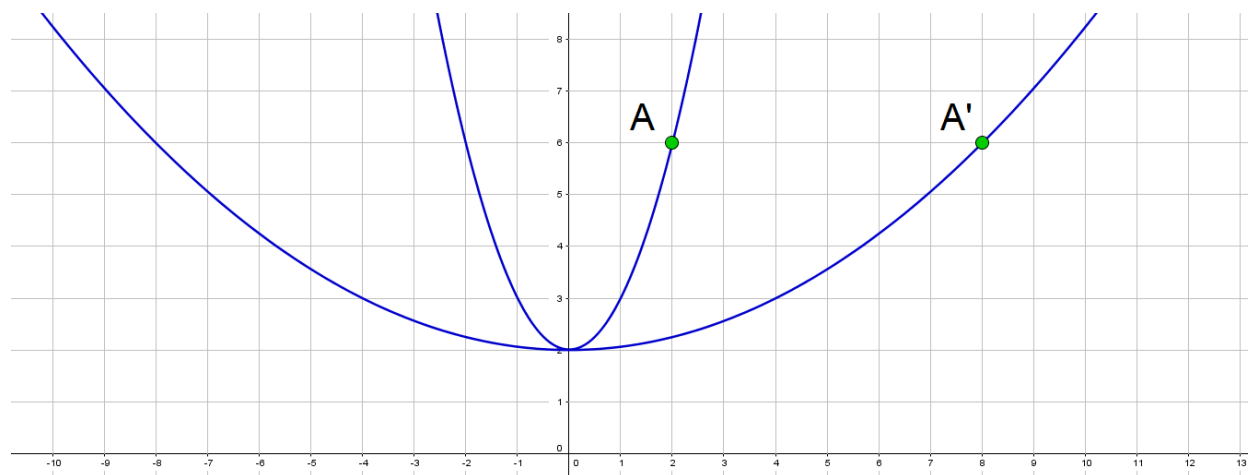
$$y = x^2 + 2$$

$$y = (4x)^2 + 2$$

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

You should notice that when  $x$  alone (rather than the whole function) is multiplied by 4, this causes the entire function to be 4 times narrower. The function *shrinks* horizontally by a factor of 4. That happens because a smaller  $x$  can now do the same job that  $x$  did before.

When  $x$  is divided by 4 the function *stretches* by a factor of 4 in the  $x$ -direction, because you now need a bigger  $x$  to produce the same result. In the image below, the  $x$ -coordinate of point  $A$  on the graph of  $y = x^2 + 2$  is 2, while the equivalent point  $A'$  on the graph of  $y = \left(\frac{x}{4}\right)^2 + 2$  has an  $x$ -coordinate of 8, which is 4 times larger.



To reflect a function over the  $x$ -axis (to turn it upside down), replace  $y$  with  $-y$  and adjust the function:  $y = |x| \rightarrow -y = |x| \rightarrow y = -|x|$ . In function notation, that would be  $g(x) = -f(x)$ .

To reflect a function over the  $y$ -axis (create a mirror image of it), replace  $x$  with  $-x$ :  $y = |x| \rightarrow y = |-x|$  or in general,  $g(x) = f(-x)$ .

Hey, wait a minute; nothing happened on that last one. Or did it...?

## Finding the Vertex and the Axis of Symmetry

The functions  $y = |x|$  and  $y = x^2$  have a bottom point called the vertex. If you flip these functions upside down they have a top point. The vertex is either the minimum or the maximum point of the function. For these basic functions the vertex is located at  $(0,0)$ , but of course it moves if you start manipulating the function.

$y = |x|$  can be shifted up by subtracting something from  $y$ , which is the same as adding something to the right side of the equation. Let's call that something  $k$ :

$$y = |x| + k$$

Here  $k$  could be a negative number, which means that the graph would shift down. The whole graph moves either  $k$  units up or  $k$  units down. That means the vertex moves to the point  $(0,k)$ . [Try that out with some real numbers for  \$k\$  to verify that the vertex does in fact end up at  \$\(0,k\)\$ .](#)

Next, we could subtract something directly from  $x$  to move the graph to the right:

$$y = |x - h| + k$$

However, if we pick a negative number for  $h$  the graph will shift to the left instead. Either way, the whole graph will move  $h$  units horizontally. The vertex will end up at  $(h,k)$ .

To stretch or compress the graph, we can multiply the right side of the equation by some number  $a$ :

$$y = a|x - h| + k$$

This does not affect the position of the vertex, which stays at  $(h,k)$ . [Check it out by picking some random \(small\) values for  \$a\$ ,  \$h\$  and  \$k\$ .](#)

The absolute value function is symmetrical. The line of symmetry, also called the axis of symmetry, is a vertical line that passes through the vertex. For the unmodified function,  $y = |x|$ , the axis of symmetry passes through  $(0,0)$ , and we can describe it by using the equation  $x = 0$ . [This says that  $x = 0$  regardless of the value of  $y$ , so it produces a vertical line through the origin.] When we manipulate the function so that the vertex moves to the point  $(h,k)$ , the axis of symmetry also shifts, to become  $x = h$ .

### Example

[Find the vertex, axis of symmetry, and the  \$x\$  and  \$y\$ -intercepts for the function  \$f\(x\) = 2|x + 5| - 4\$ .](#)

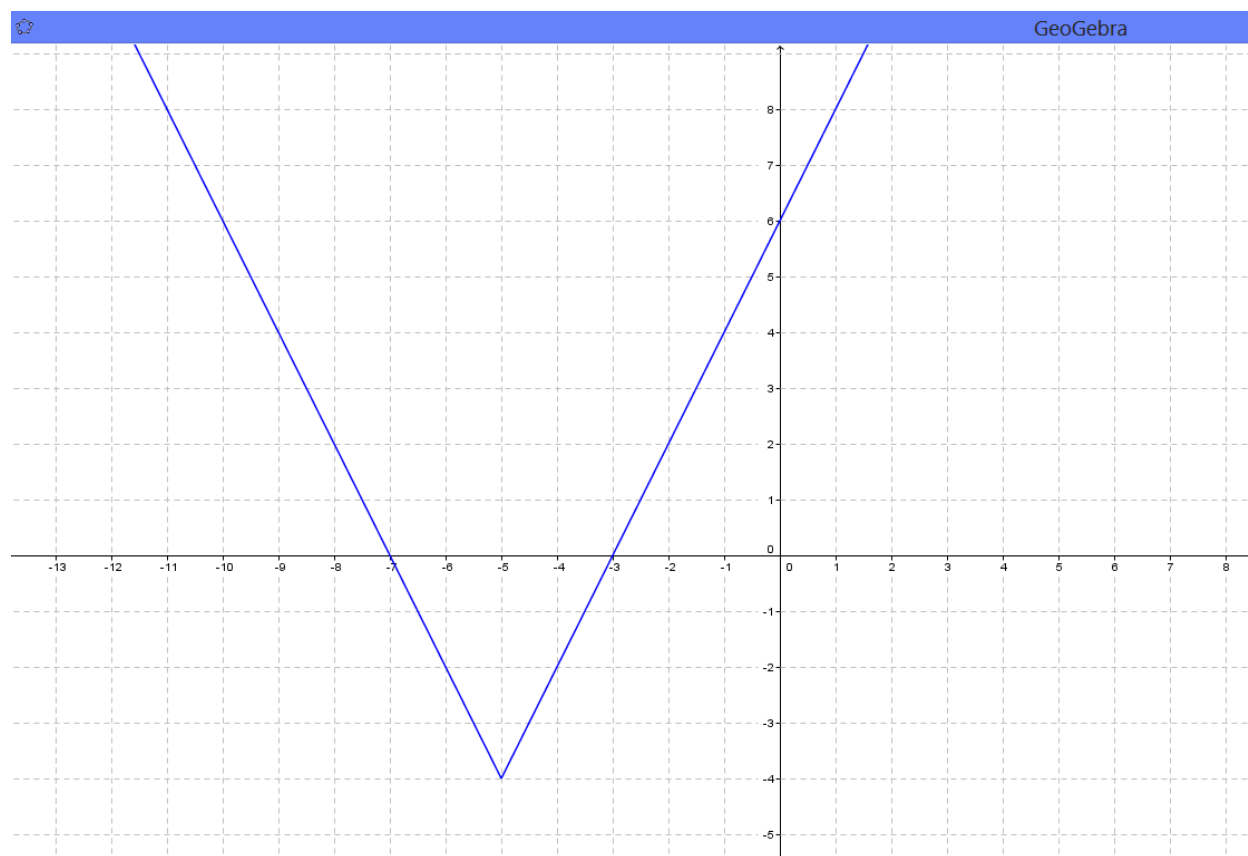
That  $f(x)$  part looks a bit scary, so I'll rewrite it as  $y = 2|x + 5| - 4$ . It looks like the basic graph of  $y = |x|$  has been shifted 5 units to the left, and 4 units down. [ $h = -5$  and  $k = -4$ ]. That would put the vertex at  $(-5, -4)$ .

The axis of symmetry has to pass through that, so it is a vertical line at  $x = -5$ .

The y-intercept always occurs where  $x$  is 0, so just substitute 0 for  $x$  to find  $y$ :  $y = 2|0 + 5| - 4 = 6$ . The y-intercept is at  $(0, 6)$ .

The x-intercept always occurs where  $y = 0$ :

$0 = 2|x + 5| - 4$ . Rearrange that to find  $x$ :  $4 = 2|x + 5|$  so  $2 = |x + 5|$ . That means either  $x + 5 = 2$  or  $x + 5 = -2$ . The x-intercepts are at  $x = -3$  and  $x = -7$ , as you can also see in the picture below. Notice that the x-coordinate of the vertex is exactly in the middle between -3 and -7.



The function  $y = x^2$  can be modified in the exact same way as  $y = |x|$ , so we can find the vertex and the axis of symmetry in the same way.

$$y = a(x - h)^2 + k$$

The vertex ends up at the point  $(h,k)$ . The axis of symmetry is the line  $x = h$ . To find the x-intercept, set  $y$  equal to 0. To find the y-intercept, let  $x = 0$ .

But wait, my function looks like this:  $y = x^2 + 6x - 16$ . Now how do I find the vertex?

If you have already learned how to complete the square (see “Factoring”) you can rewrite this equation as  $y = x^2 + 6x + \dots - 16$ . Add 9 to both sides to make a nice square:  
 $y + 9 = x^2 + 6x + 9 - 16$ . Now write it as a square:  $y + 9 = (x + 3)^2 - 16$ , which simplifies to  
 $y = (x + 3)^2 - 25$ . The vertex will be located at  $(-3, -25)$ . However, if you have practiced completing the square on more complex equations you know it can be a bit of a pain. Let’s look at an easier alternative.

For the simple parabolas  $y = x^2$  and  $y = ax^2$ , the only way that  $y$  can be zero is if  $x$  is zero. That means that these parabolas just touch the x-axis at one point, which is also the location of the vertex. If these simple parabolas are shifted left or right there will still be only one x-intercept. If you move the parabola so that it shifts down, there will be two x-intercepts. They will be located where  $y$  is zero. Because the parabola is symmetrical, the axis of symmetry will always be between those zero points. Just remember that if you are looking for the location of the vertex: if you know the x-intercepts, the x-coordinate of the vertex will be in between them.

How do you normally find the x-intercepts of  $y = x^2 + 6x - 16$ ? Well, you would set  $y$  equal to zero to get  $0 = x^2 + 6x - 16$ , which you can then solve by factoring or by using the quadratic formula. Factoring gives  $(x + 8)(x - 2) = 0$ , so  $x = -8$  or  $x = 2$ . The x-coordinate of the vertex is in between those, so take the average:  $(-8 + 2)/2 = -3$ . When  $x = -3$ ,  $y = (-3)^2 + 6(-3) - 16 = -25$ .

If you use the quadratic formula to solve  $ax^2 + bx + c = 0$ , you find that the x-intercepts are at

$\frac{-b - \sqrt{b^2 - 4ac}}{2a}$  and  $\frac{-b + \sqrt{b^2 - 4ac}}{2a}$ . In between those is the x-coordinate of the vertex,  $\frac{-b}{2a}$ .

For the function  $y = x^2 + 6x - 16$ ,  $a = 1$  and  $b = 6$ , so  $\frac{-b}{2a} = \frac{-6}{2} = -3$ . This is the x-coordinate of the vertex, and we can plug that back into  $y = x^2 + 6x - 16$  to find the y-coordinate, which is -34 in this case. Sometimes the x-coordinate is a fraction, which makes it a bit of a pain to find  $y$  from the equation.

### **Example**

Find the vertex of the graph of  $f(x) = 3x^2 - 4x + 5$ . Write the function in vertex form.

I like to rewrite that in an x-y format:  $y = 3x^2 - 4x + 5$ . The x-coordinate of the vertex is  $\frac{-b}{2a} = \frac{-(-4)}{2(3)} = \frac{4}{6}$ , or  $\frac{2}{3}$ . You can plug that into  $y = 3x^2 - 4x + 5$  to get  $y = 3\left(\frac{2}{3}\right)^2 - 4\left(\frac{2}{3}\right) + 5 = 3\left(\frac{4}{9}\right) - \frac{8}{3} + 5$ .

You will need to simplify and create common denominators:  $\frac{4}{3} - \frac{8}{3} + \frac{15}{3} = \frac{11}{3}$ .

The vertex form looks like  $y = a(x - h)^2 + k$ , and you can now plug in h and k:

$$y = a\left(x - \frac{2}{3}\right)^2 + \frac{11}{3}$$

The constant a must be 3, in order to produce the  $3x^2$  in the function.

There were a lot of fractions involved in finding k, the y-coordinate of the vertex, so you may want to take a bit of a shortcut instead. It is easy to see from the original function that the y-intercept is at (0, 5). Use that point in the vertex formula to find k:

$$y = a(x - h)^2 + k$$

$$5 = 3\left(0 - \frac{2}{3}\right)^2 + k.$$

$$5 = 3\left(\frac{4}{9}\right) + k.$$

$$5 = \frac{12}{9} + k$$

$$5 = \frac{4}{3} + k$$

$$k = 5 - \frac{4}{3}$$

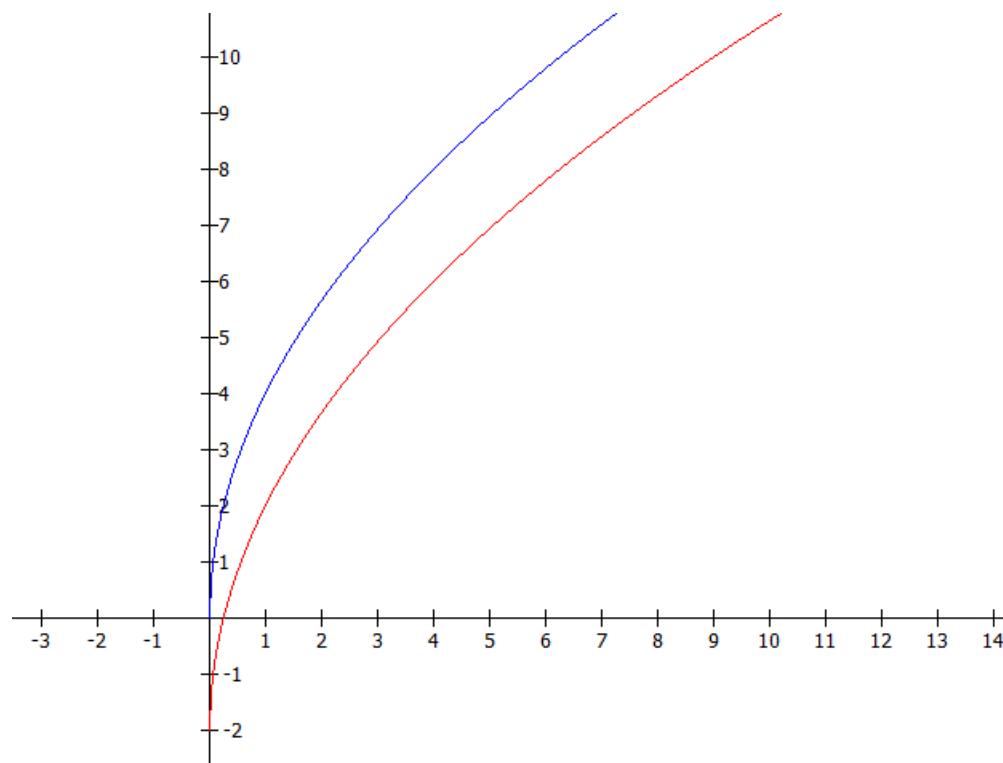
$$k = 3\frac{2}{3}, \text{ or } \frac{11}{3} \text{ just like we found earlier.}$$

$$y = 3\left(x - \frac{2}{3}\right)^2 + \frac{11}{3}.$$

## Order of Operations for Function Transformations

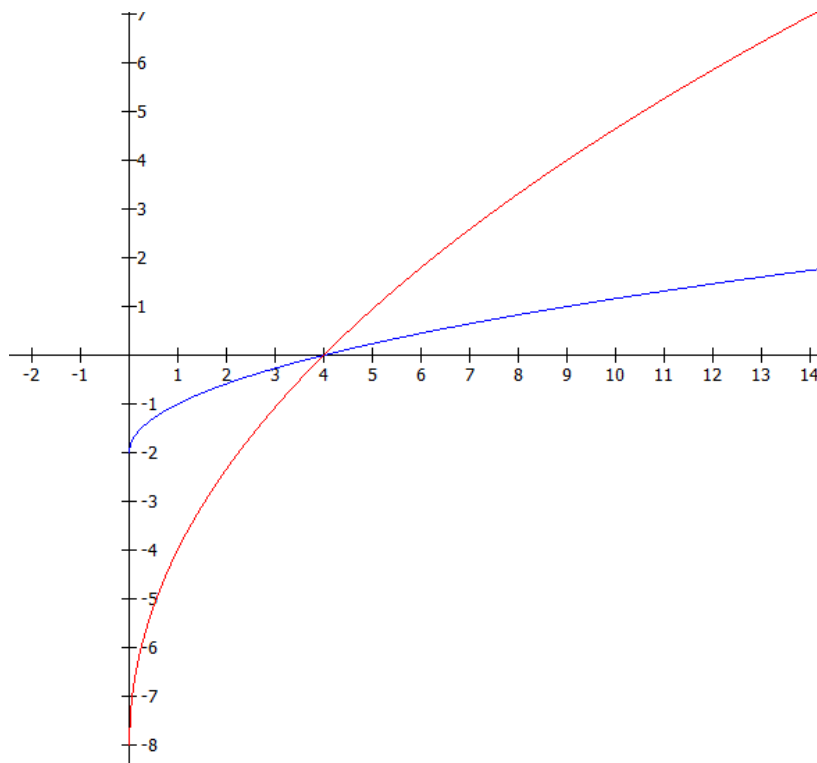
### 1. Vertical Shifts

Look at your function equation carefully so you don't make order of operation errors involving vertical shifts. **Do vertical shifts after vertical stretches or compressions.** Consider  $y = 4\sqrt{x} - 2$ . This first picture shows the graph of  $y = \sqrt{x}$  multiplied by 4 (blue curve), and then shifted down by 2 units (red curve):



This second picture shows the graph of  $y = \sqrt{x}$  shifted down by 2 units (blue curve), and then stretched by a factor of 4 through multiplication (red curve):





The correct order is the first one; multiply by 4 first and then subtract 2. That is what  $y = 4\sqrt{x} - 2$  really means. The red curve in the second picture shows  $y = 4(\sqrt{x} - 2)$  which is really  $y = 4\sqrt{x} - 8$ .

For the function  $y = -3(x - 2)^2 + 1$  the multiplication by  $-3$  is not intended to include the  $+ 1$  part, so account for this in your transformations. You can flip the graph of  $(x - 2)^2$  and then stretch by a factor of 3, or stretch it first and then flip it. You can even start with  $y = x^2$ , flip it, stretch it, and then move it to the right by 2 units. **However, the shift up must be done last.** The multiplication by  $-3$  applies only to  $(x - 2)^2$ , not to  $+ 1$ .

## **2. Reflection over the Y-Axis: $-x$**

When you replace  $x$  in any function by  $-x$ , it causes the entire graph to flip over the  $y$ -axis, which is a change that may or may not be visible. Draw the graph first, and then replace  $x$  with  $-x$ . If your problem already contains this replacement by  $-x$ , put  $x$  back in, sketch the graph, and then replace  $x$  with  $-x$  to see the reflection.

For example, if you are asked about the transformations in  $y = |-x + 4|$ , consider  $y = |x + 4|$  first. This is the graph of  $y = |x|$  moved 4 units to the *left*. If you then replace  $x$  with  $-x$ , the graph is reflected over the  $y$ -axis. That makes it look like the graph of  $y = |x|$  just moved 4 units

to the right, but you should describe this as a horizontal shift to the left, followed by a reflection.

### **3. Horizontal Stretches and Compressions**

If you are expected to do horizontal stretches and compressions you have to also be careful about your horizontal shifts. Consider  $y = 4\left(\frac{x}{5} + 2\right)^{\frac{1}{3}} - 3$ . Notice that here 2 is not added directly to  $x$ . Order of operations says that  $\frac{x}{5}$  must be done first, so 2 is being added to  $\frac{x}{5}$ . The result of that cannot be predicted from our simple formula of “ $x + 2$  shifts the graph two units to the left”. Rewrite the equation so we can see more clearly what is happening to  $x$ :

$$4\left(\frac{x}{5} + 2\right)^{\frac{1}{3}} - 3 = 4\left(\frac{x+10}{5}\right)^{\frac{1}{3}} - 3 = 4\left(\frac{1}{5}(x + 10)\right)^{\frac{1}{3}} - 3$$

Now we can see that there is really a shift to the left of 10 units.

## Polynomial Functions

An  $n^{\text{th}}$  degree polynomial has  $n$  zeros.

For synthetic division, use the number that would be the actual root of the polynomial. If you are dividing by  $(x - 2)$ , use 2 in your synthetic division.

Once you have found one zero check the polynomial that results from the division for additional zeros.

Use the Rational Zeros Theorem to find potential *rational* zeros. Divide the polynomial by its first coefficient, and then look for factors of the constant (last) term. Do not simplify the fraction.

Complex zeros and “square root” zeros, which are not rational, always come in pairs.

A polynomial with only positive terms has no negative zeros. The maximum number of positive or negative zeros of a polynomial is determined by how many sign changes occur when you insert a positive or negative value for  $x$ .

To quickly find the value of a polynomial when you choose a specific number for  $x$ , divide the polynomial by that number using synthetic division. The remainder will be the value you want.

An  $n^{\text{th}}$  degree polynomial has at most  $n-1$  turns.

When a function changes from positive to negative, and vice versa, it must cross the  $x$ -axis which results in a zero.

The graph of a polynomial crosses the  $x$ -axis where there is a zero of odd multiplicity, and just touches it where there is a zero of even multiplicity. The curve flattens where a zero has multiplicity 3 or higher.

## Graphing Polynomial Functions

### 1. Mark the zeros.

A polynomial of degree  $n$  has  $n$  roots, or zeros. For example, a cubic polynomial like  $x^3 + x^2 - 14x - 24$  is created by multiplying  $(x + 2)(x + 3)(x - 4)$ . If you want to know where the function  $f(x) = x^3 + x^2 - 14x - 24$  crosses or touches the  $x$ -axis, you need to find for which values of  $x$  the function has a value of 0. If you can factor this polynomial you can find the zeros:  $(x + 2)(x + 3)(x - 4) = 0$ , which means that  $x$  should be -2, -3, or 4. If you know the zeros you can mark them as points on the  $x$ -axis.

Although an  $n^{\text{th}}$  degree polynomial always has  $n$  roots, don't be surprised if you see a fifth degree polynomial function that has only one zero. If some of the roots are complex they will not be shown on the graph, because that would involve using imaginary numbers. Complex roots are found using the quadratic formula, and **they always come in pairs** due to the  $\pm$  sign in the formula. Those pairs always match, like  $(x - 3i)$  and  $(x + 3i)$ . It would also be impossible for a complex root to occur singly, because then the imaginary number wouldn't cancel out in the multiplication. Just try creating a polynomial by multiplying something like  $(x + 5)(x + 1)(x - 3i)$  and you'll see what I mean. So, a fifth degree polynomial could have 5 real roots, or 3 real roots and 2 complex ones, or 1 real root and 4 complex roots. There may also be repeated (duplicate) roots. The number of times a root is repeated is called its multiplicity. Even if all 5 roots of a fifth degree polynomial are real numbers, they may be repeated so that you see only one zero on the graph of the function.

### 2. Mark the y-intercept

The constant term is the  $y$ -intercept, since all the terms containing  $x$  will be zero when  $x = 0$ .

### 3. Consider the end behavior.

The first, or leading, term of a polynomial function dominates at large positive or negative values of  $x$ . The leading term has the highest degree, and that term changes in value a lot more with changing values of  $x$ . If the leading term is positive and its power is odd the value of the whole polynomial will be negative for large negative values of  $x$ , and positive for large positive values of  $x$ . We say that the "end behavior" is "down and up", since the left end of the graph points down and the right end keeps going up. That situation reverses if the leading term is *negative* and odd: the end behavior will be up and down. If the leading term is positive and has an even power, the function's value will be positive at both very large and very small values

of  $x$ ; the end behavior is “up and up”. Or, if the leading term is negative with an even power, both ends of the graph will point down.

#### 4. Decide if the graph crosses the x-axis at the zeros.

The graph crosses the x-axis where there is a zero of odd multiplicity, and just touches it where there is a zero of even multiplicity. To understand why this happens, look at values of  $x$  close to the zero value. If  $(x - 4)$  is one of the roots, the value of this term is zero at  $x = 4$ , which makes the entire polynomial have a value of zero. Just before that, say at  $x = 3.9$ , the value of the term  $(x - 4)$  will be negative, making the entire polynomial negative before another term can interfere. Just after that, at  $x = 4.1$ , the term will be positive and make the polynomial positive. This results in the graph crossing the x-axis. Now let's look at the case where the zero has an even multiplicity. If the term  $(x - 4)$  is a root that is repeated twice, we can write it as  $(x - 4)^2$ .  $(x - 4)^2$  is positive at any value of  $x$  close to 4, whether that value is slightly less than 4 or slightly greater. The function stays positive and just touches the x-axis at the zero. If the zero  $x = 4$  has a multiplicity of 3, the factored form of the polynomial contains  $(x - 4)^3$ . This expression is negative when  $x < 4$ , and positive when  $x > 4$ . The function will cross the x axis at  $x = 4$ .

#### 5. Sketch the graph

Draw the ends of your graph with an arrow, to indicate that the function will continue up or down. Guide the graph through the zeros, either crossing or touching the x-axis. When you have to turn around between zeros, you are generally not expected to know how high or low the graph reaches before it makes the turn. To get an idea, you can determine a  $y$ -value for a random  $x$  that lies between the zeros.

The curve flattens where a zero has multiplicity 3 or higher. If you want to get fancy you can show that in your graph. Compare the graphs of  $y = x^3$  and  $y = x^5$ , and also compare  $y = x^2$  with  $y = x^4$ . You can see how the curve flattens. Do some calculations with sample values of  $x$  close to 0 to see what causes this flattening.

### Number of Turns

Any graph can have one zero, since all it has to do is cross the x-axis. However, in order for there to be another zero the graph has to turn. That is, the direction of the curve has to change from going up to going down, or from going down to going up. For each additional zero there

must be another turn. If there are, for example, 5 potential zeros, then there are 4 potential turns. A 5<sup>th</sup> degree polynomial may have 5 zeros, 3 zeros, or 1 zero. There may be 4 turns, 2 turns, or no turn. Keep in mind that a turn in the graph does not necessarily result in a zero. The graph may turn, and then turn again before it reaches the x-axis. That is two turns that are “wasted” in terms of creating a zero. So, do not assume that a 5<sup>th</sup> degree polynomial with 3 zeros has only 2 turns, since it may have either 2 or 4 turns. The only thing you could say for sure is that there must be *at least* 2 turns if there are 3 zeros.

## The Intermediate Value Theorem

That sounds complicated, but it is really a big “DUH!!” If a function is continuous (it is a line with no gaps or holes), its value has to change from positive to negative, or negative to positive, in order to cross the x-axis. Suppose that we check the function value at say  $x = 1$ , and find that it is positive. Next we evaluate the function at  $x = 2$  and find that it is negative. We reason that the function line must have crossed the x axis somewhere in between  $x = 1$  and  $x = 2$ , so there is a zero between 1 and 2. Although we mostly use the Intermediate Value Theorem to look for zeros, it also applies to “intermediate” values other than zero. For example, if a continuous function has a value of 5 at the start of an interval, and a value of 10 later on, then somewhere it must have had a value of 9, at least once, because it couldn’t get from 5 to 10 without going through 9.

## Long Division and Synthetic Division

To find the zeros of a polynomial we can factor it, and division can help us find those factors. If you divide  $x^3 + x^2 - 14x - 24$  by  $x + 2$  and the remainder is zero, then  $x + 2$  must be a factor of  $x^3 + x^2 - 14x - 24$ . The result of the division will be a quadratic polynomial, which can either be factored directly or by using the quadratic formula.

First let’s look at plain long division with polynomials. That works very much like regular long division.

### Example

Divide  $x^3 + x^2 - 14x - 24$  by  $x + 2$ .

$$\begin{array}{r}
 \underline{x^2 - x - 12} \\
 x + 2 \mid x^3 + x^2 - 14x - 24 \\
 \underline{-(x^3 + 2x^2)} \\
 -x^2 - 14x
 \end{array}$$

$$\begin{array}{r}
 -(-x^2 - 2x) \\
 -12x - 24 \\
 -(\underline{12x - 24}) \\
 0
 \end{array}$$

The remainder is zero, so  $x + 2$  is a factor. The result of the division is  $x^2 - x - 12$ , which factors into  $(x + 3)(x - 4)$ . This tells us that  $x^3 + x^2 - 14x - 24 = (x + 2)(x + 3)(x - 4)$ .

### Example

Divide  $2x^5 + 3x^4 + 25x^2 - 1$  by  $x + 3$ .

Notice that the term with  $x^3$  is missing. That really means that its coefficient is zero, so write it in as  $0x^3$  to make your division easier.

$$\begin{array}{r}
 \underline{2x^4 - 3x^3 + 9x^2 - 2x + 6} \\
 x + 3 \mid 2x^5 + 3x^4 + 0x^3 + 25x^2 + 0x - 1 \\
 - (2x^5 + 6x^4) \\
 \hline
 -3x^4 + 0x^3 \\
 - (-3x^4 - 9x^3) \\
 \hline
 9x^3 + 25x^2 \\
 - (9x^3 + 27x^2) \\
 \hline
 -2x^2 + 0x \\
 - (-2x^2 - 6x) \\
 \hline
 6x - 1 \\
 - (6x + 18) \\
 \hline
 -19
 \end{array}$$

The remainder is -19.  $x + 3$  is not a factor of  $2x^5 + 3x^4 + 25x^2 - 1$ .

It is easy to make mistakes when you do long division with polynomials because you are often subtracting negative numbers. Synthetic division is a shortcut that changes the subtraction into addition, and it is more efficient because you don't have to write out all those  $x$ 's.

To divide by  $x + 3$ , we use the number -3 in our synthetic division. (To divide by  $x - 3$  we would use 3.)

Use -3 to do the synthetic division. First write out all of the coefficients of the polynomial to be divided, with their proper signs. Remember that some of the coefficients may be zero which causes the corresponding term to be invisible. **Caution: you must write the zero coefficients into your synthetic division, which is easy to forget!** The polynomial in this example has zero coefficients for the terms with  $x^3$  and  $x$ . Write it out like this:

$$\begin{array}{r|rrrrrr} -3 & 2 & 3 & 0 & 25 & 0 & -1 \end{array}$$


---

Now we are ready to start the division. Take the first coefficient, 2 in this case, and put it below the line. Next, multiply 2 times -3, and place the result of that below the next coefficient. Add 3 and -6, and place the result below the line. This gives you the next multiplication: -3 times -3. Put the result below the next coefficient in line, and add again. Use the result for the next multiplication, and so on, like this:

$$\begin{array}{r|rrrrrr} -3 & 2 & 3 & 0 & 25 & 0 & -1 \\ & & -6 & 9 & -27 & 6 & -18 \\ \hline & 2 & -3 & 9 & -2 & 6 & -19 \end{array}$$

The very last number below the line is the remainder. The other numbers represent the coefficients of the polynomial that is the result of the division. Its degree will always be one less than the degree of the polynomial you started with because you have divided by a simple polynomial of degree 1.

The result of the above division is:

$$2x^5 + 3x^4 + 25x^2 - 1 \text{ divided by } x + 3 = 2x^4 - 3x^3 + 9x^2 - 2x + 6 \text{ plus } -19/(x + 3).$$

Check your answer by multiplying the result of your synthetic division by  $x + 3$ , and add the remainder -19. You should get the original polynomial back.

If the remainder is zero, you know that you have found a factor of the original polynomial. Look at the result of your division and factor the new polynomial if possible.

Unfortunately synthetic division only works when you are dividing by a simple expression like  $x + 3$ . If you want to divide your polynomial by something with a higher power of  $x$ , like  $x^2 + 1$ , you must use long division.

It is possible to use synthetic division to divide by any linear expression such as  $3x - 5$ . Here the value for  $x$  that would make  $3x - 5$  equal to zero is  $\frac{5}{3}$ , so we want to use that fraction. When we



do, we are effectively dividing whatever polynomial we are using by  $x - \frac{5}{3}$  instead of  $3x - 5$ .

$3x - 5$  is  $3(x - \frac{5}{3})$ , so  $x - \frac{5}{3}$  is three times smaller than what we actually should divide by. **To compensate for that, you must first divide your polynomial by 3**, and then you can use  $\frac{5}{3}$  in the synthetic division.

## Finding the Value of $P(a)$ Using Synthetic Division

Polynomial functions are sometimes indicated by  $P(x)$  rather than  $f(x)$  to emphasize that the function is a polynomial rather than another type of function. If you have a polynomial function like  $P(x) = x^5 - 12x^4 + x^3 + 2x^2 - x + 15$ , it can be a bit of a pain to find the value of the function for a specific value of  $x$ . You can do this quickly and easily through synthetic division. The principle is this:

If you divide 14 by 3, the answer is 4 with a remainder of 2. That means that 14 equals 3 times 4, plus 2. In the same way, if you divide a polynomial by something like  $(x - 5)$ , you get an answer and a remainder (which may be 0). This means that your polynomial can be written as  $(x - 5)$  times *answer* + remainder. If you then take  $x$  to be 5, the polynomial turns into 0 times *answer* + remainder. So,  $P(5) = \text{remainder}$ .

You may wonder how this is useful, since you could just plug  $x = 5$  into the polynomial to find the value. Well, sometimes that is not so easy. Consider  $P(x) = x^5 - 12x^4 + x^3 + 2x^2 - x + 15$ . Even if you have a calculator it takes time to substitute  $x = 5$ , and you could make mistakes. If you don't have a calculator things could get difficult. With synthetic division the calculations are simpler:

$$\begin{array}{r|rrrrrr}
 5 & 1 & -12 & 1 & 2 & -1 & 15 \\
 & & 5 & -35 & -170 & -840 & -4205 \\
 \hline
 & 1 & -7 & -34 & -168 & -841 & -4290
 \end{array}$$

Your answer is  $x^4 - 7x^3 - 34x^2 - 168x - 841$ , with a remainder of -4190. Therefore,  $x^5 - 12x^4 + x^3 + 2x^2 - x + 15 = (x - 5)(x^4 - 7x^3 - 34x^2 - 168x - 841) - 4190$ .

When  $x = 5$ , this reads:

$$(5 - 5)(x^4 - 7x^3 - 34x^2 - 168x - 841) - 4190 = 0 - 4190.$$

So now we know that  $P(5)$  is -4190, which is the remainder you get when you divide  $P(x)$  by  $(x - 5)$ .

In general, when you divide  $P(x)$  by  $(x - a)$ , you get some other polynomial  $Q(x)$ , plus a remainder  $R$ .  $P(x) = (x - a) \cdot Q(x) + R$ .

Then set  $x = a$  so we can find  $P(a)$ :  $P(a) = (a - a) \cdot Q(a) + R = 0 + R$ . This means that  $P(a) = R$ .

For any polynomial  $P$ ,  $P(a)$  is equal to the remainder that is left when you divide  $P$  by  $(x - a)$ . If  $a$  happens to be a zero of the polynomial,  $P(a)$  is equal to zero, and the remainder is zero as we would expect.

## Finding the Zeros

The zeros of a polynomial are those values of  $x$  that make the polynomial have a value of 0. If you can factor the polynomial it is easy to see the zeros:

$$x^3 + x^2 - 8x - 12 = 0$$

$$(x + 2)(x + 2)(x - 3) = 0$$

The zeros are -2, -2 and 3.

If you consider complex factors, all polynomials can be factored completely. This means that an  $n$ th degree polynomial has  $n$  zeros. The easiest way to find the zeros of a polynomial is to use a graphing calculator or graphing software. This shows the *real* zeros. You can then divide by those zeros to find any complex zeros that you can't see on the graph. Be careful: if the function just touches the  $x$ -axis at one spot there is a definitely a repeated zero present. You can divide a polynomial of degree 3 or higher by this zero at least 2 times to get a remainder polynomial that may or may not have complex zeros. For example, graph the function  $y = x^3 + x^2 - 8x - 12$ . You can see the repeated zero at  $x = -2$  on the graph. Divide by  $(x + 2)$  twice [or by  $(x + 2)^2$  once] to see that the polynomial factors as  $(x + 2)(x + 2)(x - 3)$ . To see why the function just touches the  $x$ -axis at the repeated zero and then heads back down, first substitute  $x = -1.9$ , and then  $x = -2.1$ . The function has a negative value on both sides of  $x = -2$

because  $(x + 2)^2$  is always positive. It is the other factor,  $x - 3$ , that determines when the function will be positive or negative.

Calculators and computers are nice to have, but they don't work by magic. Someone first has to think of ways to do something, and then write a program that a machine can execute. We can find zeros of a polynomial manually although it can get complicated and tedious. Here we will just consider some of the basic principles of how this can be done.

Suppose a polynomial has only integer roots, for example:  $a$ ,  $b$  and  $c$ . We can write this polynomial as  $(x - a)(x - b)(x - c)$ . [The value of this polynomial will be zero when  $x = a$ , or  $x = b$  or  $x = c$ .] When we multiply this out, the first term will be  $x^3$  and the last term will be  $-abc$ . The roots of the polynomial will always be factors of the constant term (the last term of the polynomial, which may be zero). By looking at all the possible factors of the constant term, we can guess at the roots. If we manage to find one root, say  $b$ , we can divide the polynomial by  $(x - b)$  to make it easier to find the other roots.

When all the roots are integers, the coefficient of the first term of the polynomial is always 1. If there is another coefficient, we can divide by it when we look for the zeros. For example:

$$3x^3 - 7x^2 - 14x + 24 = 0 \quad \text{can be changed to } 3\left(x^3 - \frac{7}{3}x^2 - \frac{14}{3}x + \frac{24}{3}\right) = 0.$$

Now we can see that at least some of the roots would be fractions. All of the roots of the part in parentheses will multiply out to  $\frac{24}{3}$ . It is very tempting, but **do not simplify  $\frac{24}{3}$  to 8, because that would cause a loss of information and hide some of the roots!** The top part of each fraction must be a factor of 24 and the bottom part a factor of 3. This is a specific example of the **Rational Zeros Theorem**. Any of the numbers suggested as possible roots by the Rational Zeros Theorem can be either positive or negative. Note that the bottom part can always be 1, so integer roots are always a possibility. The actual roots of this particular polynomial are -2, 3, and  $\frac{4}{3}$ , and the complete factorization is

$$3(x + 2)(x - 3)\left(x - \frac{4}{3}\right) \quad \text{which can also be written as } (x + 2)(x - 3)(3x - 4)$$

Please note that the Rational Zeros Theorem works for polynomials that have rational zeros (zeros that can be written as a ratio, which includes integers). When you use the quadratic formula to find zeros, you'll often end up with square roots that cannot be converted to a rational number. **Even though the graph of a polynomial shows real zeros where the function crosses or touches the x-axis these may not be rational zeros, and you may search for them in vain by using the Rational Zeros Theorem.**

Synthetic division gives us a good way to test all the possible rational zeros. If we think that  $z$  is a possible zero of the polynomial, we want to divide it by  $x - z$  to see if the remainder is 0. Use  $z$  in the synthetic division. Usually it is easiest to start with the positive integers, but if all the terms of your polynomial are positive it will not have a positive zero. For example, if you look carefully at  $5x^3 + 3x^2 + 8x + 17$ , you will see that no positive value for  $x$  will make this polynomial zero. You may use  $z$  in the synthetic division even if it is a fraction, and it will give a remainder of zero if  $z$  is in fact a root. However, the result of your division is not actually accurate when you use a fraction (see synthetic division). That doesn't really matter when you only need a root.

Once you have been doing synthetic division for a while, you may notice that if you take a polynomial with a **positive leading term** and divide it by larger and larger positive numbers, the resulting numbers below the line eventually all become positive. Once that happens, there is no way that these numbers can become negative or zero again as we keep dividing by a larger and larger positive number. Try it out for yourself. Because of this, once the numbers below the line are all positive or zero, we know that there are no zeros larger than the number we are dividing by. We call this number the *upper bound*, and don't waste time looking for larger positive zeros. If your polynomial happens to have a negative leading term, just divide it by  $-1$ ; it will still conveniently have the same zeros. A somewhat similar thing happens when we divide our polynomial by smaller and smaller negative numbers. The numbers below the line eventually become alternating positive or negative numbers (consider 0 positive or negative as needed). When you see that happening you should know that there is no point in searching for smaller zeros. This is the *lower bound*. To take advantage of the upper and lower bounds when looking for zeros, start your search somewhere in the middle of the range of potential positive or potential negative zeros.

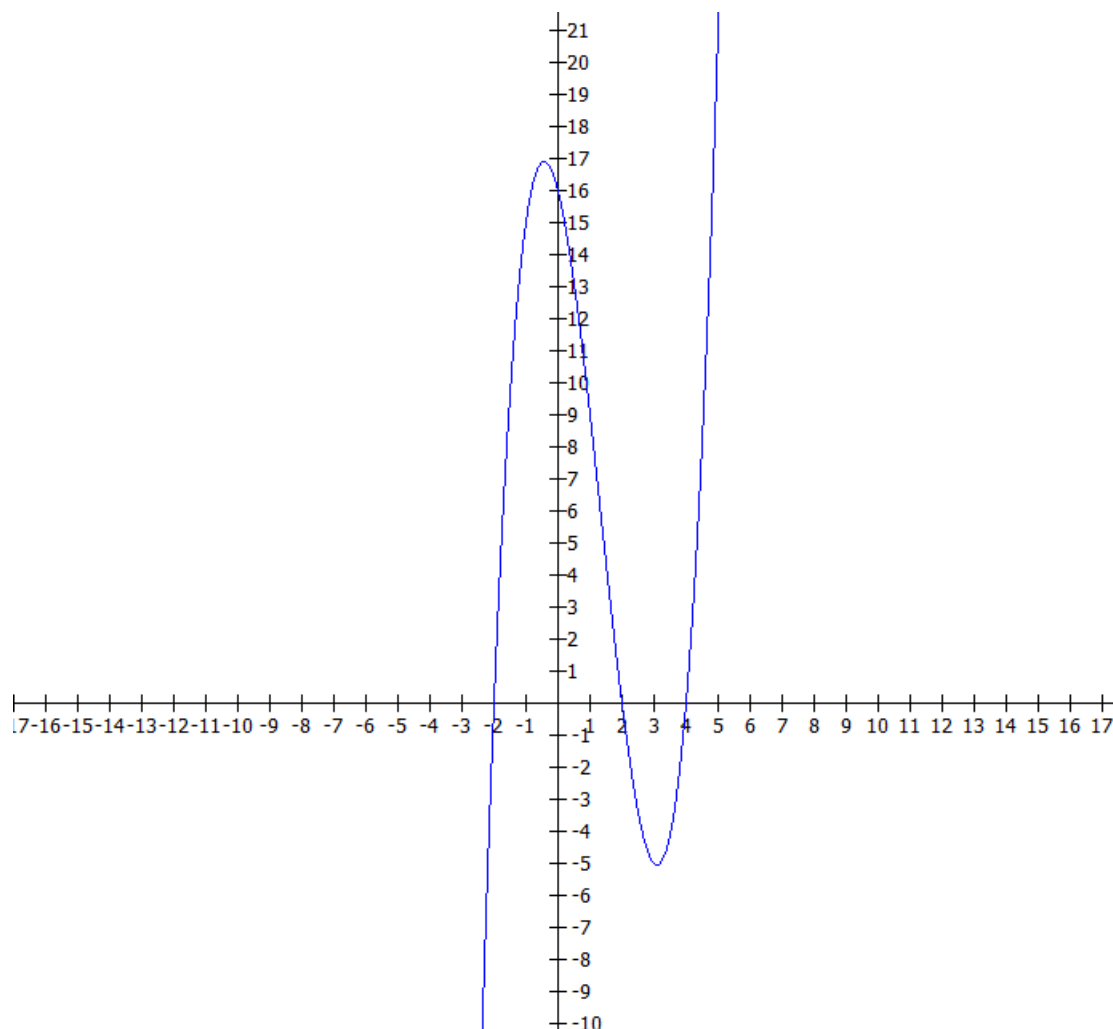
When you find one zero, look at the result of your division. This new polynomial should also have some zeros, so start looking for them. The new polynomial will have fewer potential rational zeros, and none of them will be outside the upper and lower bounds you determined for the original polynomial. This may allow you to eliminate some possibilities right away. Also, this polynomial will have its own upper and lower bounds that can guide you. If all the terms of the new polynomial are positive, there are no sign changes (so no positive zeros), and you need to start looking for negative zeros. When you find another zero, there is again a new polynomial that results from the division. Eventually you are down to a quadratic, or second-degree polynomial, and you can factor it or use the quadratic formula to find the remaining two zeros. Thanks to the quadratic formula, complex zeros and zeros with a square root that can't be simplified always come in pairs.

Use a graphing calculator or graphing software to check that you have found the correct number of real zeros. Make sure you look at the degree of your polynomial and account for all the zeros that it should have. An  $n$ -th degree polynomial has  $n$  zeros, some of which may be complex zeros that always come in pairs.

## Descartes' Rule of Signs

Because all of the possible zeros of a polynomial can be either positive or negative, it sometimes helps to know how many potential positive and negative zeros there are. In particular, if there is no possibility that there are any positive zeros at all you don't need to waste your time checking for them. Descartes' Rule of Signs can help you with that.

Poor Mr. Descartes did not own a graphing calculator or a computer because these items did not exist yet. When he needed to graph a function he first had to invent the Cartesian coordinate system, and then draw the function by hand, most likely using the methods we already looked at. From his perspective, the rule of signs probably seemed natural.



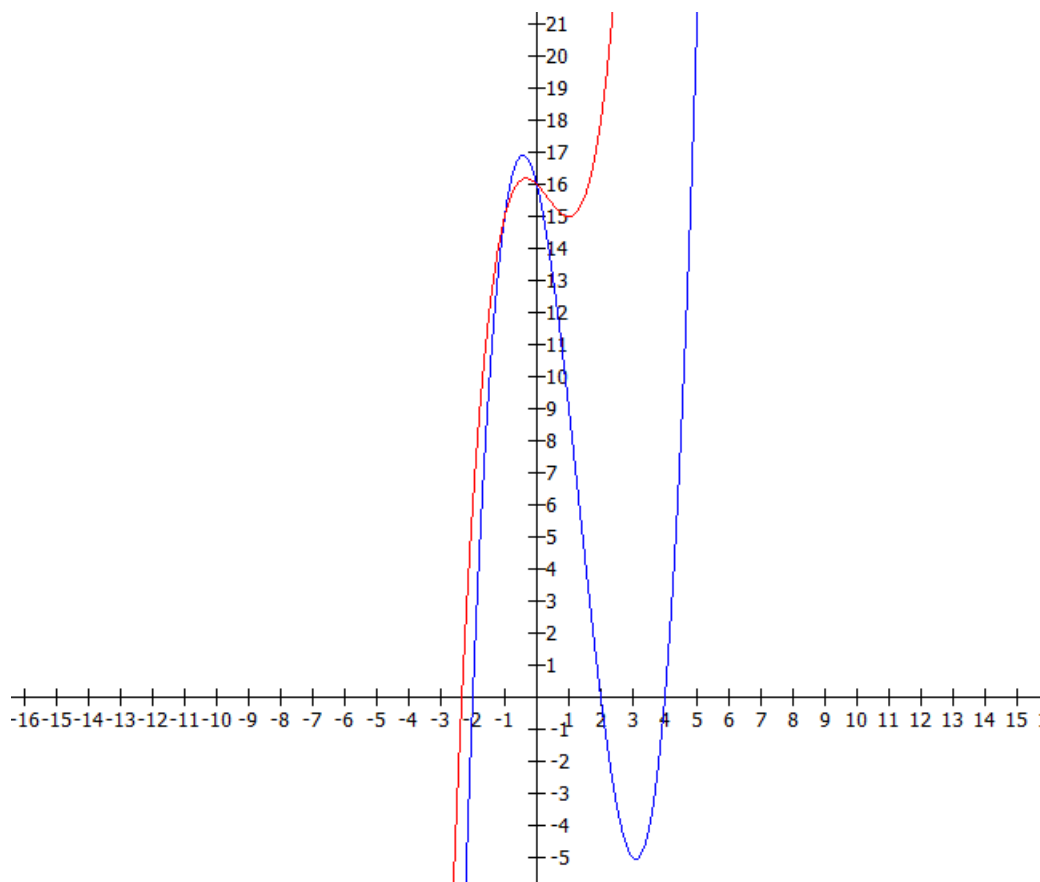
Let's look at a polynomial function with three roots. We will know that it has three roots because we are going to construct the polynomial ourselves by multiplying  $(x + 2)(x - 2)(x - 4)$ . You can see that there will be one negative root, at  $x = -2$ , and two positive roots. We can quickly multiply  $(x + 2)(x - 2)$  to get  $x^2 - 4$ . Then we multiply that by  $(x - 4)$  to get  $x^3 - 4x^2 - 4x + 16$ . The picture above is the graph of this polynomial, and it shows the expected roots.

The behavior of this function would not be a total surprise to you if you were used to calculating the values by hand. Let's do that now. When  $x = 0$ , all terms are 0 except the last. The value is 16, a relatively large positive number. When  $x = 1$ , we get  $1 - 4 - 4 + 16$ . The value is still positive, but the two negative terms  $-4x^2$  and  $-4x$  are starting to pull it in a negative direction. By the time we get to  $x = 2$  we have  $8 - 16 - 8 + 16$  and the function is at zero. Once  $x$  equals 3, the  $-4x^2$  term is very prominent, pulling the function to its maximum negative value:  $27 - 36 - 12 + 16$ . Past this point, the first term starts to play a bigger role. Because it takes  $x$

to the third power this term grows very rapidly with increasing values of  $x$ . It soon brings the function to zero ( $64 - 64 - 16 + 16$ ), and now there is no turning back.  $x = 5$  gives  $125 - 100 - 20 + 16 = 21$ , and by the time we are at  $x = 10$  the first term really starts to leave the others behind:  $1000 - 400 - 40 + 16$ . What determines the zeros of this function is the change in direction that occurs as first the end positive term, then the negative terms, and then the lead positive term become prominent in setting the value of the function. In fact, **the maximum number of positive zeros is determined by how many “sign changes” occur in the function when you substitute a positive number for  $x$ .** “Sign changes” are changes in the signs of the terms as you read from left to right:  $+x^3 - 4x^2 - 4x + 16$ . Here the changes have been marked in red to show where they occur. There are two sign changes, so the maximum number of positive zeros is 2.

Now let's look at what happens at negative values for  $x$ . At  $x = -1$ , we get  $(-1)^3 - 4(-1)^2 - 4(-1) + 16$ , which is  $-1 - 4 + 4 + 16$ . Here the function is still positive. Once  $x = -2$ , we have  $-8 - 16 + 8 + 16$ . The first two terms end up negative and they quickly start pulling the function down to zero and beyond. Larger negative values of  $x$  will do nothing to turn that around, and the function's value remains negative. For this function, negative numbers “see” only one sign change. The function is positive at zero, and then crosses the  $x$ -axis as it gets more negative. To see that there is one sign change for negative numbers, substitute  $-x$  for  $x$  in the function:  $(-x)^3 - 4(-x)^2 - 4(-x) + 16 = -x^3 - 4x^2 + 4x + 16$ . **The maximum number of negative zeros is determined by how many “sign changes” occur in the function when you substitute a negative number for  $x$ .**

Is there always a zero for every sign change in the function? No, not really. In the function we just looked at the  $x^2$  and  $x$  terms had a big advantage because they each had a coefficient of 4, giving them more influence at low values of  $x$ . When that coefficient is gone they hardly make a dent against the large positive term 16, even though they are working together to pull the function down. Here the red line shows the function  $x^3 - x^2 - x + 16$ , and you can see that there is only one zero:



There is still a “dip” in the function on the positive side, but the loop does not reach all the way down past the x-axis. That effectively removes two of the potential zeros. If the loop is long enough that it just touches the x-axis it will result in a repeated zero at that point. A prediction of a single zero on the positive or negative side will always work for a cubic polynomial function because that doesn’t involve a loop. The function simply crosses the x-axis once, which it has to do because of its predictable end behavior.

So, we can say that we can find the **maximum** number of positive or negative zeros by looking at sign changes. If there are actually fewer zeros, that is caused by one or more loops of the curve failing to reach the x-axis, resulting in 2, 4, 6, or some even number *fewer* zeros than expected. Don’t forget either that one or more of the zeros of a polynomial could actually be 0, which is neither positive nor negative.

When you do not get as many zeroes as you would expect, the missing zeroes are actually complex numbers generated by the negative root in the quadratic formula. These complex solutions always come in pairs because there is both a positive and a negative complex root for a negative number.



## Rational Expressions and Functions

Rational expressions can be simplified by finding the least common denominator of all the fractions.

The value of  $x$  that would make a factor of the denominator zero must be excluded from the domain of a rational function. There is a vertical asymptote here, unless the same factor is also present in the numerator in which case there is an infinitely tiny “hole” in the graph.

For rational equations, multiply both sides by the least common denominator to get rid of the fractions. Remember to make sure that your final solution does not give a zero value for the denominator of any of the fractions.

A fraction is equal to zero when its denominator is equal to zero.

### Horizontal Asymptotes (limits at infinity)

1. If the power of the leading term (leading term = term with the highest power) on the top is smaller than the power of the leading term on the bottom, the rational function will always approach 0. The  $x$ -axis will be a horizontal asymptote.
2. If the leading term on the top has the highest power, the function will approach positive or negative infinity.
3. If the powers of the top and bottom leading terms are equal, the function approaches a definite number which is determined by the coefficients of the leading terms. There will be a horizontal asymptote.

If the power of the top leading term is one higher than the power of the bottom leading term, the function will have a *slant asymptote*.

For rational inequalities always start by getting a zero on one side. Next, put everything on the other side over a common denominator. There are equality points where the numerator is zero. There are also potential changeover points where the denominator is zero. Use test values to see where the inequality holds.

Check your work by using a graphing calculator or graphing software.

Rational expressions are those that can be written as a ratio of two polynomials. In practical terms, this means that we'll be dealing with a fraction that has  $x$  in the denominator, and usually in the numerator too.

## Adding, Subtracting, and Simplifying Rational Expressions

The first thing that you are likely to learn to do with rational expressions is to simplify them. It is very much easier to see what is going on when an expression is composed of a single fraction.

For example, you may be asked to change  $\frac{x+3}{x(x-2)} - \frac{x-3}{x(x+2)}$  to a single fraction. That looks like

a lot of work, so let's solve an easier problem:  $\frac{1}{10} + \frac{1}{15}$ . As you know from basic arithmetic, we need a common denominator to be able to add these fractions [the fractions must both be the same size]. To do that we must find a number that both 10 and 15 "go into". We could just multiply 10 and 15 to get 150 so we can be sure that it will work out, but that's overkill in this case. The number 30 is much more suitable since it is the smallest number that is divisible by both 10 and 15. It is the LCD, which stands for least common denominator [or liquid crystal display, depending on what you are reading]. How can we find the LCD? Well, let's look at what makes up the numbers 10, 15, and 30:

$$10 = 2 \cdot 5$$

$$15 = 3 \cdot 5$$

$$30 = 2 \cdot 3 \cdot 5$$

You need to stop here and make sure you understand how the number 30 "contains" both the number 10 and the number 15. Create your own numerical examples if you need to.

Now let's go back to  $\frac{x+3}{x(x-2)} - \frac{x-3}{x(x+2)}$ .

What is the smallest expression that contains  $x(x-2)$  and  $x(x+2)$ ? The factors we need are  $x$ ,  $(x-2)$ , and  $(x+2)$ . It is not necessary to use two copies of  $x$ . The expression we want for the LCD is  $x(x-2)(x+2)$ . This is divisible by both  $x(x-2)$  and  $x(x+2)$ . You may think you should multiply out this expression right away, but it is actually to your advantage to leave the common denominator in its factored form. That way you can quickly check for any factors that cancel out on the top and bottom at the end. To get the first fraction to have a denominator

equal to the LCD, we must multiply the bottom by  $(x + 2)$ , so we have to do that to the top also to avoid changing the fraction:

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

The second fraction needs to be multiplied by  $(x - 2)$  on the top and bottom:

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

Now the denominators are the same. Because we are subtracting the numerators, we have to multiply out the expressions on the top of each fraction:

$$\frac{(x+3)(x+2)}{x(x-2)(x+2)} - \frac{(x-3)(x-2)}{x(x+2)(x-2)} = \frac{x^2+5x+6}{x(x-2)(x+2)} - \frac{x^2-5x+6}{x(x-2)(x+2)} = \frac{x^2+5x+6-(x^2-5x+6)}{x(x-2)(x+2)} = \frac{10x}{x(x-2)(x+2)}$$

Since we left the denominator in factored form, we can easily see that  $x$  can be canceled out. We can divide the top and bottom by  $x$ , provided that  $x$  is not zero. We can never divide by zero, but  $x$  could not have been zero to start with since it appears in the denominator of one of the original fractions.

$$\frac{10}{(x-2)(x+2)} \quad x \neq 0, x \neq 2, x \neq -2 \quad \text{This is the same as} \quad \frac{10}{x^2-4} \quad x \neq 0, x \neq 2, x \neq -2$$

Notice that we have to specify that  $x$  cannot be zero in the answer because you can no longer see it by looking at the rational expression. Also,  $x$  cannot be 2 or -2.

Sometimes you need to factor one or more of the denominators so that you can find the least common denominator. For example:

$$\frac{2x+8}{x^2+9x+20} + \frac{x}{(x+5)} =$$

Factor the quadratic to see that the LCD is  $(x + 4)(x + 5)$ :

$$\frac{2x+8}{(x+5)(x+4)} + \frac{x}{(x+5)} = \frac{2x+8}{(x+4)(x+5)} + \frac{x(x+4)}{(x+5)(x+4)} = \frac{2x+8+x^2+4x}{(x+5)(x+4)} = \frac{x^2+6x+8}{(x+5)(x+4)}$$

Now the top part can be factored:

$$\frac{x^2+6x+8}{(x+5)(x+4)} = \frac{(x+4)(x+2)}{(x+5)(x+4)} = \frac{(x+2)}{(x+5)} \quad x \neq -4, x \neq -5$$

We may divide the top and bottom by  $(x + 4)$  to cancel it out, provided that  $(x + 4)$  is not zero. Notice that a value of  $-4$  for  $x$  would not be allowed in the original expression anyway. You can use any value for  $x$  other than  $-4$  and  $-5$  to check your work. Try using  $x = 10$  to check that

$$\frac{2x+8}{x^2+9x+20} + \frac{x}{(x+5)} = \frac{(x+2)}{(x+5)}$$


---

If your rational expression appears as a single fraction that has other fractions contained within it, there is a special trick you can use. Because it is already one fraction, you now have the option of multiplying both the top and the bottom by the same thing. If we pick that thing strategically, we can actually get rid of all of the fractions on the top and bottom in a single operation. The trick is to multiply both the top and bottom of your fraction by the LCD of all the fractions you want to eliminate. Take a look at an example:

$$\frac{\frac{1}{x} + \frac{3}{y}}{\frac{1}{x} + 1}$$

Here the LCD of the little fractions is  $xy$ . We can multiply the top and the bottom of the main fraction by  $xy$  to get rid of all the little ones:

$$\frac{\frac{1}{x} + \frac{3}{y}}{\frac{1}{x} + 1} \cdot \frac{xy}{xy} = \frac{\frac{xy}{x} + \frac{3xy}{y}}{\frac{xy}{x} + xy} = \frac{y+3x}{y+xy} = \frac{y+3x}{y(1+x)}$$

This is much faster than eliminating these little fractions individually, but we can still do so if we want or if we don't remember the trick to use.

Let's do it the other way too. First eliminate the fractions on the top:

$$\frac{1}{x} + \frac{3}{y} = \frac{y}{xy} + \frac{3x}{xy} = \frac{y+3x}{xy}$$

Then do the bottom one, where the LCD is  $1 \cdot x$ :

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

Now we have:

$$\frac{\frac{y+3x}{xy}}{\frac{1+x}{x}}$$

To divide by a fraction you multiply by its reciprocal:

$$\frac{y+3x}{xy} \div \frac{1+x}{x} = \frac{y+3x}{xy} \cdot \frac{x}{1+x} = \frac{x(y+3x)}{xy(1+x)} = \frac{y+3x}{y(1+x)}$$

You can see that this other way takes longer, so if you have a lot of these problems to do you will want to remember the trick.

## Rational Equations

When you have an equation, you can multiply both sides by the LCD to quickly clear all those annoying fractions. Factor if necessary to find the LCD:

$$\frac{5}{x+2} + \frac{3}{x-5} = \frac{13}{x^2-3x-10}$$

$$\frac{5}{x+2} + \frac{3}{x-5} = \frac{13}{(x+2)(x-5)}$$

The LCD is  $(x+2)(x-5)$  so we multiply both sides by it:

$$(x+2)(x-5) \cdot \left( \frac{5}{x+2} + \frac{3}{x-5} \right) = \left( \frac{13}{(x+2)(x-5)} \right) \cdot (x+2)(x-5)$$

$$5(x-5) + 3(x+2) = 13, \quad x \neq -2 \text{ and } x \neq 5$$

That is a nice trick, but again, you may not remember to use it. Let's do it another way too:

Multiply each fraction on both sides of the equation by whatever it takes to get that least common denominator. Don't start multiplying everything out completely, since the goal is to eliminate factors:

$$\frac{5(x-5)}{(x+2)(x-5)} + \frac{3(x+2)}{(x+2)(x-5)} = \frac{13}{(x+2)(x-5)}$$

$$\frac{5(x-5)+3(x+2)}{(x+2)(x-5)} = \frac{13}{(x+2)(x-5)}$$

Notice that the denominators are now equal, so the numerators must be too:

$$5(x-5) + 3(x+2) = 13.$$

Now we can just multiply things out and solve:

$$5x - 25 + 3x + 6 = 13$$

$$8x - 19 = 13$$

$$8x = 32, \text{ so } x = 4$$

Just remember to check that your final solution does not give a zero value for the denominator of any of the fractions! Sometimes that may mean that there is no solution to the equation at all.

In general, finding a common denominator so that you only have to worry about the numerator works well for solving rational equations. Alternatively, it is also possible to create one fraction on each side of the equals sign so that you can cross-multiply.

### Example

$$\frac{4}{x-1} + \frac{14}{x+2} = \frac{15}{x}$$

$$\frac{4(x+2)}{(x-1)(x+2)} + \frac{14(x-1)}{(x+2)(x-1)} = \frac{15}{x}$$

$$\frac{4x+8+14x-14}{(x-1)(x+2)} = \frac{15}{x}$$

$$\frac{18x-6}{(x-1)(x+2)} = \frac{15}{x} \quad \text{Now cross-multiply:}$$

$$x(18x - 6) = 15(x - 1)(x + 2)$$

$$18x^2 - 6x = 15x^2 + 15x - 30$$

$$3x^2 - 21x + 30 = 0$$

$$x^2 - 7x + 10 = 0$$

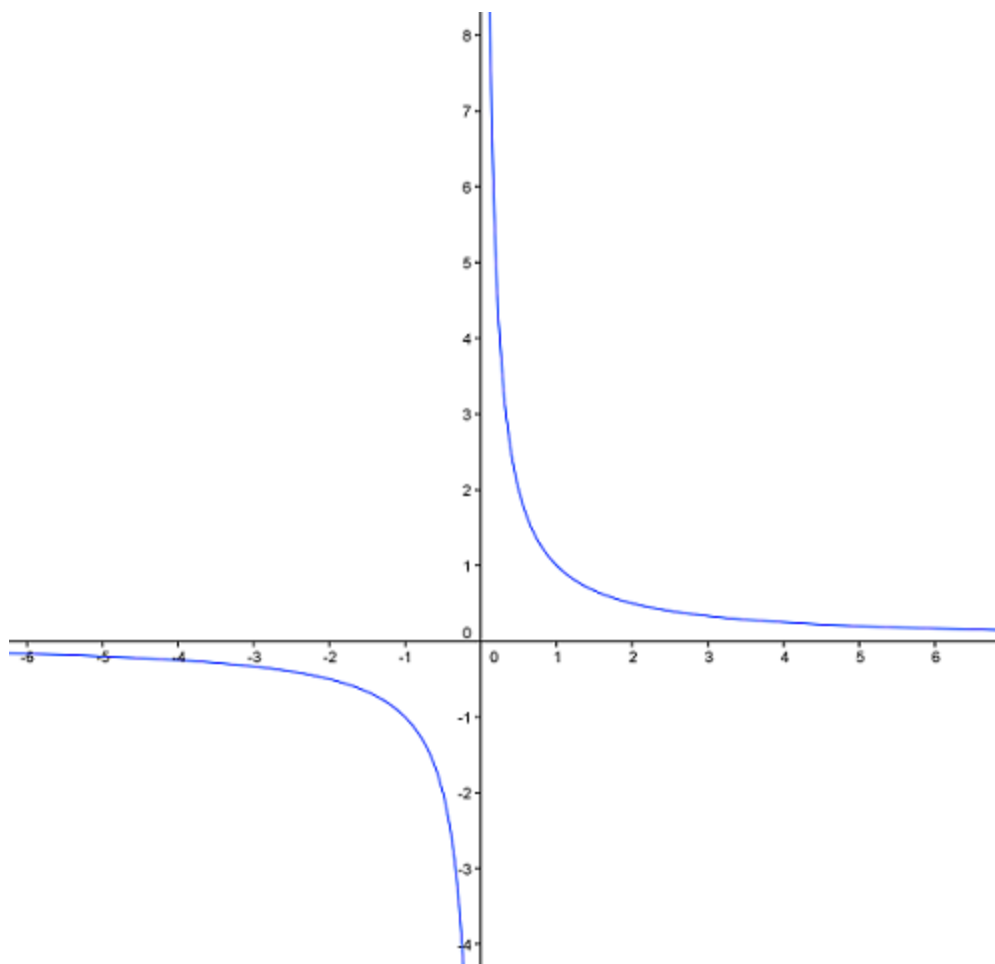
$$(x - 2)(x - 5) = 0$$

$$x = 2 \text{ or } x = 5 \quad \text{Both of these solutions work in the original equation.}$$

Just a word of caution: If you need to divide both sides of an equation by an expression containing  $x$ , make sure that it is not zero. For example, if you want to divide both sides of an equation by  $x + 4$ , immediately write  $x \neq -4$ , before you even do the division.

## Vertical Asymptotes

The most important thing to remember about rational functions is that we must never, ever divide by zero. Therefore we have to very carefully exclude any values of  $x$  that would make the denominator zero. If you look at the graph of the function  $y = \frac{1}{x}$ , you can see that this smart function already knows that since it very carefully avoids the value  $x = 0$ :



For this function the line  $x = 0$  is a *vertical asymptote*, which is a line that the function gets close to but never reaches. How does this clever function magically know to avoid  $x = 0$ ? Think about it before you continue reading.

---

Thanks to modern technology, vertical asymptotes can seem quite mysterious. You're just getting used to the idea of rational functions when all of a sudden these strange vertical lines start to appear. Early pioneers of rational functions noticed these lines and named them, but they were not very surprised by them. If you calculate function values by hand you can see exactly what causes vertical asymptotes.

Let's take a closer look at the function  $f(x) = \frac{1}{x}$ . We can easily see that inserting  $x = 0$  here is a big no-no, because it would cause a division by 0. This function has no output value for  $x = 0$ . What is more interesting is what happens on either side of 0. We can't use 0 for  $x$ , but we can use a value that is really close to 0, such as 0.001. Inserting this for  $x$ , we get  $\frac{1}{.001}$ , which is 1000. Dividing by such a small number causes the function to output a large positive value. The closer we get to 0 the larger this value gets.  $\frac{1}{.00001} = 100,000$ . Eventually it approaches positive infinity. On the other side of 0 we can pick a small negative number like -.001 to insert for  $x$ . Now the output value is -1000. The closer we get to 0 the more negative the function value, so it goes to negative infinity. On either side of the non-permissible value  $x = 0$ , the function approaches negative or positive infinity, creating a vertical asymptote.

Is there always a vertical asymptote where there would be a zero in the denominator? No, there is a reason why there may not be one. Consider the function  $f(x) = \frac{x^2-1}{x-1}$ . If you factor the top part, you see that it is composed of  $(x-1)(x+1)$ . The function still does not exist at  $x = 1$ , but now there is a difference in what happens on either side of 1. At  $x = 0.999$  we get -0.001 on the bottom, but the top is  $(-0.001)(0.999 + 1)$ . The division by the small negative number cancels out, and the function returns a value of 1.999. On the other side of 1, at  $x = 1.001$ , the same thing happens. The division by 0.001 cancels to give a function value of 2.001. As we get closer and closer to 1 on either side, the function value gets closer and closer to 2, even though it never actually gets there. There is no asymptote; there is just a "hole" in the graph at the point (1,2). The function  $f(x) = \frac{(x-1)(x+1)}{x-1}$  is not the same as  $f(x) = x + 1$  because they have different domains, but the "hole" at (1,2) is infinitely tiny. Textbooks often indicate this type of discontinuity by marking a little open circle on the graph. If you can't actually see this hole when you use a graphing calculator or software, remember that it really is infinitely small so it should be invisible no matter how much you zoom in.



## Range of a Rational Function

The domain restriction on a rational function affects the range. If the function has a little hole in its graph, the  $y$  value at the hole must be excluded from the range. If there is a vertical asymptote it may take the range to positive or negative infinity, or both.

### Example

Find the range of  $f(x) = \frac{x+1}{x^2-1}$ .

Here we need to make sure that  $x^2 - 1$  is not zero, so we must exclude  $x = 1$  and  $x = -1$ . Another problem we have is that it is not immediately clear what happens to this function when  $x$  is a very large positive or a very large negative number. Since there is a difference of two squares in this function, the easiest way to figure things out is to factor and divide:  $\frac{x+1}{x^2-1} = \frac{x+1}{(x+1)(x-1)}$ .

We end up with the simplified function  $f(x) = \frac{1}{x-1}$ ,  $x \neq 1$  and  $x \neq -1$ .

Now it is easier to see that when  $x$  is a very large positive number the output gets closer and closer to zero, and when  $x$  is a very large negative number the output gets closer and closer to zero too. Try this out with a few actual numbers so you can see what is happening. We end up with a smaller and smaller positive or negative fraction. It is not possible for the function to actually reach 0, so that has to be excluded from the range. Because the function never reaches 0, the  $x$ -axis is considered a **horizontal asymptote** of the function. An asymptote is a line that a function gets close to but never reaches.

The other interesting thing that happens with this function is when we approach the value  $x = 1$ . Taking a value for  $x$  close to 1 on the right (positive) side, such as 1.001, we see that we get a large positive number:

$$\frac{1}{1.001-1} = 1000$$

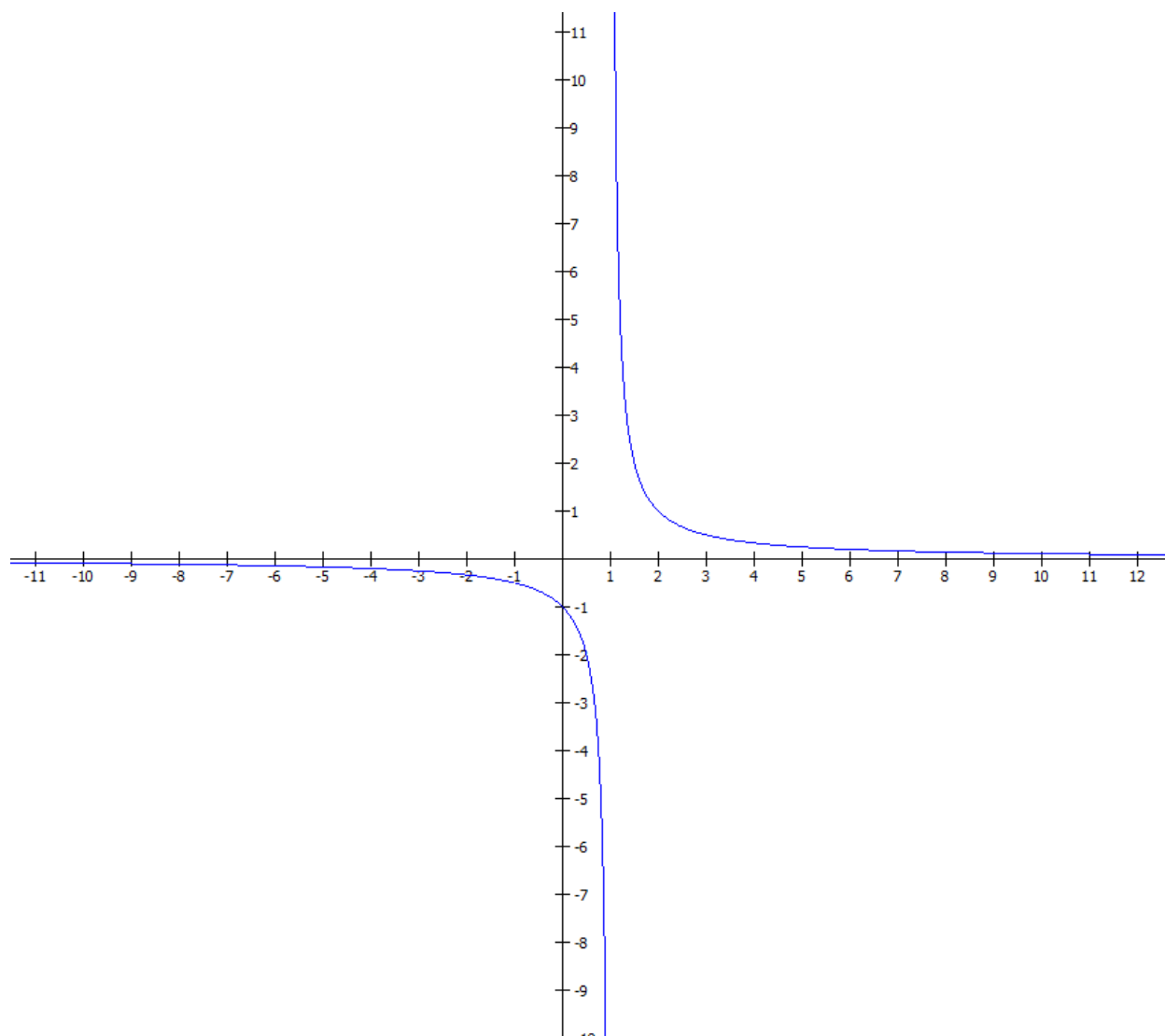
The closer we get to 1, the larger the result:  $\frac{1}{1.00001-1} = 100,000$ , and so on, all the way to  $+\infty$ . Approaching 1 from the left or negative side results in a larger and larger negative number:  $\frac{1}{0.999-1} = -1000$ , and  $\frac{1}{0.99999-1} = -100,000$ , which eventually takes us to  $-\infty$ . The line  $x = 1$  is a **vertical asymptote** of this function. Vertical asymptotes tend to appear at points where the function does not exist (the denominator is 0).

Next we have to check what happens near  $x = -1$ . Because we used the trick of factoring and dividing, we can plug  $-1$  directly into the modified function  $g(x) = \frac{1}{x-1}$ ,  $x \neq 1$  even though we could not put it into the original,  $f(x) = \frac{1}{x-1}$ ,  $x \neq 1$  and  $x \neq -1$ .  $g(-1) = \frac{1}{-1-1} = \frac{1}{-2}$ .

Even though  $g(x)$  will accept the value  $x = -1$  our original function does not. The best it can do is approach the value  $-\frac{1}{2}$  really, really closely.  $f(x) = \frac{x+1}{x^2-1}$  has an infinitely small hole in it at  $(-1, -1/2)$ , so  $-\frac{1}{2}$  must be excluded from the range. The infinitely small hole is plugged when we replace  $f(x)$  by the modified version  $g(x)$ .

So, while the function output goes from  $-\infty$  to  $+\infty$ , the two values that must be excluded from the range are  $-\frac{1}{2}$  and  $0$ .

Here is a picture of  $f(x) = \frac{x+1}{x^2-1}$ . Notice that the hole at  $(-1, -1/2)$  is not visible.



## Horizontal Asymptotes (Limits at Infinity)

In the previous section, we found that the value of the function  $f(x) = \frac{x+1}{x^2-1}$  approaches zero for very large or very small values of  $x$ . We did this by factoring  $x^2 - 1$  into  $(x + 1)(x - 1)$  and dividing, so we could see more easily what happens when  $x$  approaches positive or negative infinity. However, you may not always be able to simplify a function in this way.

In general, when a function is composed of one polynomial divided by another polynomial, there is a handy trick we can always use to find what happens at very small or very large values of  $x$ . This trick is based on the fact that  $\frac{1}{x}$  gets smaller and smaller as  $x$  gets larger and larger. It never actually reaches zero, but we can keep getting closer and closer to 0 by taking larger and larger values of  $x$ . Zero is considered the *limit* of the expression  $\frac{1}{x}$  as  $x$  goes on to infinity. Zero is also the limit of  $\frac{1}{x}$  as  $x$  goes to negative infinity. The word limit is sometimes defined as the value that a function approaches but can never reach. Infinity is hard to imagine, so we could have some philosophical discussions about limits, but for practical purposes the idea is really useful.

Because limits are so useful in math, people have developed a special notation for them. The limit of  $\frac{1}{x}$  as  $x$  goes to infinity can be written in a very compact way as  $\lim_{x \rightarrow \infty} \frac{1}{x}$ .

Let's see how it works.

### Example

For  $f(x) = \frac{x+1}{x^2-1}$ , find  $\lim_{x \rightarrow \infty} f(x)$  and  $\lim_{x \rightarrow -\infty} f(x)$ .

The notation here may look a bit intimidating, but the question is just asking what happens to  $f(x) = \frac{x+1}{x^2-1}$  at very large positive or very negative values of  $x$ . (Determine the *end-behavior* of  $f(x)$  as  $x$  approaches positive or negative infinity.) Since we did this earlier we already know the answer, and we can just focus on doing this in a different way.

To take advantage of the limit of  $\frac{1}{x}$  as  $x$  gets very large or very negative, **divide the top and bottom polynomial by the highest power of  $x$  in the denominator**. Note that this means dividing every single term, because it is easy to forget some. Here  $x^2$  is the highest power of  $x$

in the denominator, so that is what we divide by. For  $f(x) = \frac{x+1}{x^2-1}$  that gives us  $f(x) = \frac{\frac{x}{x^2} + \frac{1}{x^2}}{\frac{x^2}{x^2} - \frac{1}{x^2}}$ .

Simplifying, we get  $f(x) = \frac{\frac{1}{x} + \frac{1}{x^2}}{1 - \frac{1}{x^2}}$ . Now, if we say that the limit of  $\frac{1}{x}$  as  $x$  goes to infinity is zero, it

makes sense that the limit of  $\frac{1}{x^2}$  as  $x$  goes to infinity would also be zero. This expression actually approaches zero faster, but it never actually gets there either (hmm...). Anyway, we can make a nice approximation by replacing all of the expressions  $\frac{1}{x^n}$  with the number 0, so that  $\frac{\frac{1}{x} + \frac{1}{x^2}}{1 - \frac{1}{x^2}}$  turns into  $\frac{0+0}{1-0}$ , which is 0. This tells you that as  $x$  gets larger and larger, the output of the function approaches 0, although it never actually gets there. Zero is not included in the range of the function. Conveniently,  $\frac{1}{x^n}$  also goes to zero as  $x$  gets to be more and more negative, giving us the same result in this case for smaller and smaller values of  $x$ . Note that if you were looking for the range rather than the end behavior you would still have to go back and check what happens near the values 1 and -1, which is a little more work if you don't simplify the function first.

Let's look at a slightly more difficult example:

### Example

Determine the end behavior of  $f(x) = \frac{x^3 + 3x^2 + 4x + 7}{x^5 + x^4 + 2x^2 + 10}$ .

What happens to this function when we have very large or very small values of  $x$ ? To find out, divide all the terms by the largest power of  $x$  in the denominator, which is  $x^5$  in this case. After

simplifying, we end up with  $f(x) = \frac{\frac{1}{x^2} + \frac{3}{x^3} + \frac{4}{x^4} + \frac{7}{x^5}}{1 + \frac{1}{x} + \frac{2}{x^3} + \frac{10}{x^5}}$ . Now we have to stop and consider what

happens to something like  $\frac{10}{x}$  when  $x$  gets very large (or very negative). If you think about that for a while you will realize that it is just the same as for  $\frac{1}{x}$ ; the expression eventually gets very close to zero. If we replace all of the expressions of the form  $\frac{a}{x^n}$  with zero, we see that the value of this function approaches 0 as  $x$  approaches positive or negative infinity.

Your textbook will likely have many examples for you to try. After you have been doing these problems for a while, you will notice a pattern:

1. If the power of the leading term (leading term = term with the highest power) on the top is smaller than the power of the leading term on the bottom, the rational expression will always approach 0. The  $x$ -axis will be a horizontal asymptote of the function, unless there is a vertical shift.

2. If the leading term on the top has the highest power, the function will approach positive or negative infinity.
3. If the powers of the top and bottom leading terms are equal, the rational expression approaches a definite number which is determined by the coefficients of the leading terms. There will be a horizontal asymptote.

All of this makes sense, because the leading terms are really the only ones that matter when  $x$  gets very large or very negative. For example, you would expect  $\frac{x^3}{x^2}$  to go to infinity as  $x$  gets larger and larger.  $\frac{x^3}{2x^2}$  also keeps getting bigger as  $x$  gets larger. It works the same way for  $\frac{x^3+3x^2+5x+12}{2x^2-15}$ . If the power of the leading term on the bottom is larger, it will overpower the top and squish it to pretty well to zero eventually. If the top and bottom powers are equal, they will balance out to 1, so look at the coefficient to find out how things end up.

## Slant Asymptotes

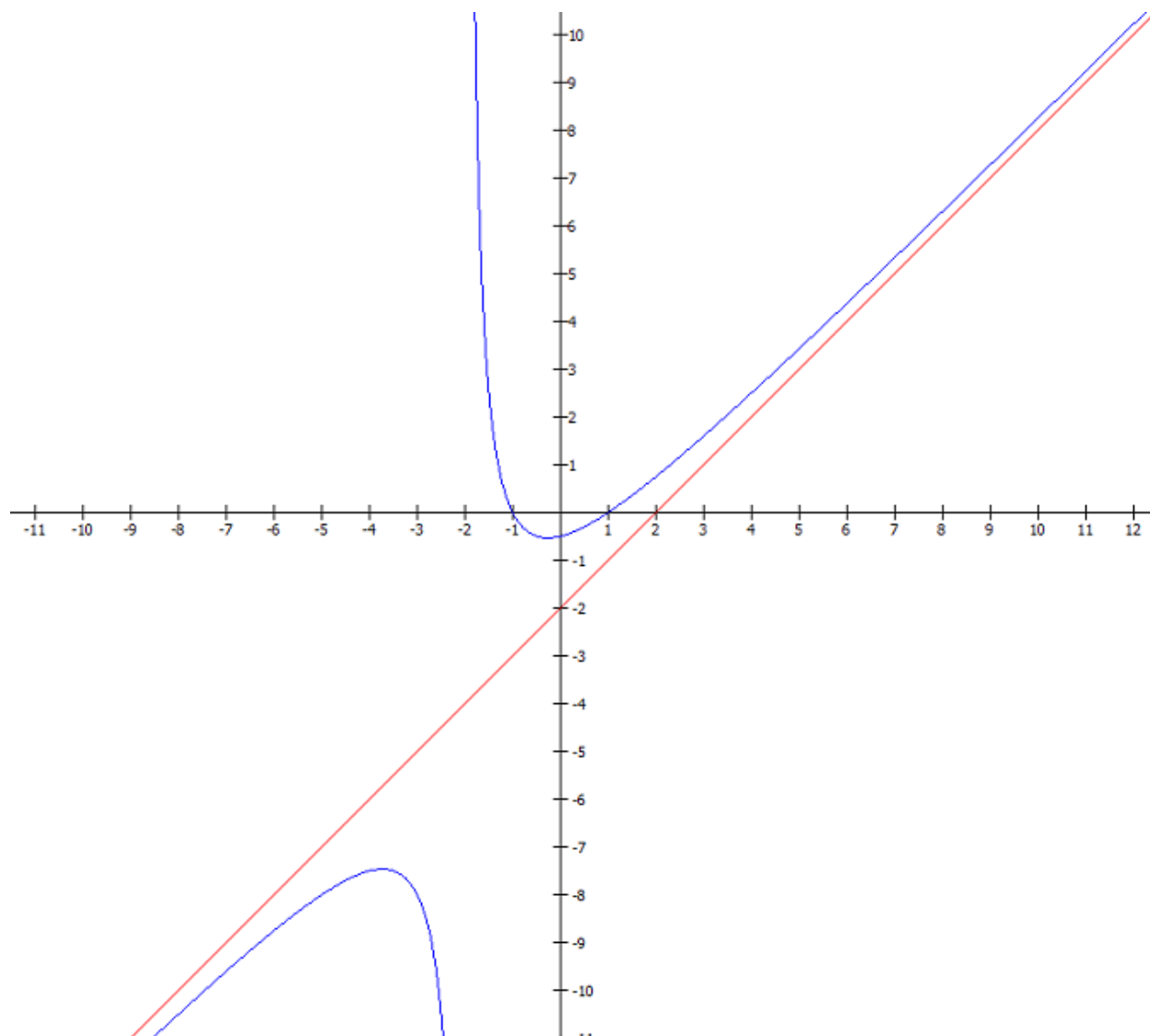
If the power of the top leading term is one higher than the power of the bottom leading term, the function will have a *slant asymptote*.

A slant asymptote is a line that is not horizontal or vertical. Like other lines, it can be described by a linear equation. The function approaches the line but never touches it. What creates this slant asymptote? Let's look at an example.

### Example

Describe the end behavior of  $f(x) = \frac{x^2-1}{x+2}$ .

We expect this function to have a vertical asymptote at  $x = -2$ , but it also has a slant asymptote. Use long division or synthetic division to divide  $x^2 - 1$  by  $x + 2$ . The result is  $x - 2$ , with a remainder of 3. This can be written as  $x - 2 + \frac{3}{x+2}$ . When  $x$  gets larger and larger, the remainder  $\frac{3}{x+2}$  gets smaller and smaller until it approaches zero. At large values of  $x$ , the function can be approximated by  $x - 2$ , although it never quite gets there. The same goes for very small values of  $x$ . The red line in this picture is  $y = x - 2$ :



Whenever a rational function has a top leading term that is one degree higher than the bottom leading term, the result of the division is a linear expression and a remainder that gets close to zero at very small or very large values of  $x$ .

## Graphing Rational Functions

1. Factor any polynomials and simplify if possible

If a factor cancels out, you need to mark a hole in the function at the point where the factor would be zero. For example, the function  $f(x) = \frac{x-3}{(x+2)(x-3)}$  is almost the same as  $g(x) = \frac{1}{(x+2)}$ , except for the fact that  $f(x)$  doesn't exist at  $x = 3$ .

## 2. Mark the vertical asymptotes

After you have cancelled out any factors that are the same, vertical asymptotes occur where there is a potential zero in the denominator.  $f(x) = \frac{1}{(x+2)}$ ,  $x \neq -2$  has a vertical asymptote at  $x = -2$ . At the vertical asymptotes, the function goes off into negative or positive infinity. Pick a test value close to the vertical asymptote to see if the function will be positive or negative there. Unless the factor that causes the vertical asymptote is raised to an even power, the function will switch signs on the other side of the asymptote.  $f(x) = \frac{1}{(x+2)}$  is negative on the left of the asymptote at  $x = -3$ , so it will be positive on the right.  $f(x) = \frac{1}{(x+2)^2}$  is positive at  $x = -3$ , so it will also be positive to the right of the asymptote.

## 3. Mark the horizontal asymptote

If the power of the leading term (leading term = term with the highest power) on the top is smaller than the power of the leading term on the bottom, the rational part of the function will always approach 0. The x-axis will be a horizontal asymptote, unless there is something added or subtracted from the rational part.  $f(x) = \frac{x-3}{(x+2)(x+7)}$  has a higher power of  $x$  on the bottom, which you can see if you multiply the factors out. The x-axis is a horizontal asymptote.

$f(x) = \frac{x-3}{(x+2)(x+7)} + 4$  has a horizontal asymptote at  $x = 4$ . If the leading term on the top has the highest power, the function will approach positive or negative infinity and there is no horizontal asymptote. If the powers of the top and bottom leading terms are equal, the function approaches a definite number which is determined by the coefficients of the leading terms. There will be a horizontal asymptote.

## 4. Mark the zeros, if any.

The zeros of a rational function occur where the numerator is equal to zero. For example,



$f(x) = \frac{(x-4)(x+5)}{(x-1)}$  has a value of zero when  $x = 4$  or when  $x = -5$ .

### 5. Decide if the graph crosses the x-axis at the zeros.

The graph crosses the x-axis where there is a zero of odd multiplicity, and just touches it where there is a zero of even multiplicity. To understand why this happens, look at values of  $x$  close to the zero value. If  $(x - 4)$  is one of the factors in the numerator, its value zero at  $x = 4$ , which makes the entire function have a value of zero. Just before that, say at  $x = 3.9$ , the value of the factor  $(x - 4)$  will be negative, and after that, at  $x = 4.1$ , the factor will be positive. This changes the sign of the function before another term can interfere, and results in the graph crossing the x-axis. Now let's look at the case where there is an even power of a factor. If the term  $(x - 4)$  is present twice, we can write it as  $(x - 4)^2$ .  $(x - 4)^2$  is positive at any value of  $x$  close to 4, whether that value is slightly less than 4 or slightly greater. The function doesn't change signs and just touches the x-axis at the zero. On the other hand,  $(x - 4)^3$  is negative when  $x$  is just a little smaller than 4, and positive when  $x$  is just a little larger than 4. This causes a change in sign and the function will cross the x-axis at  $x = 4$ .

### 6. Sketch the graph

Draw the ends of your graph with an arrow, to indicate that the function will continue up, down, or horizontally. Guide the graph through the zeros, either crossing or touching the x-axis. If you have to turn around between zeros, you are generally not expected to know how high or low the graph reaches before it makes the turn. To get an idea, you can determine a y-value for a random  $x$  that lies between the zeros.

## Rational Inequalities

Rational inequalities have an unknown in the denominator, and that unknown may appear in the numerator also.  $\frac{x+3}{5} > 4$  is not considered a rational inequality, because we can just

multiply both sides by 5 to get  $x + 3 > 20$ . However,  $\frac{1}{x} > 2$  is not so easy to change into a regular inequality. We might be tempted to multiply by  $x$  to get  $1 > 2x$ , but there is a bit of a problem with that. Try  $-\frac{1}{4}$  as a test value:  $1 > 2 \cdot -\frac{1}{4}$  is true, but  $\frac{1}{-1/4} > 2$  is not. The fact is that  $x$  is an

unknown that can be either positive or negative, so if we multiply by  $x$  we should turn the inequality sign around if  $x$  is negative, and leave it the same if  $x$  is positive. This would lead to two separate inequalities, one for a negative  $x$  and one for a positive  $x$ .

So, can we still create a common denominator so we can just work with the top part?

Let's try  $\frac{x-2}{x-5} \geq 3$ .

First, I will create a common denominator:

$$\frac{x-2}{x-5} \geq \frac{3(x-5)}{x-5}$$

Now I should be able to say that  $x - 2 \geq 3(x - 5)$ . That works out to  $x - 2 \geq 3x - 15$ , and I can subtract  $x$  from both sides:  $-2 \geq 2x - 15$ , which means that  $13 \geq 2x$  and  $x \leq 6.5$ . Any  $x$  less than or equal to 6.5 should work, so I can just pick the simplest one:  $x = 0$ . Unfortunately, when we substitute  $x = 0$  into the original equation we get  $\frac{0-2}{0-5} \geq 3$ . Our solution doesn't work once we include the denominator, which is negative when  $x = 0$ ! That makes sense, because dividing by a negative number like that flips the inequality sign. We need a different approach.

**First, we want to get a zero on one side of the inequality.** For  $\frac{x-2}{x-5} \geq 3$  we should subtract 3 on both sides:  $\frac{x-2}{x-5} - 3 \geq 0$ . Next, we put everything over a common denominator *so we can deal with the numerator and the denominator separately*:

$$\frac{x-2}{x-5} - \frac{3(x-5)}{x-5} \geq 0$$

$$\frac{x-2-(3x-15)}{x-5} \geq 0$$

$$\frac{x-2-3x+15}{x-5} \geq 0$$

$$\frac{-2x+13}{x-5} \geq 0$$

If the numerator and the denominator are both positive or both negative the inequality is true, and if they have opposite signs the inequality is false. **The expression can only change signs where the numerator is zero or where the denominator would be zero.** If the numerator is zero, the value of the whole expression is zero:  $\frac{0}{x-5} = 0$ . To find where this happens, we set the numerator equal to 0:  $-2x + 13 = 0$ , which gives a value of 6.5 for  $x$ . This is the point where the

function  $y = \frac{-2x+13}{x-5}$  would touch the x-axis, so this is where we look for the expression to potentially change from positive to negative or the other way around. Using 6 as a test value, I see that  $\frac{-2x+13}{x-5}$  is positive when  $x < 6.5$ . When  $x$  is equal to or greater than 6.5, the expression is negative and the inequality is true. However, the function  $y = \frac{-2x+13}{x-5}$  has another trick up its sleeve. It does not exist at  $x = 5$ , which means that it has a vertical asymptote there. Function values may be the same on either side of an asymptote, or they may be opposite in sign. Since  $x - 5$  is not being raised to some even power,  $\frac{-2x+13}{x-5}$  will change signs on either side of the asymptote. We already know that the function is positive on the right of the asymptote at  $x = 6$ , so on the left side the function is negative and the inequality is false. The solution is the interval  $(5, 6.5]$ . Note that 6.5 is included, while 5 is not.

Even though  $x = 5$  is not in the domain of the function, it obviously has to be considered when we are solving the inequality. **In general, we need to find the zeros of the denominator as well as the zeros of the numerator.** Order these numbers from least to greatest, and check what happens in between them.

## Partial Fraction Decomposition

1. You must start with a proper fraction.
2. The resulting decomposed fractions must also be proper fractions.
3. If the original expression contains a power of a prime polynomial in the denominator, consider only the degree of this prime polynomial when determining the degree of the numerator of your partial fractions, not the power to which it is raised. A separate fraction will account for each power.

Rational expressions can be combined into a single rational expression, but it is also possible to reverse the process.

We have already seen how to combine an expression like  $\frac{5}{x+2} - \frac{5}{x+5}$  into a single rational expression like this:

$$\frac{5(x+5)}{(x+2)(x+5)} - \frac{5(x+2)}{(x+5)(x+2)} = \frac{5(x+5) - 5(x+2)}{(x+2)(x+5)} = \frac{5x+25-5x-10}{(x+2)(x+5)} = \frac{15}{(x+2)(x+5)}.$$

Let's see if we can also do that the other way around.

Starting with  $\frac{15}{(x+2)(x+5)}$ , we would like to decompose that into  $\frac{A}{x+2}$  and  $\frac{B}{x+5}$ , but how to find the numbers at the top? Algebra helps us find unknowns by giving them names, so let's call our unknowns in this case A and B. By convention capital letters are always used for this purpose. They always start at A and go down the alphabet as far as needed.

$$\frac{15}{(x+2)(x+5)} = \frac{A}{x+2} + \frac{B}{x+5}$$

Now we apply the same process that allowed us to add the two fractions in the first place:

$$\frac{15}{(x+2)(x+5)} = \frac{A(x+5)}{(x+2)(x+5)} + \frac{B(x+2)}{(x+5)(x+2)}$$

$$\frac{15}{(x+2)(x+5)} = \frac{A(x+5)+B(x+2)}{(x+2)(x+5)}$$

Since the denominator is the same on both sides of the equation, the numerators must be equal:

$$15 = A(x + 5) + B(x + 2)$$

This expression has to be true for any value of  $x$ , so we can just pick a convenient number. If  $x = -2$ , the part with  $B$  disappears:

$$15 = A(-2 + 5) + B(0)$$

$$15 = 3A$$

$$A = 5$$

When we select  $x = -5$ , we can easily find  $B$ :

$$15 = 0 + B(-5 + 2)$$

$$15 = -3B$$

$$B = -5$$

Using some rather complex-looking mathematical reasoning, it can be shown that this kind of decomposition is always possible, although you may end up with fractions in your numerators. The proof for this involves certain restrictions, which you will have to memorize as rules.

### 1. You must start with a proper fraction.

If the degree of the numerator is not lower than the degree of the denominator, do a long division first.

### 2. The resulting decomposed fractions must also be proper fractions.

Chances are that your course will make sure there will simple first degree polynomials in the denominator, but if not you will have to allow for the numerator to have a higher degree.

### Example

Decompose  $\frac{3x^2+2}{x(x^2+1)}$  by using the method of partial fractions.

Here one of the fractions will have a degree of 2, so we will allow for the numerator to have a degree of 1:

$$\frac{3x^2+2}{x(x^2+1)} = \frac{A}{x} + \frac{Bx+C}{x^2+1}$$

$$\frac{3x^2+2}{x(x^2+1)} = \frac{A(x^2+1)}{x(x^2+1)} + \frac{x(Bx+C)}{x(x^2+1)}$$

$$\frac{3x^2+2}{x(x^2+1)} = \frac{A(x^2+1) + x(Bx+C)}{x(x^2+1)}$$

So now we know that  $A(x^2 + 1) + x(Bx + C) = 3x^2 + 2$ , for any  $x$ . When  $x = 0$ , that looks much simpler:

$$A(0 + 1) + 0 = 0 + 2$$

$$A = 2$$

Although there is not such a convenient value for  $x$  to find  $B$ , we can use what we know about  $A$ :

$$2(x^2 + 1) + x(Bx + C) = 3x^2 + 2$$

$$2x^2 + 2 + Bx^2 + Cx = 3x^2 + 2$$

$$2x^2 + Bx^2 + Cx = 3x^2$$

Since the coefficient of  $x^2$  has to be 3,  $B$  must be equal to 1 so that  $2x^2 + 1x^2$  will be equal to  $3x^2$ . Also, there is no term with  $x$  on the right, so  $C$  must be 0. This illustrates an important point, which is that it is always possible for one or more of the unknowns to be zero.

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

**3. If the original expression contains a power of a prime polynomial in the denominator, consider only the degree of this prime polynomial when determining the degree of the numerator of your partial fractions, not the power to which it is raised. A separate fraction will account for each power.**

### Example

Decompose  $\frac{2x}{(x+1)^2}$  into two partial fractions.

$x + 1$  is a prime polynomial, since it cannot be factored. The maximum degree of the numerator of the partial fractions is determined by the degree of  $x + 1$ , which is 1. The numerators can just be numbers (degree = 0).

Each power of  $(x + 1)^2$  must be accounted for, so we have to create two partial fractions with numerators  $(x + 1)^1$  and  $(x + 1)^2$ . If we had  $(x + 1)^3$  instead, we would have to create three fractions with denominators  $(x + 1)$ ,  $(x + 1)^2$  and  $(x + 1)^3$ .

$$\frac{2x}{(x+1)^2} = \frac{A}{x+1} + \frac{B}{(x+1)^2}$$

$$\frac{2x}{(x+1)^2} = \frac{A(x+1)}{(x+1)^2} + \frac{B}{(x+1)^2}$$

$$\frac{2x}{(x+1)^2} = \frac{A(x+1) + B}{(x+1)^2}$$

$$2x = A(x+1) + B$$

When  $x = 0$ :

$$0 = A + B$$

$A = -B$  Now use this information to substitute for  $A$  (or  $B$  if you prefer):

$$2x = -B(x+1) + B$$

$$2x = -Bx - B + B$$

$$B = -2$$

Since  $A = -B$ ,  $A = 2$

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

## Exponential Functions and Logarithms

Basic exponential function:  $y = b^x$ .

$y = b^x$  always passes through the point (0,1), horizontal asymptote at  $y = 0$ .

Shifted exponential function:  $y = cb^{x-h} + k$ , horizontal asymptote at  $y = k$ .

When you use the number  $e$  as the base of a simple exponential function, the function increases at a rate equal to its value at any given point.

Basic logarithmic function:  $y = \log_b x$

$y = \log_b x$  always passes through the point (1,0), vertical asymptote at  $x = 0$ .

Shifted logarithmic function:  $y = c \log_b (x - h) + k$ , vertical asymptote at  $x = h$

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

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

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

$$\text{Change of base formula: } \log_a x = \frac{\log_b x}{\log_b a}$$

The logarithmic function is the inverse of the exponential function:  $b^{\log_b x} = x$

Simple interest:  $I = Prt$

Total amount of money  $A$  that you have after  $y$  years if the interest is compounded  $n$  times per year:  $A = P\left(1 + \frac{r}{n}\right)^{ny}$

Continuously compounded interest formula:  $A = Pe^{rt}$

$A = C\left(\frac{1}{2}\right)^{\frac{y}{h}}$   $A$  is the amount remaining after  $y$  years for a substance with a half-life of  $h$  years if the original amount was  $C$ .



## Exponential Functions

When people say that something is increasing exponentially, they mean that it is getting bigger very fast, although they may not always use the term in a mathematically correct sense. An example of a true exponential increase is the initial growth of a bacterial colony. Each bacterium divides in two to reproduce, so the more bacteria there are the faster the colony grows: 1 bacterium  $\rightarrow$  2 bacteria  $\rightarrow$  4 bacteria  $\rightarrow$  8 bacteria  $\rightarrow$  16 bacteria .... The function  $y = x^2$  increases very quickly with increasing values of  $x$ , but  $y = 2^x$  increases a lot faster. When  $x = 2$ ,  $x^2$  is equal to  $2^x$ . When  $x = 10$  however,  $x^2$  is 100 while  $2^x = 1024$ . For the function  $y = 2^x$ , the value of the function is multiplied by 2 every time  $x$  increases by 1.

Basic exponential functions have the form  $y = b^x$ . Here  $b$  is the “base” that is raised to whatever power you choose for  $x$ . The base  $b$  cannot be a negative number because of alternating positive and negative values, and missing points for some fractional exponents [ $(-4)^{1/2}$  for example would be a problem since that is  $\sqrt{-4}$ .] On the other hand, negative values for  $x$  are perfectly fine to use. Remember that  $b^{-3}$  means  $\frac{1}{b^3}$ , which is always a positive value because  $b$  is never negative.

All basic (unmodified) exponential functions have graphs that are entirely above the  $x$ -axis. They get really, really close to the  $x$ -axis, but they can never actually touch it (the  $x$ -axis is a horizontal asymptote). There is no exponent that you can put on the base  $b$  that gives you a value of zero; try it out for yourself. Since  $b^0 = 1$ , the graph always passes through the point  $(0,1)$ , for any value of  $b$ . When  $b > 1$ , the value of an exponential function increases quickly (exponential growth). Values for  $b$  that are between 0 and 1 are interesting because they result in rapidly decreasing values (exponential decay). We can't use 0 or 1 as values for the base, because these numbers would give straight lines rather than exponential curves.

## Modified Exponential Functions

We can modify basic exponential functions in several ways. This is often useful when we want to create a function to model exponential growth that we see in experimental data. When we look at such data, we usually find that they don't have a value of 1 at the starting point. There may be some initial amount or a certain number of organisms, and then there is an exponential increase or decay over time. You can take the function  $y = b^x$  and multiply it by some constant, say 10, to get  $y = 10b^x$ . Now the function values will start at 10 when  $x = 0$ :  $y = 10 \cdot b^0 = 10 \cdot 1 = 10$ . For the general constant  $c$ , the function  $y = cb^x$  will have a value of  $c$  at

$x = 0$ . Instead of passing through the point  $(0,1)$  the graph will pass through the point  $(0,c)$ . This multiplication does not affect the horizontal asymptote.

Just like any other graph, the graph of  $y = cb^x$  can be shifted up or down by adding or subtracting some constant. To shift the graph of  $y = 10(3)^x$  down by 5 units, just write  $y = 10(3)^x - 5$ . Before the shift, the  $x$ -axis was a horizontal asymptote of the function. When you shift the graph down by 5 units, the horizontal asymptote also shifts down, to the line  $y = -5$ . In general, shift the graph of  $y = cb^x$  up or down by writing  $y = cb^x + k$ , where  $k$  is a positive or a negative number. The horizontal asymptote will then be the line  $y = k$ .

One example of a modified exponential function is Newton's Law of cooling. This law lets you calculate the temperature of a hot object with an initial temperature  $T_i$ , at any given time  $t$ . How fast the object cools depends on the difference between its initial temperature and the temperature of its surroundings,  $T_s$ :

$$T(t) = T_s + (T_i - T_s)e^{kt}$$

where  $k$  is a small negative constant that can be determined experimentally. Notice that this is an exponential function that has been shifted up by an amount equal to  $T_s$ , the temperature of the surroundings. If you are working with experimental data you will want to subtract  $T_s$  from all of the measurements so that you can determine the exponential function more easily.

For any function of  $x$ , we can subtract something directly from  $x$  in the formula. If we put in  $(x - 2)$  instead of  $x$ , that means that  $x$  now has to be 2 units bigger to do the same job that it did before. The curve shifts to the *right*. That is a little counterintuitive, since we are subtracting and we might expect that the graph would shift to the left if we don't stop and think about it carefully. If we add something to  $x$  in the formula, the graph shifts to the *left*. The general formula looks like this:  $y = cb^{x-h}$ , or  $y = cb^{x-h} + k$  if you are shifting up or down also.

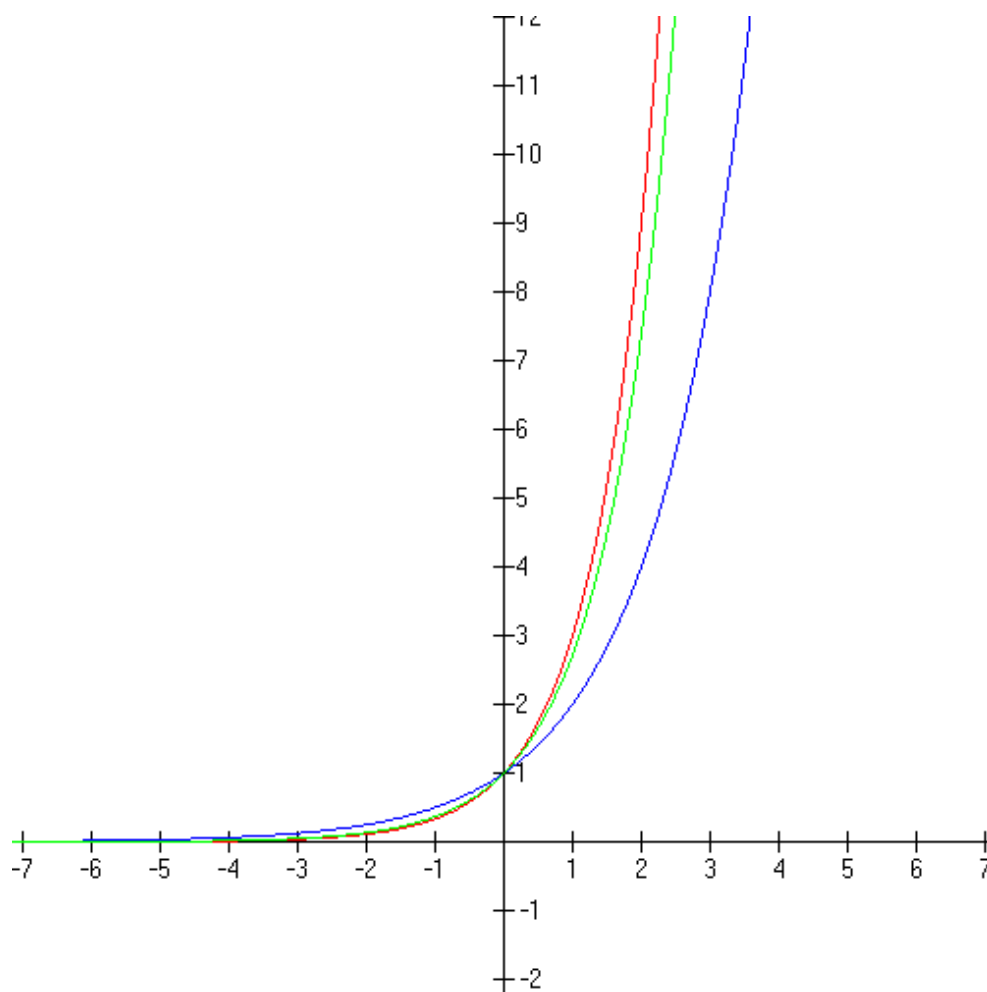
## The Number $e$

Exponential functions increase faster and faster as their  $y$ -value increases. The rate of increase at a particular point depends on the value of the function at that point, just like the number of new rabbits there will be in a growing colony a year from now depends on how many rabbits there are right now. Consider the function  $y = 2^x$ . When  $x = 1$  the value of the function (the  $y$  value) is 2. Because the function increases ever faster as it goes, we want to approximate the rate of increase by looking at the value of the function between  $x = 1$  and a point very close to

that, say  $x = 1.001$ . At  $x = 1.001$ ,  $y = 2^{1.001} = 2.00138677$ . The function value has increased by 0.00138677. That means that  $y$  has increased by 0.00138677 units for an increase of 0.001 units in  $x$ . The *average rate of increase* is defined as the change in  $y$  divided by the change in  $x$ , or  $0.00138677 \div 0.001$  which works out to 1.38677. When  $x = 3$ ,  $y = 8$ , and at  $x = 3.001$   $y$  is equal to 8.00554710. Now the rate of increase is approximately  $(8.0055471 - 8) \div 0.001 = 5.54710$ . So, when the value of the function is 2, the rate of increase is 1.38677, and when the value of the function is 8, the rate of increase is 5.54710. The function  $y = 2^x$  increases at a rate that is less than its value, at any given point. If you do the same calculations for the function  $y = 3^x$ , you will find that at any given point the function is increasing faster than its value at that point – see table below. There is a base, a number between 2 and 3, for which the function  $y = b^x$  increases at a rate exactly equal to itself at any point. This special number is the number  $e$ . Like  $\pi$ ,  $e$  has an infinite number of digits. The number  $e$  is 2.7182818228..., and it can be found on your TI 83 or 84 graphing calculator above the division key. The table below uses spreadsheet calculations to find the rate of increase for the three functions  $y = 2^x$ ,  $y = e^x$ , and  $y = 3^x$  more accurately by measuring over very small intervals. Notice that at  $x = 1$ , the value of the function  $y = e^x$  is approximately 2.718, and it is increasing at a rate of 2.718. When  $x = 2$ , the function value is 7.389, and the rate of increase is also 7.389, and so on.

$x$	$y = 2^x$	Rate of Increase	$y = e^x$	Rate of Increase	$y = 3^x$	Rate of Increase
1	2	1.386342	2.718282	2.718418	3	3.296018
1.0001	2.000139		2.718554		3.00033	
2	4	2.772685	7.389056	7.389426	9	9.888054
2.0001	4.000277		7.389795		9.000989	
3	8	5.54537	20.08554	20.08654	27	29.66416
3.0001	8.000555		20.08755		27.00297	
4	16	11.09074	54.59815	54.60088	81	88.99248
4.0001	16.00111		54.60361		81.0089	

This image shows the graphs of  $y = 2^x$  (blue),  $y = e^x$  (green), and  $y = 3^x$  (red):



The number  $e$  is a “natural” base for an exponential function, providing a perfectly balanced rate of increase. This amazing number can also generate a spiral that occurs naturally in many places in the real world. The logarithmic spiral is generated by graphing the equation  $r = e^\theta$ . This is an equation in polar coordinates instead of Cartesian coordinates, and it uses the radius of the curve as it relates to the angle  $\theta$  that the radius makes with the positive part of the  $x$ -axis. Polar coordinates greatly simplify the equations of many curves. In the picture below, you can see the radius at a 45 degree angle drawn as a red line. I actually used the modified equation  $r = e^{0.00349\theta}$  to generate the curve so it would produce a nice image. If you substitute 45 degrees for  $\theta$ , you can see that the particular radius represented by the red line has a length of about 1.17 units. As the angle  $\theta$  increases, the radius also increases in such a way that the curve keeps its same shape. Below the graph is a photograph of a Nautilus shell, which allows

the animal to keep the same shape as it grows.

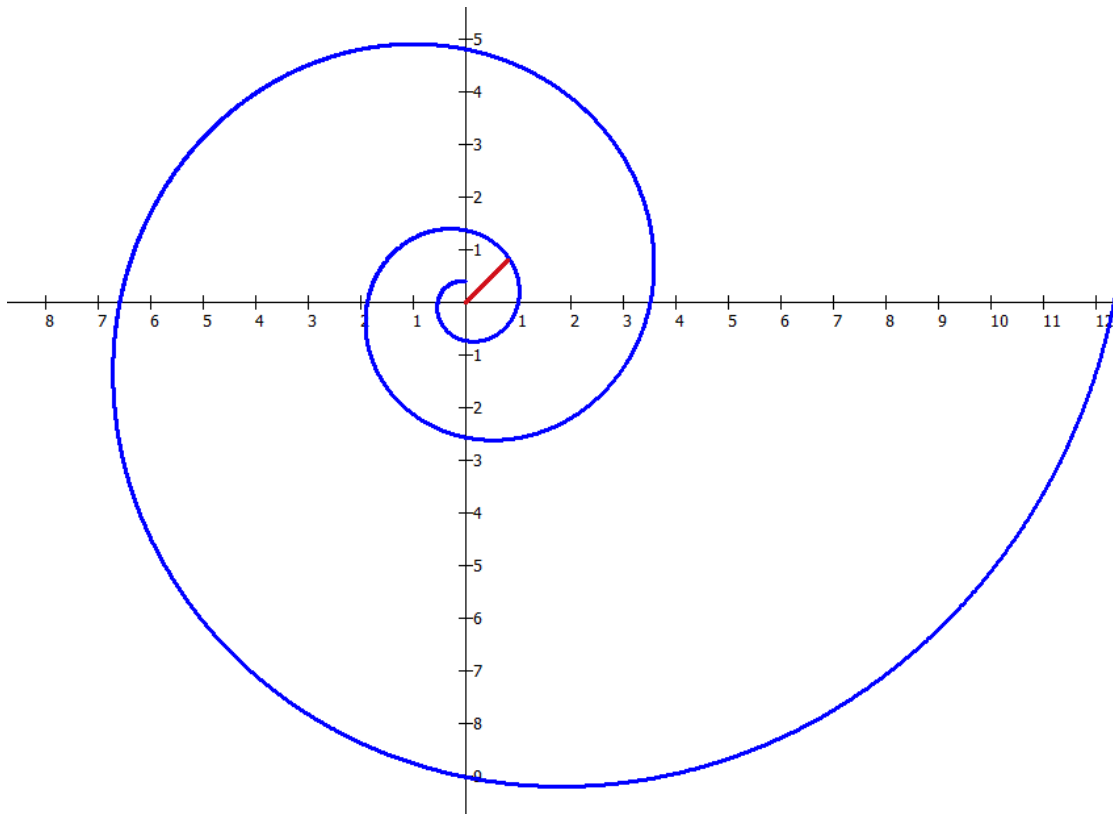


Photo by Chris 73

The number  $e$  is related to the number  $\pi$  and the imaginary number  $i$  by the equation  $e^{\pi i} = -1$ .

## Compound Interest and Other Increases (and Decreases)

One way of getting interest is through simple interest. With simple interest, your interest is paid out to you directly and not added to your account. If you do not add more money to your account (or withdraw from it), each interest payment is the same. Calculating simple interest is straightforward. First you determine the principal, which is the original amount of money that you deposited. You multiply the principal  $P$  by the interest rate  $r$ , to get the interest  $Pr$ . For example, if you have \$100 and the interest rate is 5% per year, the interest will be 100 times 5%, or 100 times 0.05, which comes out to \$5. If you invest your money for a longer period at a simple interest rate, your total interest will be  $I = Prt$ , where  $t$  is the number of years, or the number of other time intervals at which the interest is paid out.

When you get compound interest on your money, that means that the interest is added to your account, and the next time interest is calculated you get interest on the new amount. This gives you interest on your interest, so your money grows more quickly. Regular bank accounts are set up this way. To simplify things, we will first look at interest that is compounded yearly (annually). The first time the interest is calculated, it is just like simple interest. Your principal is multiplied by the interest rate  $r$  to get the amount of interest. Now you have your principal  $P$ , and also the interest,  $Pr$ , where  $r$  is expressed as a fraction or decimal. For example, if the interest rate is 5%, use 0.05 for  $r$ . The amount in your account is  $P + Pr$ , which can also be written as  $P(1 + r)$ , so your money has been multiplied by  $1 + r$ . The multiplication by 1 is what preserves your original amount, and the multiplication by  $r$  provides the interest. This multiplication by  $1 + r$  happens every year. For year 2, you have  $P(1 + r)$  in your account. Again your money is multiplied by  $1 + r$ , so now you have  $P(1 + r)(1 + r)$ . After 3 years, the money in your account is  $P(1 + r)(1 + r)(1 + r)$ . We can write that as  $P(1 + r)^3$ . After  $y$  years, the total amount is  $P(1 + r)^y$ .

Most banks compound your interest monthly rather than yearly. Unfortunately when they do this calculation, they don't multiply your money by  $(1 + \text{the yearly rate})$ ; you only get  $1/12^{\text{th}}$  of the annual rate at this time. However, over the course of a year, your money is multiplied by  $1 + r/12$  twelve times, which is  $P$  times  $(1 + \frac{r}{12})^{12}$ . After  $y$  years of this monthly compounding, your money has been multiplied by  $(1 + \frac{r}{12})^{12y}$  times, for a total of  $P$  times  $(1 + \frac{r}{12})^{12y}$ .

Do your own calculations to see that after 3 years of 5% simple interest on \$100 you would have \$115.00 (assuming you didn't spend your interest payment when you received it). With interest compounded yearly, an initial amount of \$100 will increase to a total of \$115.76. The same principal invested at 5% compounded monthly for 3 years would grow to \$116.15. The more frequently the interest is compounded the faster your money increases, because you get interest on your interest more often. If interest is compounded  $n$  times yearly, then after  $y$  years you have  $P(1 + \frac{r}{n})^{ny}$ . Each year your money is multiplied by  $(1 + \frac{r}{n})^n$ . As  $n$  gets very large the expression  $(1 + \frac{r}{n})^n$  reaches a limit, because the extra bits of interest we are compounding on get smaller and smaller. This is where we again meet the number  $e$ , or 2.7182818..... As  $n$  gets bigger and bigger,  $(1 + \frac{r}{n})^n$  gets closer and closer to  $e^r$ . This can be shown by using larger and larger values of  $n$  (grab a calculator and try it out), or through calculus. If the bank's computer system was busy calculating your interest day and night, after  $y$  years you would have an amount very close to  $Pe^{ry}$  in your account (where  $P$  is your principal and  $r$  is the interest rate).

There are other problems that are very similar to compound interest problems. These are cases where the increase or decrease depends on the quantity already present, such as the population growth of a city, or the number of people remaining in a stadium at a given time after a sports event. The increase or decrease is given as a percentage.

If the rate is stated as a percentage per unit of time, such as "an annual increase of 12%", use the compound interest formula  $A = P(1 \pm r)^t$ .  $A$  is the amount at time  $t$ ,  $P$  is the original amount, and  $r$  is the rate of increase or decrease. You would use the  $+$  sign when there is an increase, and the minus sign if there is a decrease. Pay attention to whether the time  $t$  is in years, months, weeks or days.

If the rate is applied continually you are expected to use the Pert formula:  $A = Pe^{rt}$ . The problem usually makes it clear that this is the case.

## Annualized Return

Math textbooks used to heavily emphasize the power of compound interest, but there can be times when interest rates are very low or even negative. At those times people don't want to put money in the bank, and look for other investments instead. When you do that, you need to be able to compare profits from various investments on a yearly basis. For example, suppose you buy \$1000 worth of shares of a company, and decide to sell them 5 years later at \$1150. Your profit is \$150 on \$1000, or 15%. Was that a good deal compared to investing in something

else? You may think that your yearly (annualized) return here was 3% per year, but it is actually a bit more complicated because of compounding. Think backwards from compound interest: what yearly rate of interest would give you \$1150 after 5 years?

$$A = P(1 + r)^y$$

$$1150 = 1000(1 + r)^5$$

Now we can calculate  $r$ , the annualized return:

$$\frac{1150}{1000} = (1 + r)^5$$

Take the 5<sup>th</sup> root on both sides:

$$\left(\frac{1150}{1000}\right)^{\frac{1}{5}} = 1 + r$$

$$r = \left(\frac{1150}{1000}\right)^{\frac{1}{5}} - 1 \approx 0.0283 \text{ or about } 2.83\%$$

That's not quite as good as you might have thought!

In general, the annualized return is:  $r = \left(\frac{\text{end value}}{\text{start value}}\right)^{\frac{1}{y}} - 1$

If you are discouraged by this and decide to sell your \$1000 worth of stock after only 4 months, at a profit of \$15, or only 1.5% you can still calculate your annualized return:

Number of years = 4 months/ 12 months or 1/3.

$$r = \left(\frac{\text{end value}}{\text{start value}}\right)^{\frac{1}{y}} - 1$$

$$r = \left(\frac{1015}{1000}\right)^{\frac{1}{1/3}} - 1$$

$$r = \left(\frac{1015}{1000}\right)^3 - 1 \approx 0.0457 \text{ or about } 4.57\% \quad \text{Notice that this is not the same as 3 times } 1.5\%.$$



## Exponential Equations

Sometimes the unknown that you have to solve for in an equation for is an exponent, or part of an exponent. Most of these problems can be solved by making the base equal on both sides of the equation:

$$5^{x+2} = 125$$

$$5^{x+2} = 5^3$$

$$x + 2 = 3, \text{ so } x = 1$$

Sometimes the base needs to be adjusted on both sides:

$$32^{x-1} = 4^8$$

$$(2^5)^{x-1} = (2^2)^8$$

$$2^{5x-5} = 2^{16}$$

$$5x - 5 = 16$$

$$x = \frac{21}{5}$$

A slightly more challenging problem will require you to remember that  $1/3$  is the same as  $3^{-1}$ , and  $2/3$  would be the same as  $(3/2)^{-1}$ :

$$\left(\frac{2}{3}\right)^{5x+1} = \left(\frac{27}{8}\right)^{x-4}$$

$$\left(\frac{2}{3}\right)^{5x+1} = \left(\left(\frac{3}{2}\right)^3\right)^{x-4}$$

$$\left(\frac{2}{3}\right)^{5x+1} = \left(\left(\frac{2}{3}\right)^{-3}\right)^{x-4}$$

$$\left(\frac{2}{3}\right)^{5x+1} = \left(\frac{2}{3}\right)^{-3x+12}$$

$$5x + 1 = -3x + 12$$

$$x = \frac{11}{8}$$

## Logarithms

Early in the 16<sup>th</sup> century, progress in astronomy was held back by the difficult calculations required to decipher the workings of the solar system. Logarithms provided a much appreciated solution. The usefulness of logarithms was discovered independently by John Napier, a Scottish baron, and Jost Bürgi, a Swiss clockmaker. Logarithms are really just exponents, and we already know that we can multiply two the same numbers raised to different powers by just adding the exponents:  $x^a \times x^b = x^{a+b}$ . In the same way, we can divide quickly and easily by subtracting the exponents:  $x^a \div x^b = x^{a-b}$ . Both Napier and Bürgi reasoned that if they could express any number as some base  $x$  raised to a power, they could multiply and divide easily, which was very important at a time when calculators had not yet been invented. Logarithms quickly became very popular. Henry Briggs proposed using the number 10 as a base, and soon extensive tables were published showing numbers and their base 10 logarithms. These base 10 logs are now available by using the LOG key on your calculator. When we use 10 as a base, the log of 100 is 2, because  $100 = 10^2$ . The log of 1000 is 3, because  $10^3 = 1000$ , and so on. Numbers that are not powers of 10 have logs that are expressed as rounded-off decimal values. For example, the base 10 log of 5 is .69897.... This means that  $10^{0.69897}$  is approximately 5.

Logarithmic functions are the inverse of exponential functions. Let's say you have the number 4, and you put it into the exponential function  $y = 10^x$ . The result of that is  $10^4$  or 10,000. Then you put 10,000 into the logarithmic function  $y = \log x$ . Now you get back to 4. That works the other way too: start with 10,000 and take the log to get 4. Put 4 into the exponential function  $y = 10^x$  to get back to 10,000. Basically logarithmic and exponential functions cancel each other out. As a result of that, something like  $10^{\log 1000}$  is 1000, and  $10^{\log x}$  is just  $x$ .

As we saw earlier, an exponential function only outputs positive values. You can only use positive numbers as an input for a logarithmic function. Logarithms are exponents, and it does not make sense to ask for say, the base 10 logarithm of -100. There is no exponent we can put on the number 10 that will give a value of -100. We can use negative exponents, like  $10^{-2}$ , but this will always produce a positive value.  $10^{-2}$  means  $\frac{1}{10^2}$  which is 0.01. **Logarithms of negative numbers or zero are not defined.**

Let's see how we would do a simple multiplication using base 10 logarithms. We will multiply  $12,879 \times 23,457$ . The log of 12,879 is approximately 4.1098821, and the log of 23,457 is about 4.3702725. This means that we can rewrite our problem as follows:  $12,879 \times 23,457 = 10^{4.1098821} \times 10^{4.3702725}$ . To multiply these two numbers we add their exponents, so we get  $10^{4.1098821 + 4.3702725}$ , which is  $10^{8.4801546} = 302,102,695$ . The real answer for  $12,879 \times 23,457$  is

302,102,703, however it would take a very long time to obtain this answer if you were multiplying without a calculator. The simple addition problem  $4.1098821 + 4.3702725$  can be done quite quickly on paper, with far less risk of error. The logarithmic values themselves used to be available in tables, so people could just look them up. Carrying these tables around and searching for values was somewhat inconvenient, so people designed slide rules. These devices featured movable rows of numbers that were spaced according to the value of their logarithms. Slide rules also allowed the more complex operations of multiplication and division to be reduced to simple addition and subtraction. The results were usually approximate, but good enough for most purposes. When Enrico Fermi created the world's first self-sustaining nuclear reaction, he was monitoring it using his slide rule. As soon as calculators became widely available people discarded their slide rules because they could get faster and more accurate results by using a calculator.

Although logarithms are no longer used to simplify calculations, they are still important in mathematics, science, engineering, and computer applications. The basic rules of using logarithms are not hard to learn, and easy to re-create using simple base 10 logarithm examples.

$10^3 \cdot 10^2 = 10^5$ . We add the exponents, so we add the logs:

$$\log(10^3 \cdot 10^2) = \log(10^3) + \log(10^2) = 3 + 2 = 5$$

General rule:  $\log(ab) = \log a + \log b$

$10^5 \div 10^3 = 10^2$ . We subtract the exponents, so we subtract the logs:

$$\log(10^5 \div 10^3) = \log(10^5) - \log(10^3) = 5 - 3 = 2$$

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

$(10^4)^2 = 10^8$ . We multiply the exponents. The exponent on the outside is always there, so grab it first:

$$\log(10^4)^2 = 2 \cdot \log(10^4) = 2 \cdot 4 = 8$$

General rule:  $\log(a^x) = x \log a$

Notice that there is nothing you can do with  $\log(a + b)$  or  $\log(a - b)$ .

I have a nice calculator, but it only does logarithms in two kinds of bases. The first, as we have already seen, is the common base 10 log, just labeled LOG. The other one is the natural logarithm, ln. Natural logarithms use the special number  $e$  as a base. The number  $e$  is also the base for the natural exponential function  $y = e^x$ . At any given point, this special function

increases at a rate that is exactly equal to its value at that point. The natural logarithm function,  $y = \ln x$ , is the inverse of the natural exponential function. Again these operations cancel each other out:  $\ln(e^x) = x$  and  $e^{\ln x} = x$ .

Logarithms can have any positive number as a base, so you will often see something like  $\log_2 5$ . This means the base 2 log of 5, and if you want to find the value of it you should ask to what power 2 should be raised to create the number 5:

$2^x = 5$  Here  $x$  is the base 2 logarithm of 5. Solve for  $x$  by taking the base 10 logarithm of both sides:

$$\log 2^x = \log 5$$

$$x \log 2 = \log 5$$

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

The reason I chose to take the base 10 logarithms on both sides is that those are easy to get using a calculator. I could also have used the natural logarithms, in which case the answer would be  $x = \frac{\ln 5}{\ln 2}$ . Both answers give the same numerical value for  $x$ , about 2.3219.

In general, if you want the base  $b$  logarithm of the number  $a$ , just call it  $x$ :

$$\log_b a = x$$

$$b^x = a$$

$$x \log b = \log a$$

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

Whenever you are doing problems with exponents, always check if you can change the base. Watch for the numbers  $4 = 2^2$ ,  $8 = 2^3$ ,  $25 = 5^2$  and  $125 = 5^3$ .

### Example

Simplify:  $2 \ln 8 - 3 \ln 4$

First put the exponents up:  $\ln 8^2 - \ln 4^3$

Change it to a division:  $\ln \left( \frac{8^2}{4^3} \right)$

Change the base:  $\ln \left( \frac{(2^3)^2}{(2^2)^3} \right) = \ln \left( \frac{2^6}{2^6} \right) = \ln 1 = 0$

### Example

Solve for x:  $\log_x 2 + \log_x 4 = 3$

$$\log_x (2 \cdot 4) = 3$$

$$\log_x 8 = 3$$

$$8 = x^3$$

$x = 8^{1/3} = 2$ . Although there are complex solutions for x they are not suitable here because these numbers cannot be used as the base of a logarithm.

### Example

Solve for x:  $\frac{5^{x^2}}{25^x} = 125$

This problem is just screaming for you to adjust the base:

$$\frac{5^{x^2}}{(5^2)^x} = 5^3$$

$$\frac{5^{x^2}}{5^{2x}} = 5^3$$

$$5^{x^2-2x} = 5^3$$

$$x^2 - 2x = 3$$

$$x^2 - 2x - 3 = 0$$

$$(x + 1)(x - 3) = 0$$

$x = -1$  or  $x = 3$ . Both solutions work.

### Example

$\log_2 x + 4 = \log_2 32$ , find  $x$ .

To solve this problem, write both sides of the equation as a power of 2:

$$2^{\log_2 x + 4} = 2^{\log_2 32}$$

$$2^{\log_2 x} \cdot 2^4 = 2^{\log_2 32}$$

$$x \cdot 16 = 32$$

$x = 2$  You can easily verify that this is correct:  $\log_2 2 + 4 = \log_2 32$ , which says that  $1 + 4 = 5$ .

## Change of Base Formula

Sometimes it is necessary to change the base of a logarithm. For example, suppose that I need a numerical value for  $\log_3 12$ . My calculator will only do base 10 logs ( $\log$ ) or natural logarithms ( $\ln$ ), so I have to change the base.

First we need to get rid of the base 3 log, so we have to have an equation. Just call the log something, like  $y$ , so you can raise both sides to a power of 3:

$$\log_3 12 = y$$

$$3^{\log_3 12} = 3^y$$

$$12 = 3^y$$

Then take the base 10 log on both sides:

$$\log 12 = \log 3^y$$

$$\log 12 = y \log 3$$

$$y = \frac{\log 12}{\log 3}$$

Practice: change the base of  $\log_a x$  to  $b$  using the method shown above.

The general formula is

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

## Graphing Logarithmic Functions

Taking a logarithm, which is finding an exponent, is the inverse of raising to a power. When two functions are inverses of each other we can graph them by reversing their  $x$  and  $y$  values, which reflects one graph over the line  $y = x$  to create the graph of the inverse function. If you think of the line  $y = x$  as a mirror, one graph is the mirror image of the other.

To graph  $y = \log_2 x$  by hand, you can make a (small) table of values for the inverse function

$y = 2^x$ :

m

<b>X</b>	<b>Y</b>
-2	1/4
-1	1/2
0	1
1	2
2	4

This helps you see that the base 2 log of  $\frac{1}{4}$  is -2, the base 2 log of  $\frac{1}{2}$  is -1, and so on. Choose these convenient values for  $x$ , and mark the corresponding  $y$ -values on your graph.

To graph more complex functions like  $y = \log_3(x - 2) + 1$ , graph the parent function,  $y = \log_3 x$  first by constructing a table for  $y = 3^x$ . Then shift your points 2 units to the right and 1 unit up.

If you try to graph logarithmic functions electronically you may run into a problem because your calculator or graphing software does not graph logarithms with a base other than 10 or  $e$ . You will have to adjust a function like  $y = \log_2 x$  yourself by turning both sides into a power of 2:

$$2^y = 2^{\log_2 x}$$

$$2^y = x$$

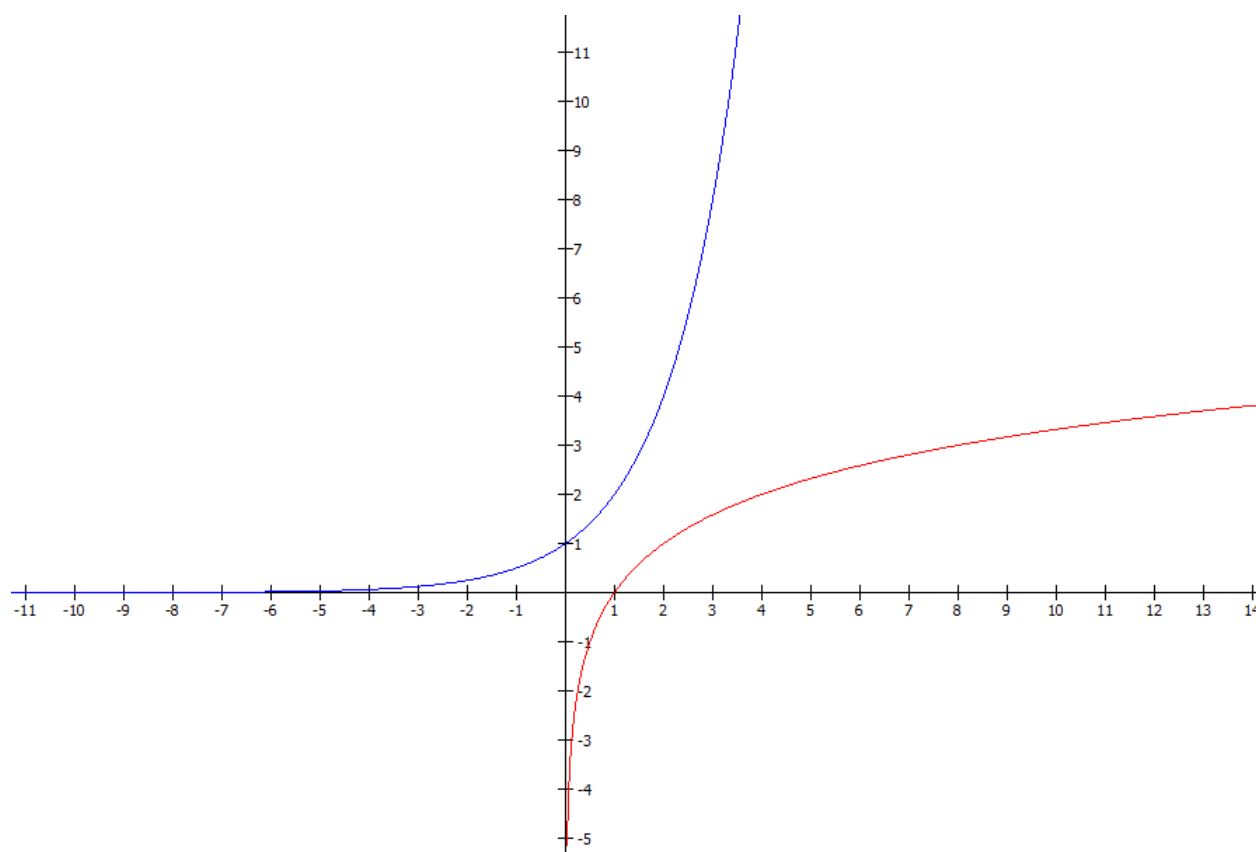
$$\log(2^y) = \log x$$

$$y \log 2 = \log x$$

$$y = \frac{\log x}{\log 2}, \text{ which you would enter as } y = \log(x)/\log(2).$$

It may seem a bit odd that  $2^{\log_2 x}$  is equal to  $x$ . The base 2 logarithmic function is the inverse of the exponential function with base 2. One function “undoes” the operation of the other. If you take any number  $x$  and find the base 2 log, and then put the result of that into the exponential function  $y = 2^x$ , you get your original number  $x$  back. Try that out with a suitable number like 8, which is  $2^3$ . The base 2 logarithm of 8 is 3. Now if you raise 2 to the power  $(\log_2 8)$  you just get 8, the number you started with. That is similar to taking the positive square root of a number and then squaring the result:  $(\sqrt{16})^2 = 16$ . To get rid of the square root you use squaring. To get rid of  $\log_2$  you raise to a power of 2:  $2^{\log_2 8} = 8$ .

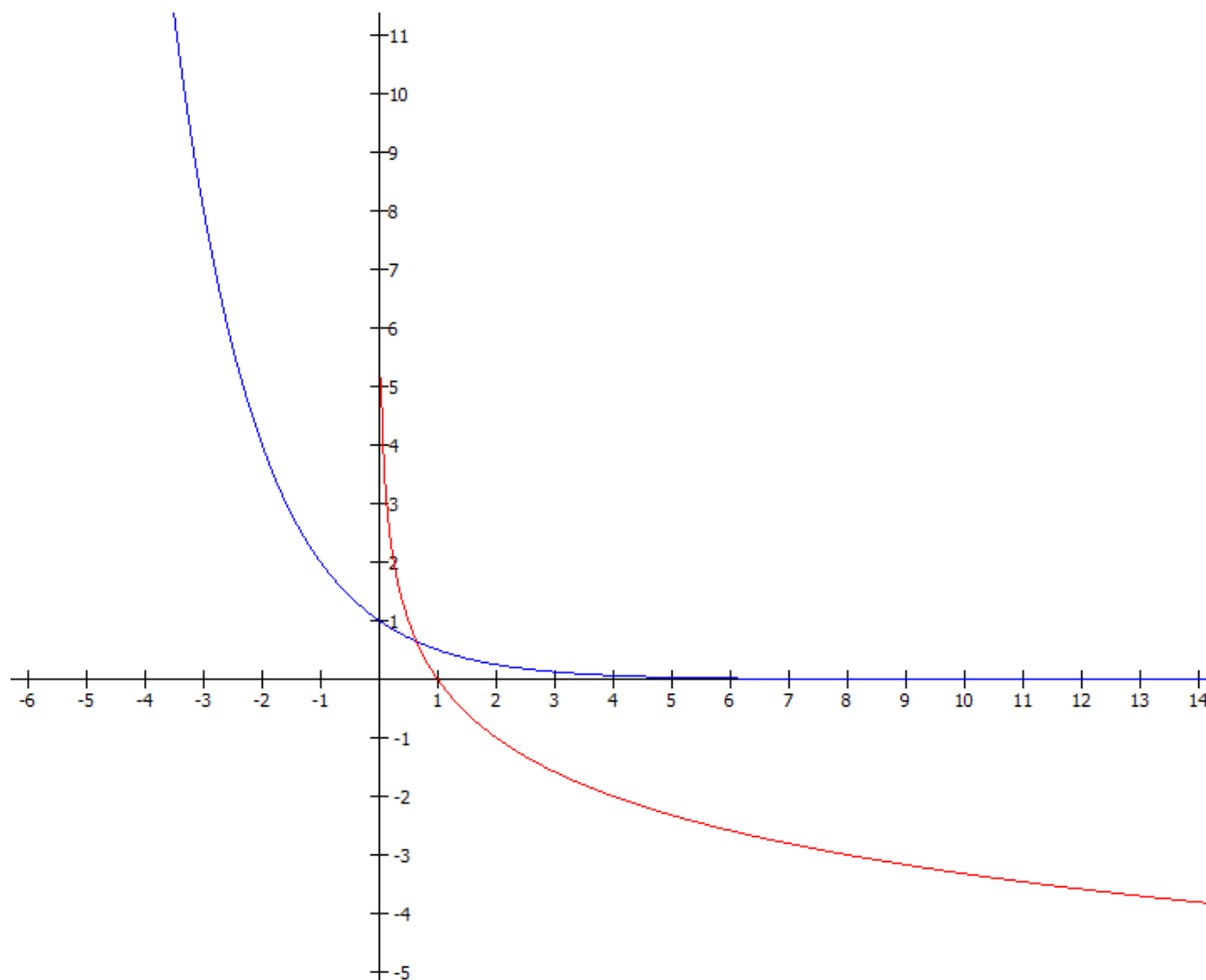
The graph below shows the functions  $y = 2^x$  (blue curve) and  $y = \log_2 x$  (red curve):





Just as a basic exponential function always includes the point (0,1) in its graph, the graph of a basic logarithmic function always passes through the point (1,0).  $\log_2 1 = 0$  since  $2^0 = 1$ . The log of 1 is always 0 for any base. Notice that when  $x$  is smaller than 1, this logarithmic function returns negative values. The base 2 logarithms of numbers less than 1 are negative, which means that we must raise 2 to a negative power to create such numbers. We can find the base 2 logarithms of very small positive numbers, but there is no logarithm of 0. It may look like the graph is touching the  $y$ -axis in the picture, but it actually only gets really, really close. The  $y$ -axis is a vertical asymptote.

The next picture shows the functions  $y = (1/2)^x$  (blue curve), and  $y = \log_{1/2} x$  (red curve). Notice that these graphs still pass through the points (0,1) and (1,0). The  $x$ -axis is a horizontal asymptote for the exponential function, and the  $y$ -axis is a vertical asymptote for the logarithmic function.



Logarithmic functions can be modified and shifted in the same way as exponential functions. We can multiply the function by a constant, to get  $y = c \log_b x$ . Because  $\log_b 1$  is always 0,  $c \log_b x$  is also 0, and the graph still passes through the point  $(1, 0)$ . To shift the graph up or down add  $k$ :  $y = c \log_b(x) + k$ . To shift right or left subtract  $h$ :  $y = c \log_b(x-h) + k$ . Note that if you shift the graph of a logarithmic function left or right, the vertical asymptote also shifts left or right. If you shift the graph  $h$  units to the right (which turns into left if  $h$  is negative), the vertical asymptote will be at the line  $y = h$  instead of at the  $y$ -axis.

## Half-Life

Half-life is a specific type of exponential decay. The half-life function is an exponential function with a base of  $\frac{1}{2}$ . Say we have an amount  $C$  of a substance that has a half-life of one year. The

amount remaining after 1 year is  $C \cdot \left(\frac{1}{2}\right)$ . If two years go by, the original amount  $C$  gets cut in half twice:  $C \cdot \left(\frac{1}{2}\right)^2$ . The remaining amount after 3 years is given by  $C \cdot \left(\frac{1}{2}\right)^3$ , and so on. After  $y$  years, the amount  $A$  that remains is given by the equation  $A = C \left(\frac{1}{2}\right)^y$ . Slightly confusing is what happens if we have a much longer half-life. First let's try a half-life of 10 years, to make things easier to imagine. It now takes 10 years before we have half of the original amount left. How can we take the original equation  $A = C \left(\frac{1}{2}\right)^y$  and make the time take 10 times as long? Let's try  $A = C \left(\frac{1}{2}\right)^{\frac{y}{10}}$ . Now 10 years have to go by to give us  $\frac{1}{2}$  the original amount. That's all there is to it. If the half-life is a large scary number like 360,000 years, you just write the equation as  $A = C \left(\frac{1}{2}\right)^{\frac{y}{360,000}}$ . If your half-life is in different units, like 5 months, you can write the equation like this:  $A = C \left(\frac{1}{2}\right)^{\frac{m}{5}}$  and your answer will be in months.

When you are actually using this equation, the question may ask something like: "Only 15% of the original amount of carbon-14 remains in a sample. Carbon-14 has a half-life of 5730 years. How old is the sample?" Notice that the original amount of C-14 is not given. This really means that you can use whatever original amount you want, so long as the remaining amount is 15% of that. A convenient value to use is 1 [kg, gram, pound, whatever; it really doesn't matter]:

$$0.15 = 1 \cdot \left(\frac{1}{2}\right)^{\frac{y}{5730}}, \text{ so we write that as } 0.15 = (.5)^{y/5730}$$

Now all we have to do is solve for the only unknown  $y$ . This can be done by taking a logarithm on both sides. Either the base 10 logarithm (LOG) or the natural logarithm (LN) will work here. Actually, any logarithm at all would work, but your calculator only does LOG or LN.

$\log 0.15 = \frac{y}{5730} \log(.5)$ . Rearranging gives  $\frac{\log 0.15}{\log 0.5} = \frac{y}{5730}$ , and your calculator should tell you that the sample is about 15,683 years old.

You should make a note of the fact that any exponential decay function has a half-life associated with it. For example, if  $y = 5^{-x}$ , that is an exponential decay function because we can rewrite it as  $y = \frac{1}{5^x}$  according to the rules for exponents, which is the same as  $y = \left(\frac{1}{5}\right)^x$ . The base is smaller than 1. Next, we can convert the base.  $\frac{1}{5}$  is  $\frac{1}{2}$  raised to some power:  $\left(\frac{1}{2}\right)^x = \frac{1}{5}$ . To find  $x$ , take a base 10 logarithm on both sides:  $x \log \frac{1}{2} = \log \frac{1}{5}$ , so  $x = \log \frac{1}{5} / \log 0.5$ . Use a

calculator to get  $x = 2.3219$  (rounded off). Now rewrite the function as  $y = \left(\left(\frac{1}{2}\right)^{2.3219}\right)^x$ , or  $y = \left(\frac{1}{2}\right)^{2.3219x}$ . If  $x$  is the half-life, it will be  $1 \div 2.3219$  or about 0.43.

## Parabolas

For  $y = ax^2 + bx + c$ , the graph is a parabola and the x-coordinate of the vertex is  $-\frac{b}{2a}$

$y = ax^2$  Opens upwards if  $a > 0$ , and down if  $a < 0$ . Vertex at  $(0, 0)$ , focus at  $(0, p)$

$y = a(x - h)^2 + k$  Same, but vertex at  $(h, k)$ , focus at  $(h, p + k)$

$x = ay^2$  Opens to right if  $a > 0$ , and to left if  $a < 0$ . Vertex at  $(0, 0)$ , focus at  $(p, 0)$

$x = a(y - k) + h$  Same, but vertex at  $(h, k)$  and focus at  $(p + h, k)$

$$a = \frac{1}{4p} \text{ so } p = \frac{1}{4a}$$

You can check your work with parabolas by using a graphing program. Use “^” for the exponents.

## Quadratic Equations and Parabolas

When we graph the simple equation  $y = x^2$  we create a parabola. The **vertex** of a parabola is its highest or lowest point. In this case the vertex or lowest point of the parabola is at  $(0, 0)$ . We can adjust the shape of the parabola by using the equation  $y = ax^2$ , where  $a$  is some arbitrary constant. If  $a$  is a positive number larger than 1, the parabola looks “narrow”, and if  $a$  is a very small positive number the parabola flattens out. If  $a$  is negative the parabola flips over so that the vertex is at the top.

Adding a number to the equation shifts the parabola up, or down if the number is negative:

$y = ax^2 + k$ . This can also be written as  $y - k = ax^2$ . Just as subtracting  $k$  from  $y$  shifts the parabola in the positive  $y$  direction, subtracting  $h$  from  $x$  will shift the parabola in the positive  $x$  direction:

$y - k = a(x - h)^2$ , or  $y = a(x - h)^2 + k$ . The vertex of this parabola is located at the point  $(h, k)$

All quadratic equations graph as parabolas. To change  $y = ax^2 + bx + c$  to the more useful form above where you can read off the position of the vertex, just complete the square (see

“Completing the Square and the Quadratic Formula”).

### Example

Find the vertex of the parabola  $y = 5x^2 + 20x - 12$

$y = 5x^2 + 20x - 12$  Temporarily remove the 5 from in front of the x:

$\frac{y}{5} = x^2 + 4x - \frac{12}{5}$  Complete the square, making sure to add 4 to *both* sides:

$$\frac{y}{5} + 4 = x^2 + 4x + 4 - \frac{12}{5}$$

$$\frac{y}{5} + 4 = (x + 2)^2 - \frac{12}{5}$$

$\frac{y}{5} = (x + 2)^2 - \frac{12}{5} - 4$  Now put the 5 back where it belongs by multiplying everything by 5:

$$y = 5(x + 2)^2 - 12 - 20$$

$$y = 5(x + 2)^2 - 32$$

The vertex is at (-2, -32)

If you are working with parabolas a lot, you may get tired of completing the square each time. You can just do it once with the general equation of a parabola, and then use the result of that each time. Here is how it works:

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

$$\frac{y}{a} = x^2 + \frac{b}{a}x + \frac{c}{a}$$

$$\frac{y}{a} + \left(\frac{b}{2a}\right)^2 = x^2 + \frac{b}{a}x + \left(\frac{b}{2a}\right)^2 + \frac{c}{a}$$

$$\frac{y}{a} = \left(x + \frac{b}{2a}\right)^2 + \frac{c}{a} - \left(\frac{b}{2a}\right)^2$$

$$y = a\left(x + \frac{b}{2a}\right)^2 + c - \frac{b^2}{4a}$$

If you compare this to  $y = a(x - h)^2 + k$ , you will see that  $-\frac{b}{2a}$  corresponds to  $h$ , the  $x$ -coordinate of the vertex. This means that we can take a shortcut. For example, the parabola  $y = 5x^2 + 20x - 12$  will have a vertex with  $x$ -coordinate  $-\frac{b}{2a}$ , or  $-\frac{20}{2 \cdot 5} = -2$ .

If we know that the  $x$  coordinate of the vertex is  $-2$ , the fastest way to find the  $y$ -coordinate is by substituting  $x = -2$  in the equation of the parabola.  $y = 5(-2)^2 - 40 - 12 = -32$ . The vertex is located at  $(-2, -32)$ .

Now we can convert  $y = 5x^2 + 20x - 12$  to the vertex form  $y = a(x - h)^2 + k$  faster than we did before.  $h$  and  $k$  are the coordinates of the vertex, which we just calculated, and  $a$  is the coefficient of the first term,  $5$ . The vertex form is  $y = 5(x + 2)^2 - 32$ .

If you know the vertex and one point on a parabola, you can get the equation of the parabola. If you are modeling data that are given to you, look to see if the vertex is obvious from the data. (Parabolas are symmetrical, so look for similar  $y$  values on both sides of the point that you suspect could be the vertex. Sometimes the data need to be modeled by one half of a parabola, so the vertex is the starting point). When we know the vertex, we use the form  $y = a(x - h)^2 + k$ . Simply fill in the values for  $h$  and  $k$ . For example, if the vertex is at  $(2, 3)$ , the equation will look like this:  $y = a(x - 2)^2 + 3$ . If you also have another point  $(x, y)$ , you can put that into this equation to find  $a$ .

### Example

The axis of symmetry of the parabola  $y = 3x^2 + bx + 4$  is the line  $x = 1$ . Find the value of  $b$ .

Here we know that the vertex is located at some point  $(1, k)$ . Let's fill in what we know:

$$y = a(x - 1)^2 + k$$

$$y = 3(x - 1)^2 + k$$

Now, how does that relate to  $y = 3x^2 + bx + 4$ ? Well, let's just multiply  $(x - 1)^2$  out:

$$y = 3(x^2 - 2x + 1) + k$$

$$y = 3x^2 - 6x + 3 + k$$

Because  $k$  is just a number, we can conclude that  $b$  must equal  $-6$ .

On the other hand, if you know that the vertex always occurs at  $x = -\frac{b}{2a}$ , you can simply fill in the values:

$$x = -\frac{b}{2 \cdot 3}$$

$$1 = -\frac{b}{6}$$

$$b = 1 \cdot -6 = -6$$

If you do not know where the vertex is but you have at least three points on the parabola, you can also find the equation. In this case you would use the form  $y = ax^2 + bx + c$ . Filling in three separate points will give you three equations where the only unknowns are  $a$ ,  $b$  and  $c$ , which you can then solve for. It is particularly helpful if one of your points has an  $x$ -coordinate of 0, since that gives you the value of  $c$  directly. For three points  $(x_1, y_1)$ ,  $(x_2, y_2)$  and  $(x_3, y_3)$  we have:

$$ax_1^2 + bx_1 + c = y_1$$

$$ax_2^2 + bx_2 + c = y_2$$

$$ax_3^2 + bx_3 + c = y_3$$

Here the three unknowns,  $a$ ,  $b$  and  $c$ , can be found by solving this system of three linear equations. This is most easily done by solving the corresponding matrix (see Matrices).

If we are provided with the coordinates of two points on a parabola, say  $(x_1, y_1)$  and  $(x_2, y_2)$ , we can find a set of possible equations for that parabola as follows:

$$y_1 = ax_1^2 + bx_1 + c$$

$$y_2 = ax_2^2 + bx_2 + c$$

Subtract these two equations to get:

$$y_1 - y_2 = a(x_1^2 - x_2^2) + b(x_1 - x_2)$$

There are many possibilities for  $a$  and  $b$ , so there are an unlimited number of equations for the parabola that passes through the given two points. To find the simplest one, take  $a = 1$ . Solve for  $b$ , and then for  $c$ .



### Example

Find the equation of the parabola that passes through the point (1,-2) and has its axis of symmetry at  $x = 3$ . The y-intercept is 8.

This clever problem disguises the fact that it is providing you with 2 points on the parabola: (1,-2) and (0,8). The third piece of information is that the x-coordinate of the vertex is 3. That gives us a hint that we should be using the vertex form,  $y = a(x - h)^2 + k$ , with  $h = 3$ :

$$y = a(x - 3)^2 + k$$

Remember that  $k$  is the y-coordinate of the vertex, not the y-intercept! Use your two points to create two equations with two unknowns:

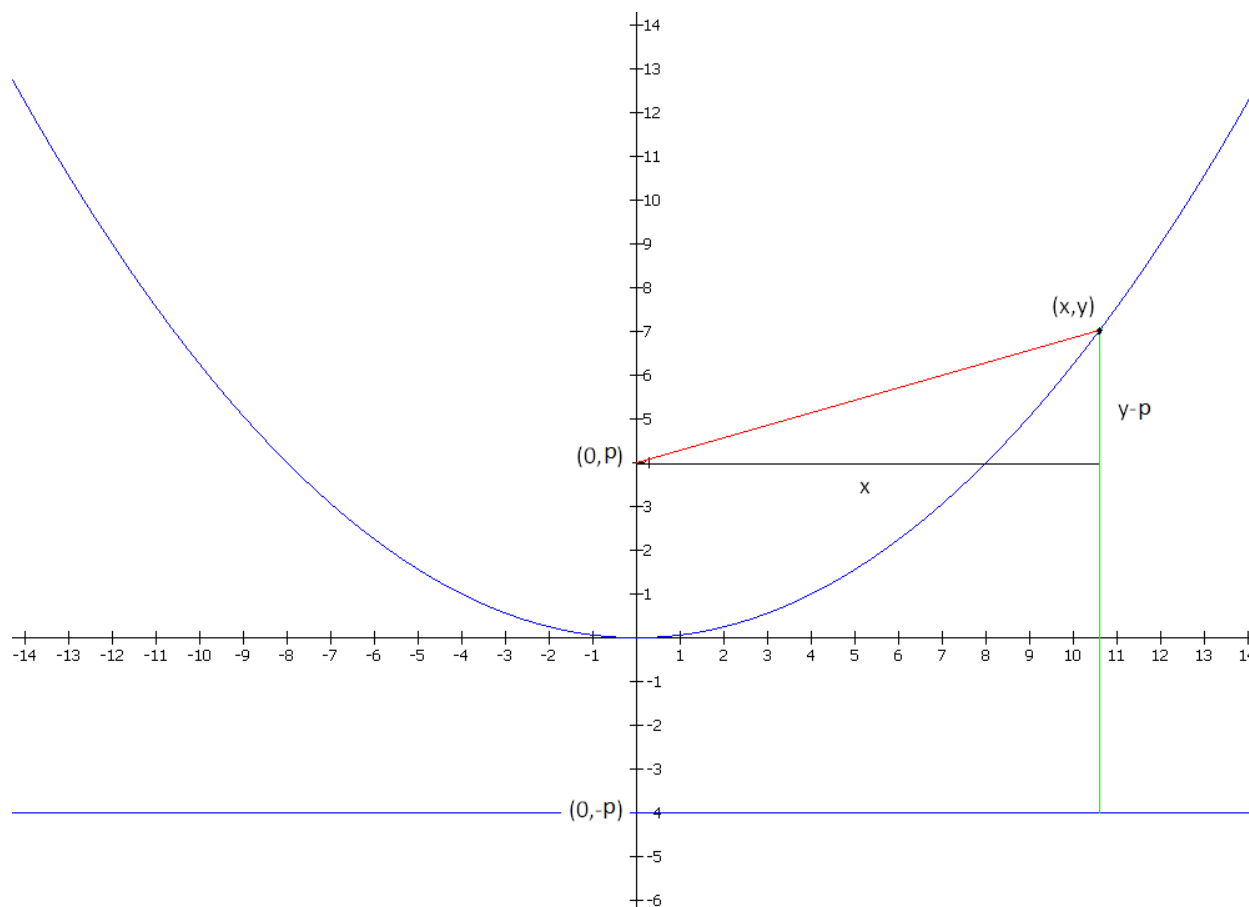
$$-2 = a(1 - 3)^2 + k$$

$$8 = a(0 - 3)^2 + k$$

Solve this system of equations to get  $a = 2$  and  $k = -10$ . The equation of the parabola in standard form is  $y = 2x^2 - 12x + 8$ .

## Basic Graphs

Parabolas are a type of conic section. If you take a cone and cut it in various ways, you can get a circle, an ellipse, a parabola, or a hyperbola. These conic sections have one or two special points called **foci**, except for the circle for which the focus is really the center. All parabolas have a single focus. This is the point that all light falling on a 3 dimensional parabolic mirror would be focused on. If you put a parabolic mirror in the sun you can boil water at the focus ( <http://www.youtube.com/watch?v=0xfTl3Ugyjk&feature=related> ). For other conic sections the distance to the focus is called  $c$ , but for a parabola this is a bit confusing because quadratic functions already have an end term called  $c$  which is not at all related to the focus. The distance from the vertex of a parabola to the focus is therefore called  $p$  by many textbooks, although others use  $c$ . Here is a picture of a parabola determined by the equation  $y = ax^2$  where  $a$  is a positive number. The distance from the vertex to the focus is  $p$  (or  $c$ , depending on your textbook), and the focus is located at  $(0, p)$ . In this picture  $p$  is 4.



The **directrix** is a line that runs perpendicular to the axis of symmetry of the parabola, at a distance of  $p$  below the vertex (or above it if the parabola opens down). In the picture above, the directrix is the blue line underneath the parabola. By definition, a parabola is the collection of all points  $(x,y)$  such that the distance from the focus to  $(x,y)$  is the same as the distance from the point  $(x,y)$  to the directrix. That means that the red line in the picture is the same length as the green line. Then of course the squares of their lengths are also equal.

It is not hard to find the length of the red line. By the Pythagorean theorem, the square of its length is equal to  $x^2 + (y - p)^2$ . The length of the entire green line is  $y + p$ , so the square of its length is  $(y + p)^2$ . Now we can say that  $x^2 + (y - p)^2 = (y + p)^2$

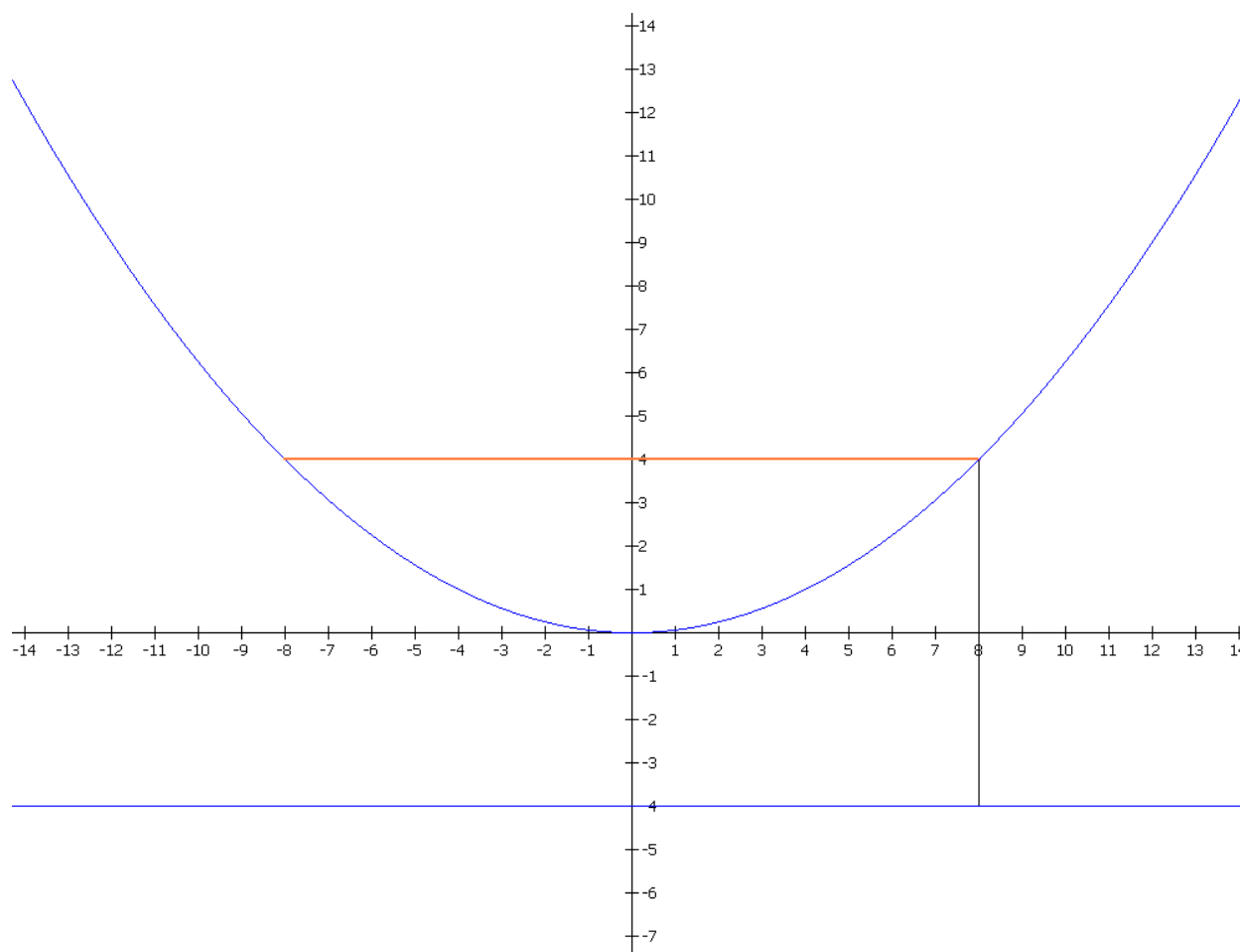
Multiplying out, you get  $x^2 + y^2 - 2py + p^2 = y^2 + 2py + p^2$ . A lot of that cancels out to give  $4py = x^2$ . We could also write that as  $y = \frac{1}{4p} x^2$ .

Now we have two equations for this parabola:  $y = ax^2$  and  $y = \frac{1}{4p} x^2$ .

This means that  $a = \frac{1}{4p}$ , so if we know  $a$  we can find  $p$ .  $4ap = 1$  so  $p = \frac{1}{4a}$ . Since many people dislike fractions, you may just see the parabola written as  $4py = x^2$ . You can use either format to find  $p$ ; just make sure that your equation is written correctly. If you have a parabola described by the equation  $y = \frac{1}{8}x^2$ , and you want to use the format  $4py = x^2$  to find  $p$ , rewrite the equation as  $8y = x^2$ . Now you can see that  $p$  is 2.

For the parabola  $y = x^2$ , or  $1y = 1x^2$ , the distance to the focus,  $p$ , must be  $\frac{1}{4}$ . For this basic parabola, the focus is located at  $(0, \frac{1}{4})$ . To get the focus at a nicer distance, I used  $y = \frac{1}{16}x^2$  to make the pictures for this section. You can leave the equation like that or write it as  $16y = x^2$  to get rid of the fraction. Either way,  $p = 4$  which puts the focus at  $(0, 4)$ . It also flattens the parabola, and the flatter the parabola is the more the focus moves outward. In the video linked above, the focus was quite far away from the mirror. When you first learn about parabolas they usually have equations like  $y = x^2$  or  $y = 3x^2$ , but once you start looking at the focus you'll want "flatter" parabolas that have some fraction in front of  $x^2$ , or alternatively an integer in front of  $y$ .

The next picture shows the **latus rectum**, which is a line through the focus that runs parallel to the directrix. It is marked in orange. Its main function seems to be to provide additional problems for students to solve. However, it does help you graph a parabola by giving an indication of how wide it should be at the area of the focus. If this item is in your course you will be asked to find the endpoints of the latus rectum, which is actually quite easy if you know where the focus is.



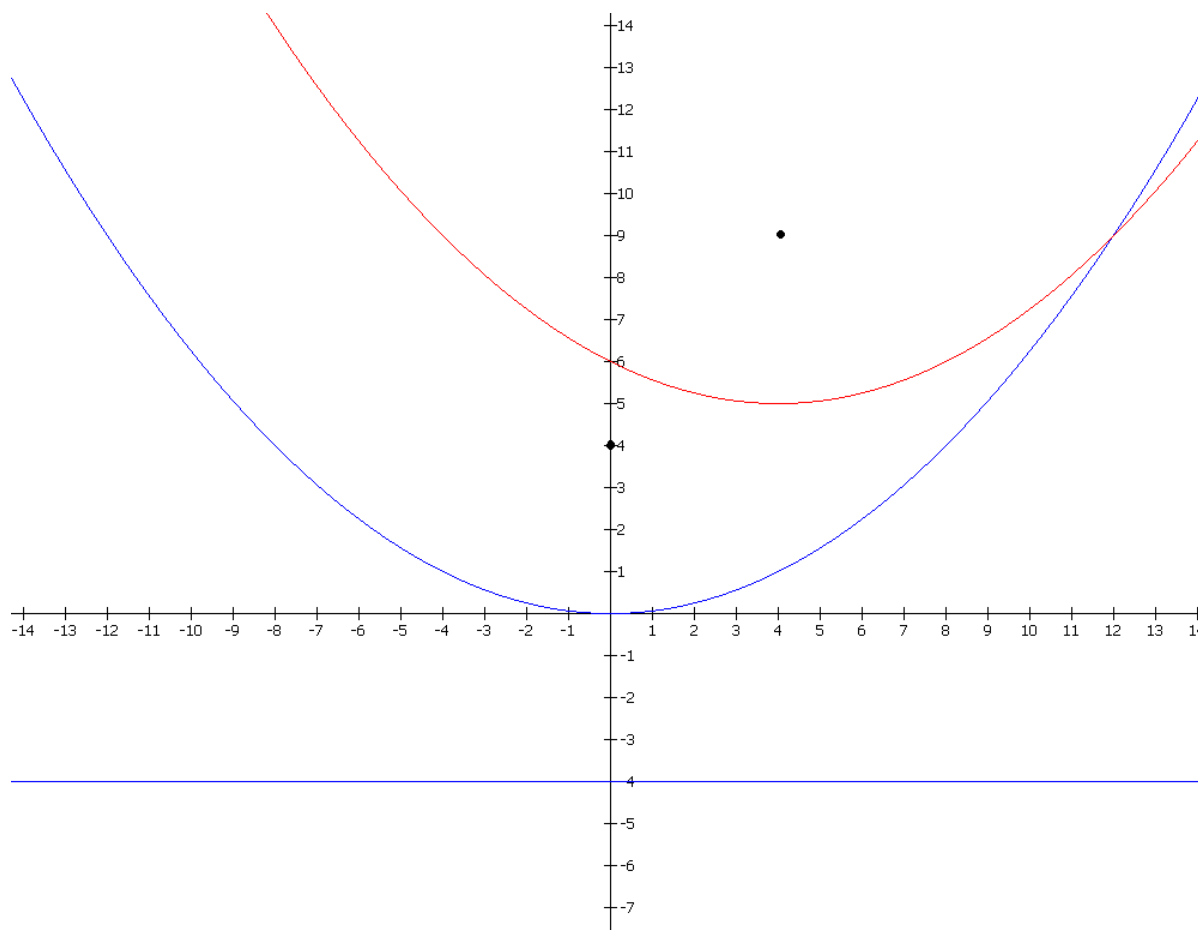
For every point on the parabola, the distance to the focus is equal to the distance to the directrix. So, by definition, half the length of the orange line is equal to the length of the black line.

Because the distance from the vertex to the focus is  $p$ , the length of the black line is  $2p$ . This means that the orange line has a total length of  $4p$ . The focus is at  $(0,p)$  for this standard parabola, so the endpoints of the latus rectum must be located at  $(-2p,p)$  and  $(2p,p)$ .

## Shifted Parabolas

When we start moving the vertex, the focus [and the latus rectum] move accordingly. Let's change  $y = \frac{1}{4p} x^2$  to  $(y - k) = \frac{1}{4p} (x - h)^2$ . You can also write this as  $4p(y - k) = (x - h)^2$ . Now the vertex is at  $(h, k)$ .

The next picture shows that when we move the vertex to  $(h,k)$  from  $(0,0)$ , the focus (marked by a black dot) moves from  $(0,p)$  to  $(h,p+k)$ .

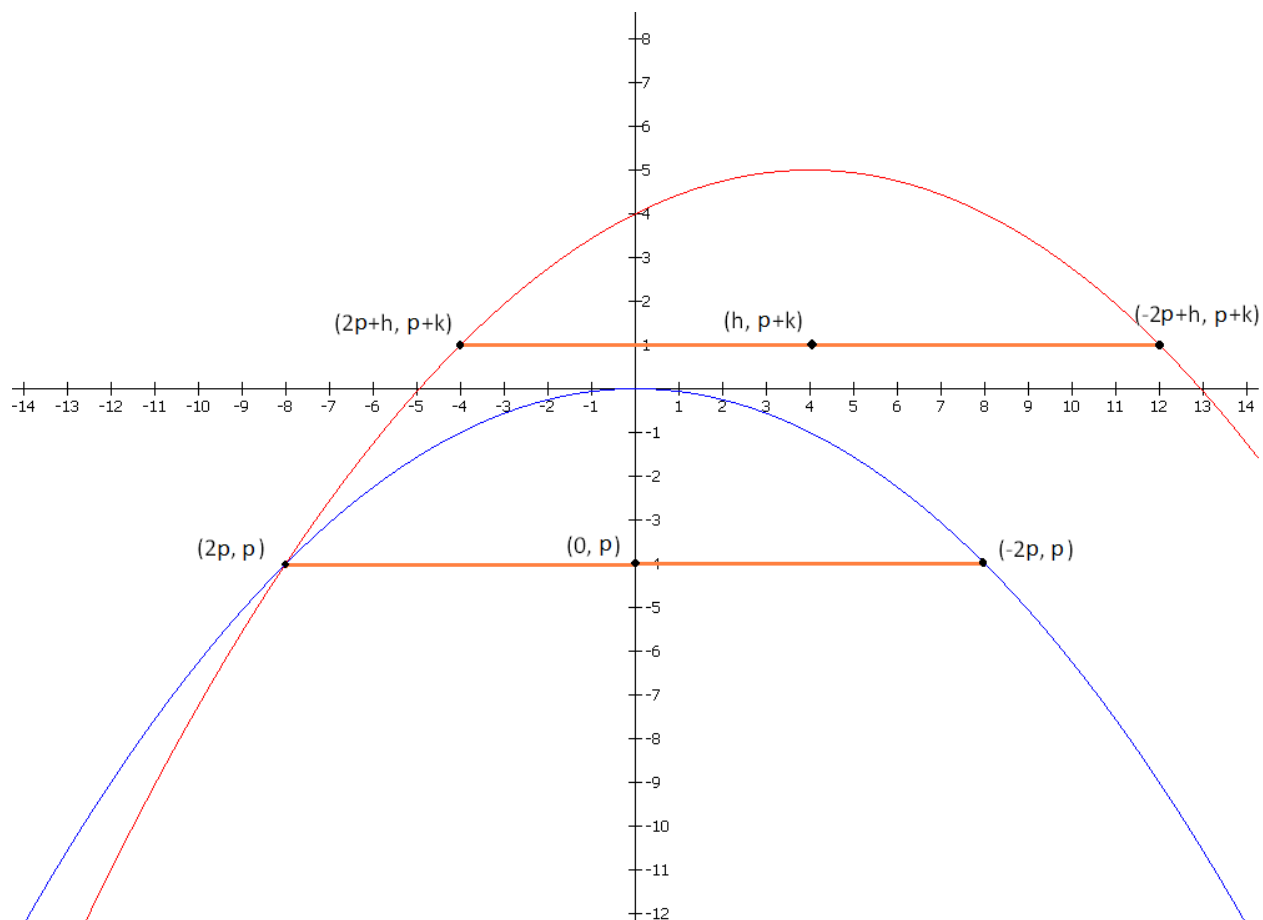


For this example I used  $y = \frac{1}{16}x^2$  and  $y = \frac{1}{16}(x-4)^2 + 5$ . So,  $h = 4$  and  $k = 5$ , which moves the focus from  $(0, 4)$  to  $(4, 9)$ . [Note that the points of the latus rectum also move  $h$  units to the right and  $k$  units up (or to the left and down if  $h$  and  $k$  are negative). They end up at  $(-2p+h, a+k)$  and  $(2p+h, p+k)$  or in this case  $(-4, 9)$  and  $(12, 9)$ .]

Now, how can we translate all of this to a parabola that opens downward? The basic equation of a parabola is  $y = ax^2$ . The parabola opens up if  $a$  is positive, and down when  $a$  is negative.

Since  $a = \frac{1}{4p}$ ,  $p$  will need to be a negative number. The distance to the focus is  $|p|$  since

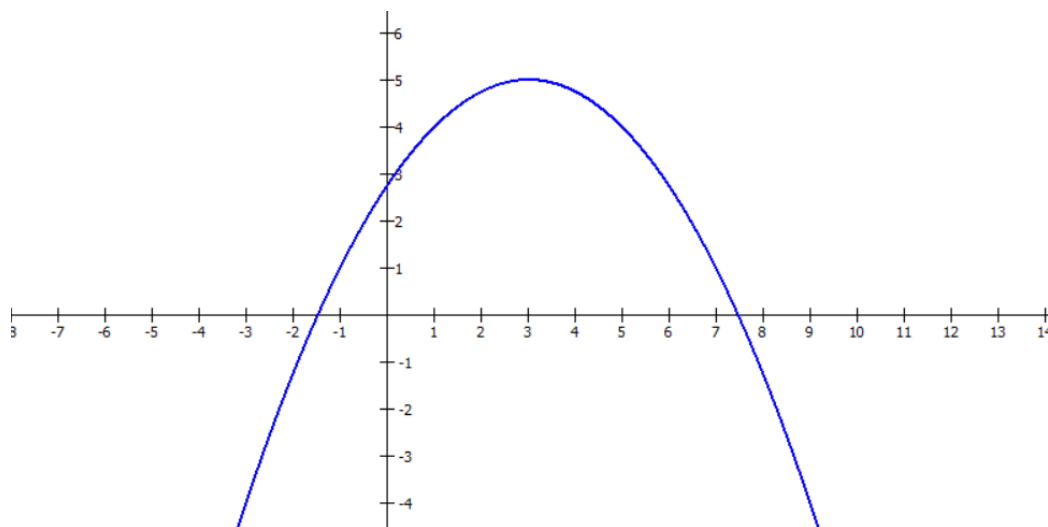
distance is always positive. The image below shows  $y = -\frac{1}{16}x^2$  and  $y = -\frac{1}{16}(x-4)^2 + 5$ .  $p$  is  $-4$ , and the vertex of the parabola shifts from  $(0,0)$  to  $(4,5)$ :



### Example

The focus of a parabola that opens down is located at the point  $(3, 4)$ . A point on the parabola is  $(7, 1)$ . Find the equation of this parabola.

A quick sketch of this parabola would show that it should look something like this:



Since the focus is at  $(3, 4)$ , the vertex will be located at  $(3, k)$ . That provides some information to put into the general equation  $y = \frac{1}{4p}(x - h)^2 + k$ , since we know that  $h = 3$ . We also have a point,  $(7, 1)$ , that we can use for  $x$  and  $y$ . However, that still leaves both  $p$  and  $k$  as unknowns. Because the parabola opens down, we know that  $p$  has to be a negative number. We can also see that there is a relationship here between  $p$  and  $k$ . The  $y$ -coordinate of vertex is  $k$ , and the  $y$ -coordinate of the focus is 4. The distance between the focus and the vertex is  $|p|$ , by definition. We can see from the image above that  $k = 4 + |p|$ . Because it is already known that  $p$  will be a negative number, the absolute value sign will change the value:  $k = 4 - p$ . Now that we have an expression for  $k$ ,  $p$  is the only unknown:

$$y = \frac{1}{4p}(x - h)^2 + k \quad \text{Remove the awkward fraction to make it easier to solve:}$$

$$4p(y - k) = (x - h)^2$$

$$4p(1 - (4 - p)) = (7 - 3)^2$$

$$4p(1 - 4 + p) = (4)^2$$

$$4p(-3 + p) = 16$$

$$-12p + 4p^2 - 16 = 0$$

$$4p^2 - 12p - 16 = 0$$

$$p^2 - 3p - 4 = 0$$

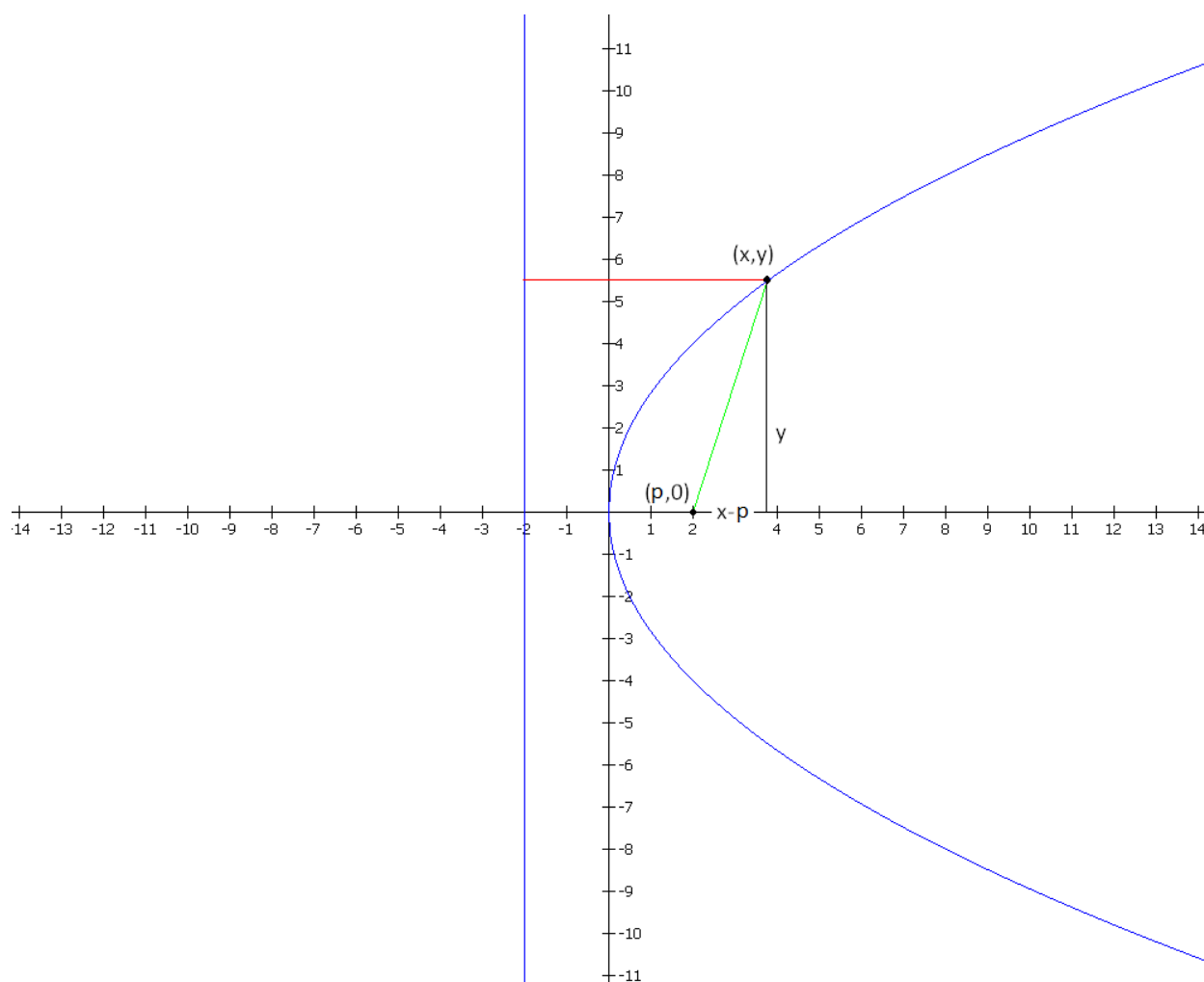
$$(p - 4)(p + 1) = 0$$

$$p = 4 \text{ or } p = -1$$

The only solution here is  $p = -1$ , since  $p$  must be negative. That means that  $k = 4 - -1 = 5$ . The equation of the parabola,  $y = \frac{1}{4p}(x - h)^2 + k$  is  $y = -\frac{1}{4}(x - 3)^2 + 5$ , or  $y - 5 = -\frac{1}{4}(x - 3)^2$ .

## Side-Opening Parabolas

Next, we will look at parabolas that open to the side. All of these parabolas have equations of the form  $x = ay^2$ . If  $a$  is a positive number the parabola opens to the right, as shown in this picture:





The focus of this parabola is located at  $(p, 0)$ , which happens to be the point  $(2, 0)$  in this case. Again we see that by definition, the square of the length of the red line equals the square of the length of the green line. The length of the red line is  $x + p$ . We can find the length of the green line by using the Pythagorean Theorem. The square of the length is  $(x - p)^2 + y^2$ . Therefore,  $(x - p)^2 + y^2 = (x + p)^2$ .

$x^2 - 2px + p^2 + y^2 = x^2 + 2px + p^2$ , so after we cancel terms we get

$y^2 = 4px$ , which is the same as  $x = \frac{1}{4p} y^2$ .

I made this parabola using the equation  $x = \frac{1}{8} y^2$ . This means that  $p = 2$  and the focus ends up at  $(2, 0)$ .

[The length of the latus rectum is still  $4p$ , which puts the endpoints at  $(p, 2p)$ , and  $(p, -2p)$ .]

Again, we can move the vertex of the parabola to  $(h, k)$ , using the equation  $(x - h) = \frac{1}{4p} (y - k)^2$ .

That moves every point over. The focus ends up at  $(p+h, k)$

If  $a$  is a negative number the parabola opens to the left. The number  $p$  will be negative, and the distance to the focus is  $|p|$ .

## Circles and Ellipses

$(x - h)^2 + (y - k)^2 = r^2$  A circle with radius  $r$  and center at  $(h, k)$

$\frac{(x-h)^2}{a^2} + \frac{(y-k)^2}{b^2} = 1, \quad a > b$  An ellipse with a horizontal major axis and center at  $(h, k)$ .

$\frac{(x-h)^2}{b^2} + \frac{(y-k)^2}{a^2} = 1, \quad a > b$  An ellipse with a vertical major axis and center at  $(h, k)$

$c$  is the distance from the center of the ellipse to the foci.  $c^2 = a^2 - b^2$ .

Foci:  $(-c, 0)$  and  $(c, 0)$  or  $(0, -c)$  and  $(0, c)$

Vertices:  $(-a, 0)$  and  $(a, 0)$  or  $(0, a)$  and  $(0, -a)$

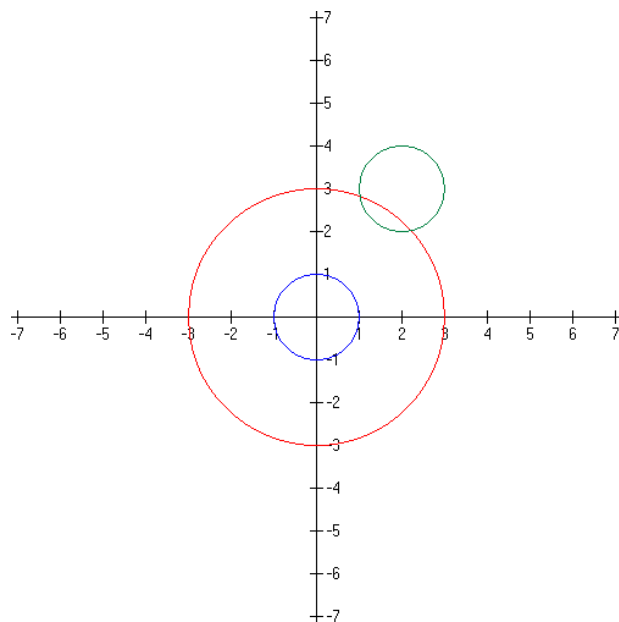
Covertices:  $(0, b)$  and  $(0, -b)$  or  $(-b, 0)$  and  $(b, 0)$

Shift these markers as needed by adding  $(h, k)$  to the coordinates.

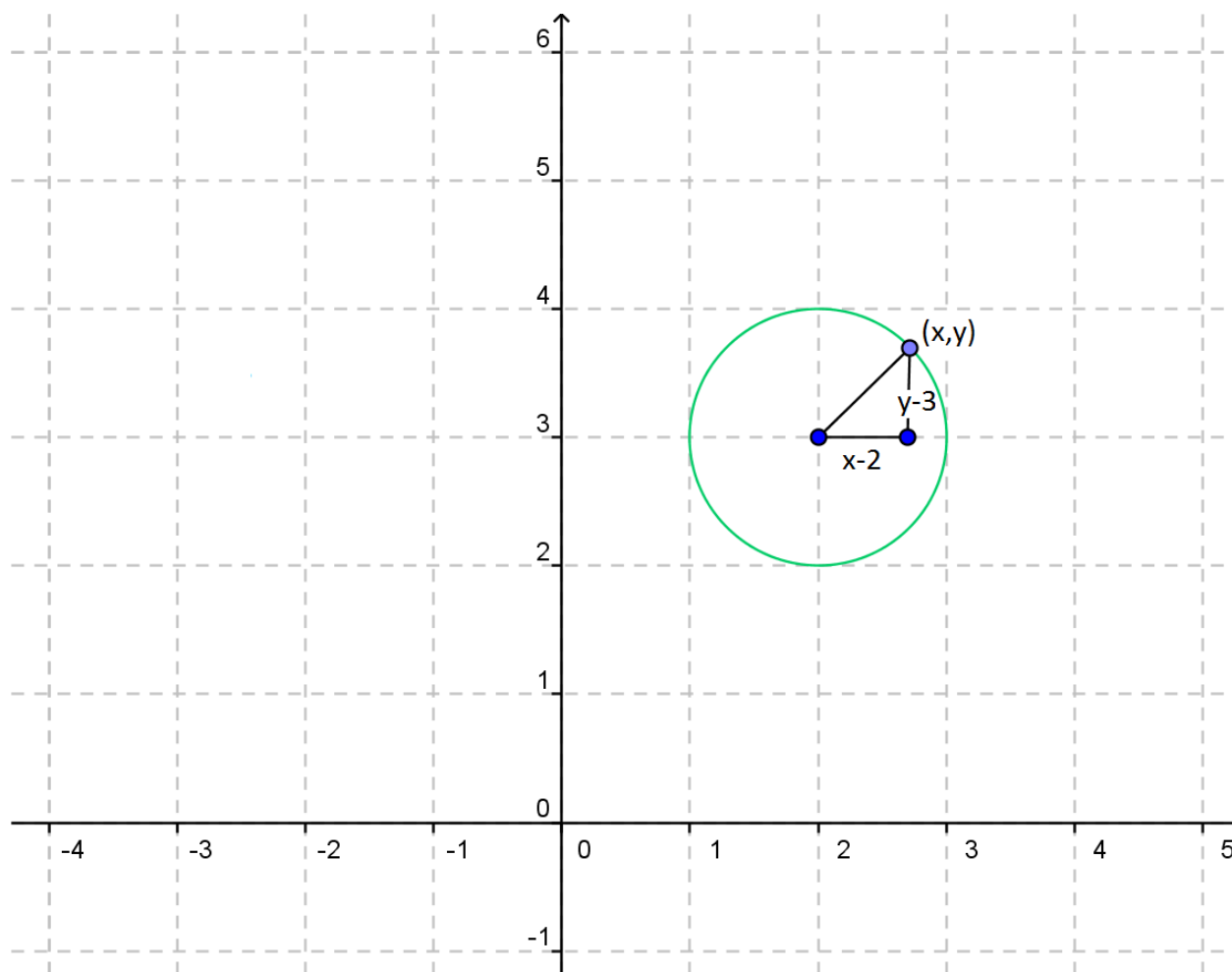
The area of an ellipse with a major half-axis of length  $a$  and a minor half-axis of length  $b$  is  $\pi ab$

Let's set up a Cartesian coordinate system and draw it magnified, so that each unit is very large. Now draw a circle, centered at the origin, with a radius of 1. If your drawing is big enough, you can pick any point  $(x, y)$  on your circle and easily see that  $x^2 + y^2 = 1$ . This just happens because the distances  $x$  and  $y$  are perpendicular to each other, creating a right triangle with a hypotenuse of length 1. We can change the equation for this circle to one that makes it easier to graph using a calculator or computer.  $y^2 = 1 - x^2$ , which means that  $y = \pm \sqrt{1 - x^2}$ . Graph this as two separate functions to create a circle.

If we want to make the circle bigger, we could change our equation to  $x^2 + y^2 = 9$ . When you graph this equation, you will see that it gives us a circle with a radius of 3. The graph is the red circle in the picture below. The general equation for a circle is  $x^2 + y^2 = r^2$ , where  $r$  is the radius. Using this formula, or the equivalent  $y = \pm \sqrt{r^2 - x^2}$ , we can create a circle as big or small as we want.



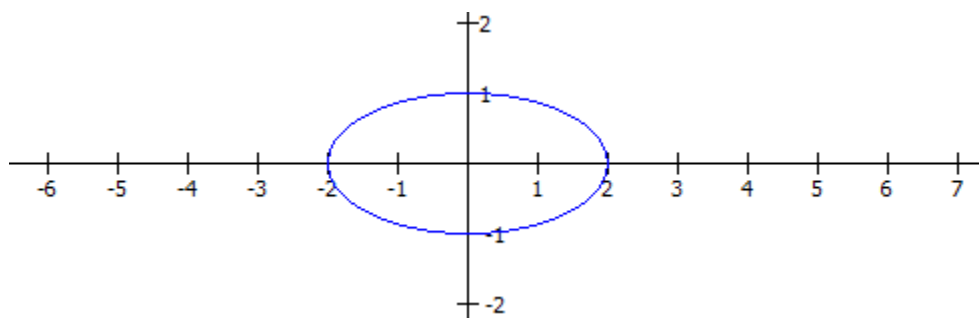
It is also possible to move the circle away from the center of our coordinate system. The green circle in the picture above has a radius of 1, and the center is located at the point (2,3). At first it may seem that the simple relationship  $x^2 + y^2 = 1$  no longer holds. However, we can account for the fact that the center has been moved. Pick a point (x,y) on the edge of the circle as shown in the picture below. The actual x distance that we use to calculate the radius is now  $x - 2$ , and the y distance is  $y - 3$ . We can say that  $(x - 2)^2 + (y - 3)^2 = 1$ .



If we want the center to be at the point  $(h, k)$ , we can write  $(x - h)^2 + (y - k)^2 = 1$ . For a circle with radius  $r$ , we would write  $(x - h)^2 + (y - k)^2 = r^2$ . To graph the circle we would rearrange that to  $(y - k)^2 = r^2 - (x - h)^2$ , and then to  $y - k = \pm \sqrt{r^2 - (x - h)^2}$ , or  $y = \pm \sqrt{r^2 - (x - h)^2} + k$ . For example, a circle with a radius of 1 centered at  $(2, 3)$  would be drawn using the equation  $y = \pm \sqrt{1 - (x - 2)^2} + 3$ .

In the previous section we enlarged our circle with radius 1 by changing its equation to  $x^2 + y^2 = 9$ . Dividing both sides by 9, we can see that this equation could also be written as  $\frac{x^2}{9} + \frac{y^2}{9} = 1$ , or alternatively:  $\left(\frac{x}{3}\right)^2 + \left(\frac{y}{3}\right)^2 = 1$ . Dividing both  $x$  and  $y$  by 3 actually caused the little circle to stretch so that its radius is 3 times larger than it was before. Now let's stretch our

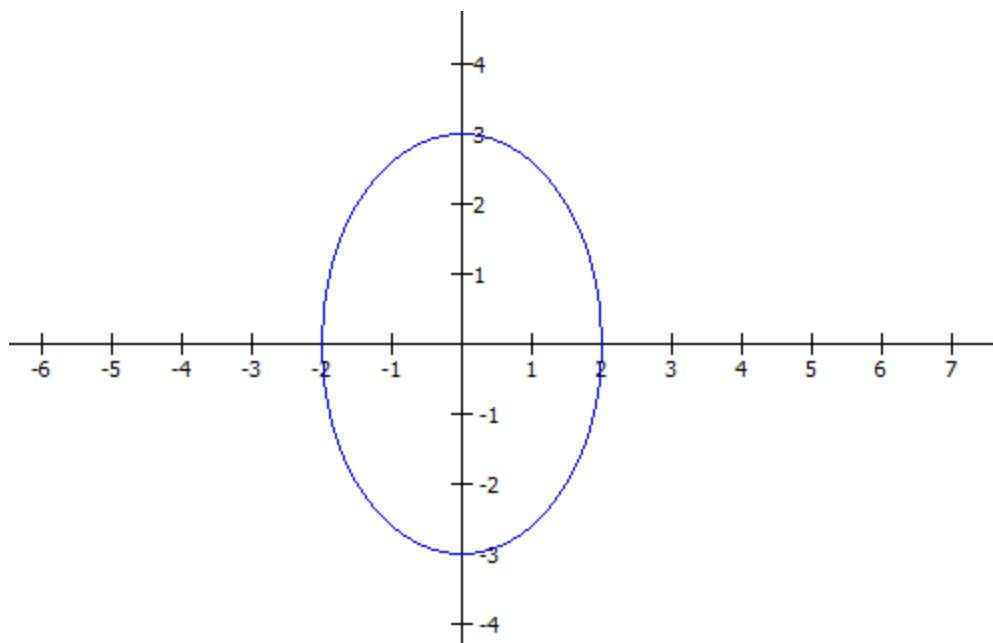
first circle just on the sides, so that it becomes wider only in the direction of the x-axis. This is surprisingly easy to do mathematically. The equation of the first circle was  $x^2 + y^2 = 1$ . Now we write:  $\left(\frac{x}{2}\right)^2 + y^2 = 1$ . What is the effect of this change? Notice that the “height” of our circle doesn’t change. When  $x = 0$ ,  $y$  is still 1, just as it was before. However, we now need a bigger  $x$  to create the same value of  $y$ . At  $y = 0$ ,  $x$  is now 2 instead of 1. The “radius” of our circle has been stretched to a length of 2 units along the x-axis. It looks like an ellipse. In fact, it *is* an ellipse.



To actually graph the equation  $\left(\frac{x}{2}\right)^2 + y^2 = 1$  you need to solve for  $y$ :  $y = \pm \sqrt{1 - \frac{x^2}{4}}$ . Graph the positive and negative parts of the equation separately.

Every ellipse has a shortest “radius” and a longest “radius” at right angles to each other. By convention, the length of the longest radius is called “ $a$ ” and the length of the shortest radius is called “ $b$ ”. Because those are not real radii we instead talk about the longest axis, which is the **major axis**, and the shortest axis, which is called the **minor axis**. The length of the major axis is  $2a$ , and the length of the minor axis is  $2b$ . It is not necessary for the major axis of the ellipse to lie along the x-axis. Let’s stretch the ellipse in the y-direction. We can do this by writing:

$\left(\frac{x}{2}\right)^2 + \left(\frac{y}{3}\right)^2 = 1$ . This will cause the radius along the y-axis to be 3 units long, because when  $x = 0$ ,  $y = 3$ . “ $a$ ” now refers to the length of the radius along the y-axis, since it is the longest. The major axis of the ellipse lies along the y-axis, and it has a length of 6 units:



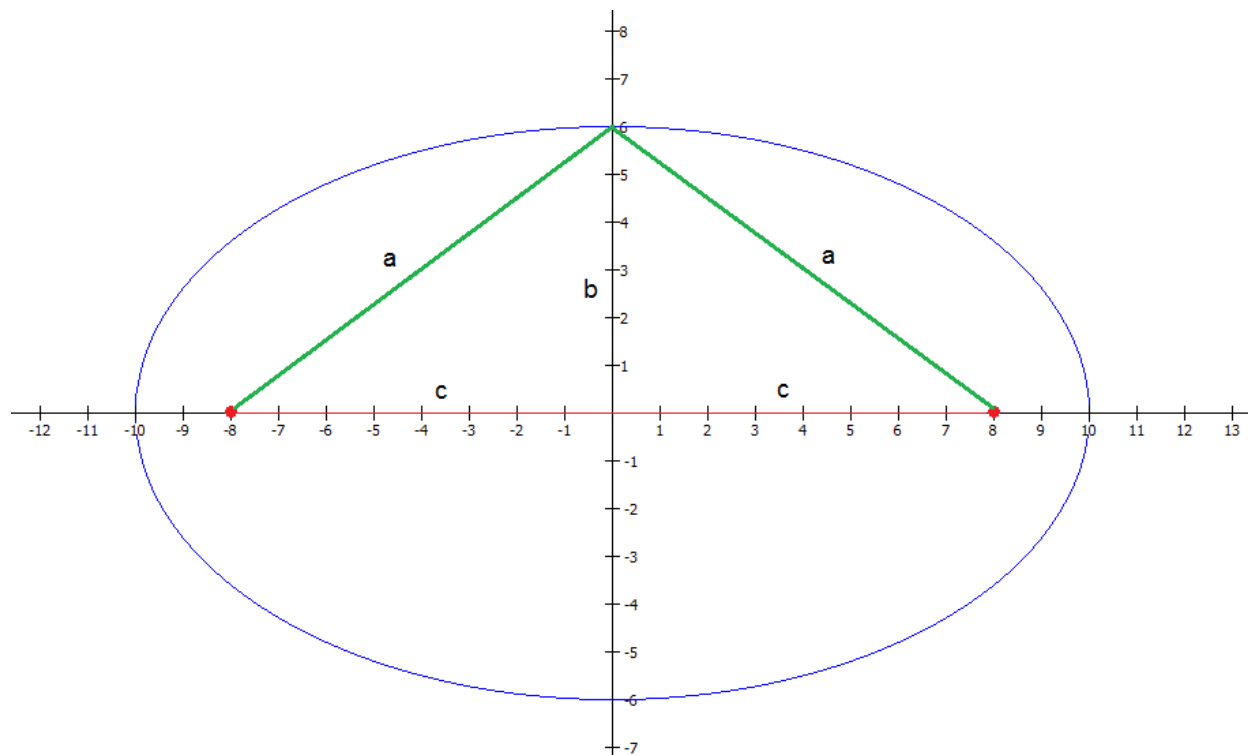
Remember that  $a$  is always larger than  $b$  for the ellipse! This is because we want to use  $2a$  for the longest axis and  $2b$  for the shortest axis. Write your equation as  $\frac{(x-h)^2}{a^2} + \frac{(y-k)^2}{b^2} = 1$  for an ellipse with a horizontal longest axis, and  $\frac{(x-h)^2}{b^2} + \frac{(y-k)^2}{a^2} = 1$  for an ellipse with a vertical longest axis.

The **vertices** of an ellipse are considered to be the endpoints of the major axis [(0,3) and (-3,0) in the picture above]. The **co-vertices** are the endpoints of the minor axis [(2,0) and (-2,0)].

## The Foci of an Ellipse

If you lived in society that was not technologically developed, you could still easily draw a circle by connecting two sticks with a string, placing one in sandy ground and drawing the circle with the other by keeping the string taut, which keeps the radius constant. In the same way, you could draw an ellipse by tying each end of a string to two sticks, and placing them in the ground so that there is still a fair bit of slack in the string. You can then use a third stick to draw an ellipse by keeping the string stretched as far as possible. The points where you place the sticks in the ground are called the foci of the ellipse. The distance from each focus to the center of the ellipse is called " $c$ ". Because the string has a fixed length, the sum of the distances from each focus to a random point on the ellipse is always the same. The length of the longest

“radius” or half-axis of the ellipse is “ $a$ ”. When your “pencil” is at this point, you can see that the length of the string is  $c + a + (a - c)$ , which is  $2a$ . When the drawing instrument is at the end of the shortest axis, the distance to each focus from this point is the same. This distance is “ $a$ ”, and the length of the shortest half-axis is “ $b$ ”, as shown in the image below:



Using the Pythagorean Theorem we can see that  $a^2 = b^2 + c^2$ , giving us the relationship between the three important measurements of an ellipse. Notice that the actual equation of the Pythagorean Theorem is  $c^2 = a^2 + b^2$ , but the measurements of the ellipse have not been named to correspond to this convention. Probably mathematicians can't imagine that anyone might be confused by anything mathematical, so they don't feel a need to worry about this. You may find it easier to remember the odd equation  $a^2 = b^2 + c^2$  if you rewrite it as  $c^2 = a^2 - b^2$ . There, now at least it looks a bit like the Pythagorean Theorem.

## Equations of Circles and Ellipses

Just to make life difficult, your textbook will give you equations for circles and ellipses that are not in a convenient standard form. In fact, you may have difficulty telling whether you are dealing with a circle or an ellipse at all. It may help you to know how authors of textbooks can

produce these equations for you to puzzle over. For example, let's take the simple circle we looked at earlier:  $(x - 2)^2 + (y - 3)^2 = 1$ . You can easily see that this is a circle with radius 1 and center at (2, 3). Now try actually expanding the squared parts:  $x^2 - 4x + 4 + y^2 - 6y + 9 = 1$ . Rearrange that to  $x^2 - 4x + y^2 - 6y = -12$ , and give it to someone you don't like to graph out. For even more fun, multiply the whole equation by something, like maybe 42:  $42x^2 - 168x + 42y^2 - 252y = -504$ . Notice that no matter what I do to this equation of a circle, the coefficients of  $x^2$  and  $y^2$  will always be the same, which gives away the shape of the figure if you know what to look for. To actually get this back to a convenient standard format you must complete the square (see "Factoring"):  $42(x^2 - 4x + \dots) + 42(y^2 - 6y + \dots) = -504$ . The missing numbers are 4 and 9 as you can easily see by cheating and looking back at how we made the equation. You must add  $42 \cdot 4$  and  $42 \cdot 9$  to both sides of the equation to get  $42(x^2 - 4x + 4) + 42(y^2 - 6y + 9) = -504 + 546$ . This gives you  $42(x - 2)^2 + 42(y - 3)^2 = 42$ , and you can take it from there.

The equation of an ellipse can also be rearranged. Take an average ellipse with its center at (h, k):  $\frac{(x-h)^2}{a^2} + \frac{(y-k)^2}{b^2} = 1$  and multiply both sides by  $a^2b^2$  to get  $b^2(x - h)^2 + a^2(y - k)^2 = a^2b^2$ . Then expand the squares, and multiply everything by a random large number. Notice that the coefficients of  $x^2$  and  $y^2$  will be different, but they will either both be positive or both be negative. This should suggest to you that the equation may represent an ellipse. Again complete the squares to get back to standard form.

Some graphing programs will accept an equation like  $x^2 - 4x + y^2 - 6y = -12$  and draw the shape for you. You can probably find one online if you search for "free graphing algebra/geometry".

### Example

Find the location of the foci of the ellipse represented by  $324(x - 1)^2 + 25(y + 4)^2 = 36$

First divide both sides by 36 so you get a 1 on the right side:

$9(x - 1)^2 + \frac{25}{36}(y + 4)^2 = 1$ . Then you must make this look like the equation for an ellipse.

Remembering that dividing by a fraction is the same as multiplying by its reciprocal, you write

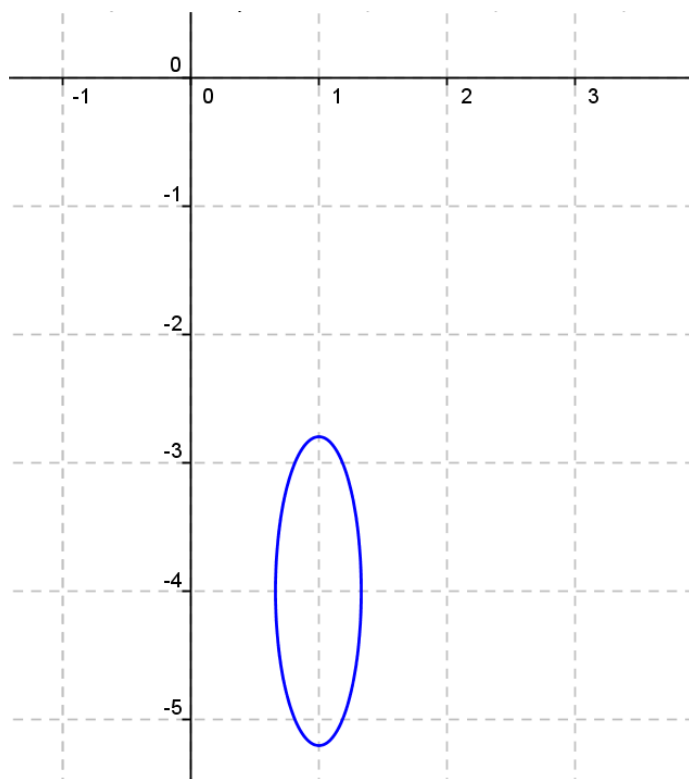
this as  $\frac{(x-1)^2}{\frac{1}{9}} + \frac{(y+4)^2}{\frac{36}{25}} = 1$ . Now you can see that this is an ellipse with its center at (1, -4). By

definition  $a$  is larger than  $b$ , so here  $36/25$  represents  $a^2$ , and  $1/9$  represents  $b^2$ . Because  $a$  and  $b$  are distances, their values are always positive. Take the positive square root ( $a$  and  $b$  are distances, so they are positive) to find that  $a$  is  $6/5$  and  $b$  is  $1/3$ . Knowing that for every ellipse  $c^2 = a^2 - b^2$ , we find that the value for  $c^2 = 299/225$ , meaning that  $c$ , the distance to the focus, is



$\frac{\sqrt{299}}{\sqrt{225}}$ , or  $\frac{1}{15}\sqrt{299}$ . Because the foci are located on the long axis of the ellipse, the vertical axis in this case, they will be found at  $(1, -4 + \frac{1}{15}\sqrt{299})$  and  $(1, -4 - \frac{1}{15}\sqrt{299})$ .

Here is a graph of  $324(x - 1)^2 + 25(y + 4)^2 = 36$ :



## Eccentricity of an Ellipse

The distance to the focus,  $c$ , and half the length of the longest axis,  $a$ , can be used together to describe how flattened and stretched an ellipse is compared to a circle. We define the eccentricity of an ellipse as  $c/a$ . If we push the ellipse back into a circle, both foci come together to become the center of the circle, and the distance  $c$  is zero. When we stretch the ellipse out really far, the focus moves out farther and farther and starts coming quite close to the edge, although it never actually gets there (the ellipse would turn into a line segment if it did). Therefore, the eccentricity  $e$  is between zero and 1:  $0 \leq e < 1$ .

## Deriving the Equation of an Ellipse (Optional)

Those people who are naturally skeptical may wonder if the ellipse that we draw using two foci and a string is really the same kind of figure that we obtained by stretching a circle in one direction or another. Fortunately we can prove mathematically that this is indeed the case. The following proof shows that for the ellipse  $(x/a)^2 + (y/b)^2 = 1$ , we can place two foci at distances “c” from the origin, so that the sum of the distances from any point (x,y) on the ellipse to the foci is always the same, and adds up to 2a. Therefore, even though we used an equation to create the ellipse, it is still the same as if we had used two foci with a string and a pencil.

Draw an ellipse with its major axis along the x-axis, and foci at (-c, 0) and (c, 0). Pick a random point (x,y) on the ellipse in Quadrant I (x and y are both positive). As we saw earlier, the sum of the distances between this point and the foci is 2a. Let’s call these distances  $d_1$  and  $d_2$ , and use the Pythagorean Theorem to describe them.

$$d_1^2 = (x + c)^2 + y^2, \text{ and } d_2^2 = (c - x)^2 + y^2$$

Therefore,

$$\sqrt{(x + c)^2 + y^2} + \sqrt{(c - x)^2 + y^2} = 2a$$

You can’t just get rid of the square roots by squaring both sides, because  $(a + b)^2 = a^2 + 2ab + b^2$ . Rearrange the equation first, and then square:

$$\sqrt{(x + c)^2 + y^2} = 2a - \sqrt{(c - x)^2 + y^2}$$

$$(x + c)^2 + y^2 = 4a^2 - 4a\sqrt{(c - x)^2 + y^2} + (c - x)^2 + y^2$$

Next, move the square root term to the left and everything else to the right so you can square again:

$$4a\sqrt{(c - x)^2 + y^2} = -(x + c)^2 - y^2 + 4a^2 + (c - x)^2 + y^2$$

$$\sqrt{(c - x)^2 + y^2} = \frac{1}{4a} (-x^2 - 2cx - c^2 - y^2 + 4a^2 + c^2 - 2cx + x^2 + y^2)$$

This simplifies to

$$\sqrt{(c-x)^2 + y^2} = \frac{1}{4a}(-4cx + 4a^2), \text{ or}$$

$$\sqrt{(c-x)^2 + y^2} = a - \frac{c}{a}x$$

$$(c-x)^2 + y^2 = a^2 - 2cx + \frac{c^2}{a^2}x^2$$

$$c^2 - 2cx + x^2 + y^2 = a^2 - 2cx + \frac{c^2}{a^2}x^2$$

$$x^2 - \frac{c^2}{a^2}x^2 + y^2 = a^2 - c^2$$

$$x^2 \left(1 - \frac{c^2}{a^2}\right) + y^2 = a^2 - c^2$$

Since  $1 - \frac{c^2}{a^2} = \frac{a^2}{a^2} - \frac{c^2}{a^2}$ , we can change that to

$$x^2 \frac{a^2 - c^2}{a^2} + y^2 = a^2 - c^2$$

Dividing both sides by  $a^2 - c^2$ , we get

$$\frac{x^2}{a^2} + \frac{y^2}{a^2 - c^2} = 1$$

Now substitute:  $b^2 = a^2 - c^2$ , which gives the standard equation for an ellipse:

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$$

## The Area of an Ellipse (Optional)

To find the area of an ellipse we can compare it with a circle that has the same radius as the longest half-axis of the ellipse.

An ellipse  $\left(\frac{x}{a}\right)^2 + \left(\frac{y}{b}\right)^2 = 1$  has a major half-axis of length  $a$  and a minor half-axis of length  $b$

A circle  $x^2 + y^2 = a^2$  has a radius of length  $a$ . Rewrite that as  $\left(\frac{x}{a}\right)^2 + \left(\frac{y}{a}\right)^2 = 1$ .

To change the circle  $\left(\frac{x}{a}\right)^2 + \left(\frac{y}{a}\right)^2 = 1$  into the ellipse  $\left(\frac{x}{a}\right)^2 + \left(\frac{y}{b}\right)^2 = 1$ , we can divide  $\frac{y}{a}$  by  $\frac{b}{a}$ :

$$\frac{y}{a} \div \frac{b}{a} = \frac{y}{a} \times \frac{a}{b} = \frac{y}{b}$$

As we saw earlier, taking the basic equation of a circle ( $x^2 + y^2 = 1$ ) and dividing  $y$  by something causes the  $y$  value of the graph to be multiplied by that number. If we divide  $y$  by 3, every  $y$  value on the graph becomes three times larger. If we would divide  $y$  by  $\frac{1}{3}$ , the circle would be squished by a factor of 3. Now we are dividing the original value  $\frac{y}{a}$  by  $\frac{b}{a}$ . By definition  $b$  is smaller than  $a$ , so the circle becomes squished by a factor of  $\frac{b}{a}$ . In fact, in this squished circle every  $y$  value is only  $\frac{b}{a}$  times as large as the original  $y$  value. You should confirm this by solving both the equation of the ellipse and the circle for  $y^2$ , to get  $y_{\text{ellipse}}^2 = b^2(1 - \frac{x^2}{a^2})$  and  $y_{\text{circle}}^2 = a^2(1 - \frac{x^2}{a^2})$  so that  $\frac{y_e}{y_c} = \frac{b}{a}$

The Italian mathematician Bonaventura Cavalieri (1598 – 1647) reasoned that we can divide both the circle and the ellipse up into infinitely tiny strips running from the top to the bottom. For each tiny strip of the ellipse the area is  $\frac{b}{a}$  times as large as the area of a tiny strip of the circle. If there are infinitely many such strips then the total area of the ellipse is exactly  $\frac{b}{a}$  times the area of the circle. The area of a circle with radius  $a$  is  $\pi a^2$ .  $\frac{b}{a}$  times  $\pi a^2$  is  $\pi ab$ , which is the area of the ellipse.

## Hyperbolas

For hyperbolas, put a first and b second, regardless of which one is larger, because a represents the space between the vertices.

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1 \quad \text{A side-opening hyperbola. Asymptotes: } y = \frac{b}{a}x \text{ and } y = -\frac{b}{a}x$$

$$\frac{y^2}{a^2} - \frac{x^2}{b^2} = 1 \quad \text{A hyperbola that opens to the top and bottom. Asymptotes: } y = \frac{a}{b}x \text{ and } y = -\frac{a}{b}x$$

2a: The distance between the vertices (length of the transverse axis).

2b: The distance from the vertex to the asymptote (length of the conjugate axis).

c: The distance from the center between the curves to the foci.  $c^2 = a^2 + b^2$ .

$$\frac{(x-h)^2}{a^2} - \frac{(y-k)^2}{b^2} = 1 \quad \text{and} \quad \frac{(y-k)^2}{a^2} - \frac{(x-h)^2}{b^2} = 1 \quad \text{are centered at } (h,k).$$

The asymptotes are easy to shift. Just subtract h from x and k from y:

$$y - k = \pm \frac{b}{a}(x - h) \quad \text{and} \quad y - k = \pm \frac{a}{b}(x - h).$$

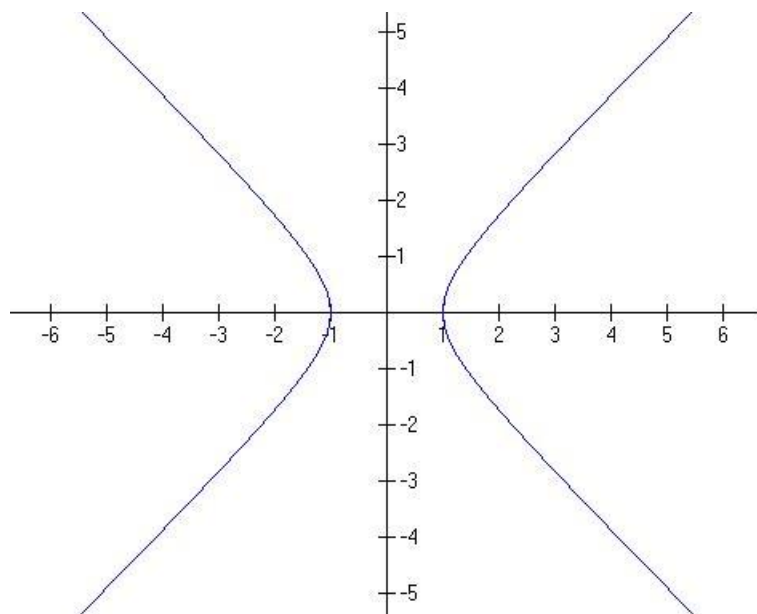
Eventually someone wondered what would happen if you take the equation  $x^2 + y^2 = 1$  and change it to  $x^2 - y^2 = 1$ . Without using a graphing calculator, we can make some predictions about the resulting shape. If you look at  $x^2 + y^2 = 1$ , you can see that both x and y have a fixed range.  $y = \pm \sqrt{1 - x^2}$ , and  $x^2$  is always a positive number. The smallest value for  $x^2$  is 0, so y can't be smaller than -1 or bigger than 1. Also, you can't pick a value for x that is larger than 1 or smaller than -1. If you did, there would be a negative number under the square root sign. Even if you didn't know that the equation represented a circle, you could predict that your graphing calculator would produce a figure with a definite size from this equation.

Now look at  $x^2 - y^2 = 1$ . If you look at this equation closely, you can see that although you can use any value for y, there is a restriction on x:  $x^2 = 1 + y^2$ . Because of the square, the smallest possible value for  $y^2$  is 0. You can make  $x^2$  very large, but not very small. In fact, the smallest

possible value for  $x^2$  here is 1, which means that  $x$  has to be greater than or equal to 1, or less than or equal to -1.

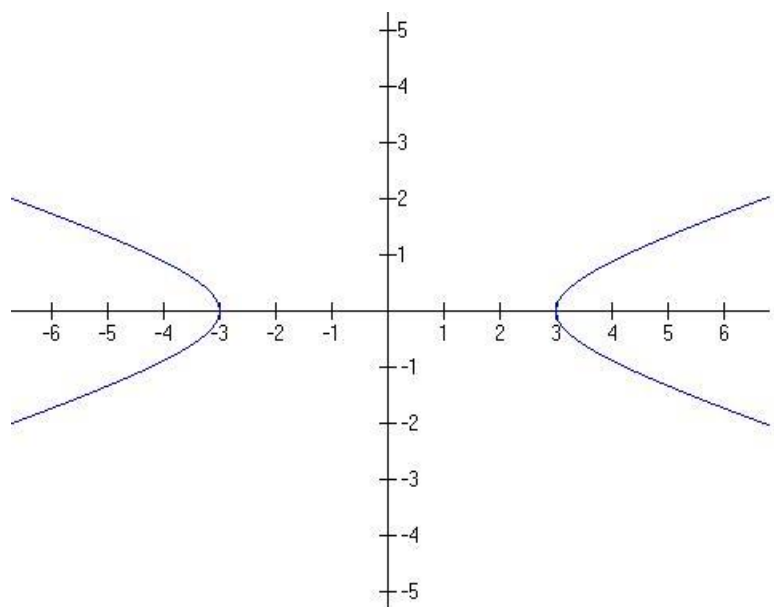
No  $x$  values larger than -1 and smaller than 1 are possible, but there are no other restrictions on  $x$ . This means that the shape defined by this formula does not have a definite size; it extends infinitely far along the  $x$ -axis in both the negative and the positive direction. For each value of  $x$ , the formula supplies two possible values for  $y$  [a positive and a negative value]. As  $x^2$  gets larger, so does  $y^2$

Let's graph this shape. Enter the equation  $x^2 - y^2 = 1$  in a graphing program, or rewrite it as  $y = \sqrt{x^2 - 1}$  and  $y = -\sqrt{x^2 - 1}$ . When you look at the resulting picture, you shouldn't really be surprised that it looks nothing like a circle. This shape is called a hyperbola. It has two branches, each infinitely large:



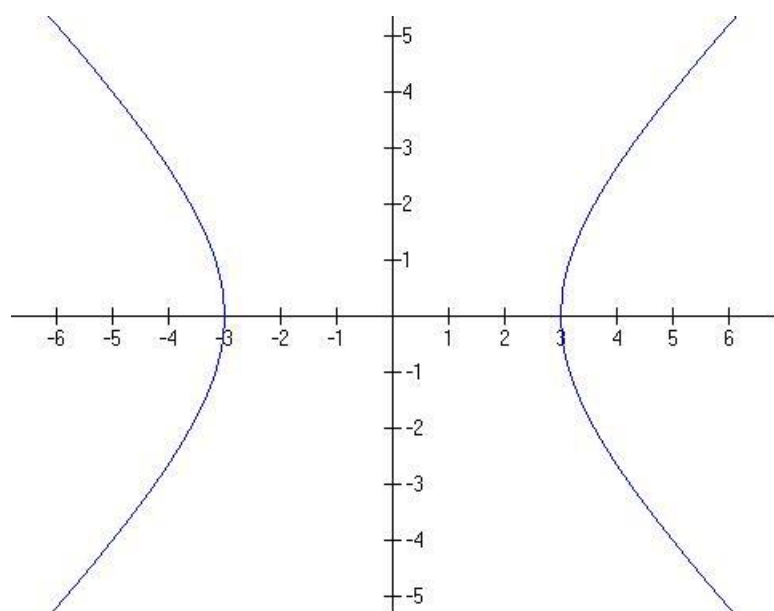
We can manipulate the shape of a hyperbola just like we did with a circle. The space between the two curves corresponds to the values of  $x$  that cannot fit in the equation. Let's try

stretching that space out a little:  $\left(\frac{x}{3}\right)^2 - y^2 = 1$ . Now  $x$  has to be smaller than or equal to -3 or bigger than or equal to 3, as in this picture:

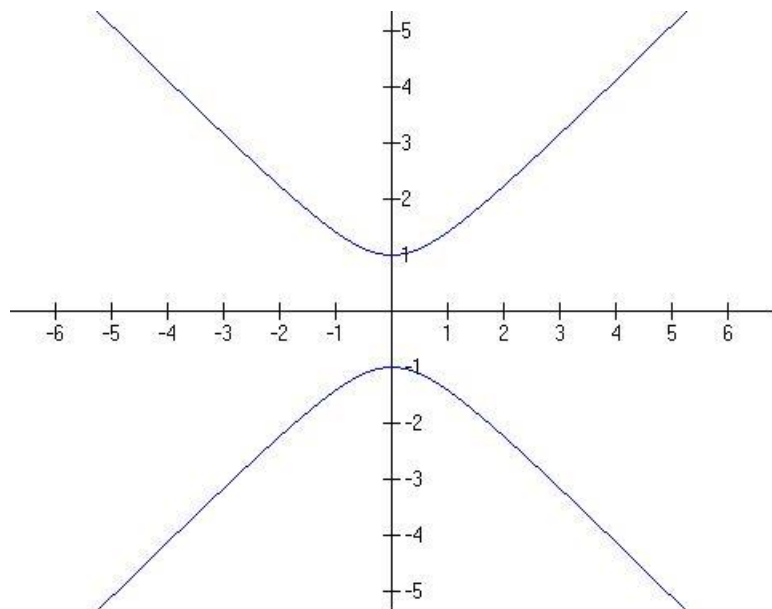


This has the unintended consequence of squishing the curves, because we are actually stretching them out in the x-direction. That is easily fixed by changing our formula to

$\left(\frac{x}{3}\right)^2 - \left(\frac{y}{3}\right)^2 = 1$  as in the following picture:



The next logical question is what happens to the shape defined by  $x^2 - y^2 = 1$  if we reverse  $x$  and  $y$ . The equation for that would be  $y^2 - x^2 = 1$ . Now there are certain values that  $y$  cannot be, while there are no restrictions on  $x$ . In this case  $y$  cannot be between  $-1$  and  $1$ . Our picture looks like this:



Either way, a hyperbola always has two vertices, which are the “points” of the curves. The imaginary line through those vertices is called the major axis or **transverse axis**. The transverse axis is either horizontal or vertical.

For the hyperbola, the distance “ $a$ ” has been defined as the distance from the center between the two branches to the closest point of each branch (the total distance between the branches is  $2a$ ). As we have already seen, this distance is determined by the number we use to divide  $x$  by [or  $y$  if it is the first term]. I know it is hard to remember, but this means that **you must use  $a$  in the first term and  $b$  in the second, regardless of which one is larger**. The general equation of

a hyperbola is  $\left(\frac{x}{a}\right)^2 - \left(\frac{y}{b}\right)^2 = 1$  or  $\left(\frac{y}{a}\right)^2 - \left(\frac{x}{b}\right)^2 = 1$ , depending on whether the curves open to the sides or to the top and bottom. You can also write these equations as  $\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1$  and

$$\frac{y^2}{a^2} - \frac{x^2}{b^2} = 1$$

The vertices of a hyperbola are just like the vertex of a parabola. They are the points of the curve that are closest to the center of the hyperbola. If you have not shifted your hyperbola



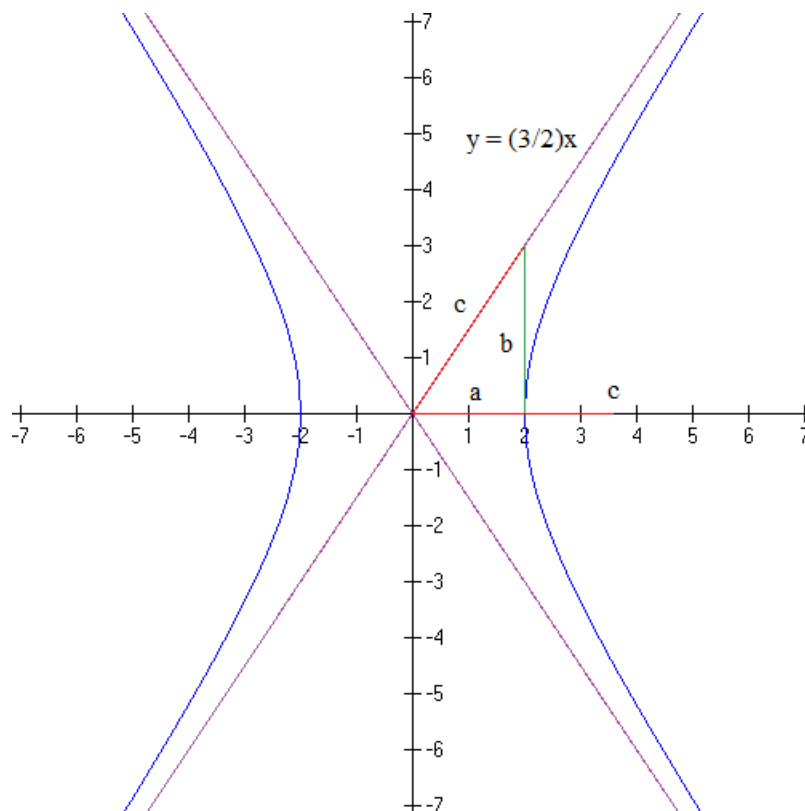
the vertices will be found at  $(a,0)$  and  $(-a,0)$  for a side-opening hyperbola, and at  $(0,a)$  and  $(0,-a)$  for a hyperbola that opens to the top and bottom.

## Asymptotes of a Hyperbola

Let's consider what happens to the top-and-bottom opening hyperbola  $\frac{y^2}{a^2} - \frac{x^2}{b^2} = 1$  as  $x$  gets very large.  $\frac{y^2}{a^2} - \frac{x^2}{b^2} = 1$  can be written as  $\frac{y^2}{a^2} = \frac{x^2}{b^2} + 1$ . As  $x$  gets bigger and bigger, so does  $y$ . Eventually, the 1 on the right side of the equation starts to look very small next to the other numbers.  $\frac{y^2}{a^2}$  becomes nearly equal to  $\frac{x^2}{b^2}$ , but it never quite gets there. If it did, then we'd have  $\frac{y^2}{a^2} = \frac{x^2}{b^2}$ . Taking both the positive and negative square roots of that on each side we get  $\pm \frac{y}{a} = \pm \frac{x}{b}$  or  $y = \pm \frac{a}{b} x$ . Because the hyperbola never reaches these values, the lines  $y = \frac{a}{b} x$  and  $y = -\frac{a}{b} x$  are *asymptotes* of the hyperbola.

Using a similar argument for the side-opening hyperbola  $\frac{x^2}{a^2} - \frac{y^2}{b^2}$ , which can be written as  $\frac{x^2}{a^2} = \frac{y^2}{b^2} + 1$ , we find that the asymptotes for this parabola are  $y = \frac{b}{a} x$  and  $y = -\frac{b}{a} x$ .

To illustrate, let's look at a "side opening" hyperbola with the equation  $\frac{x^2}{4} - \frac{y^2}{9} = 1$



The purple lines are the asymptotes of the hyperbola:  $y = -\frac{3}{2}x$  and  $y = \frac{3}{2}x$ .

We define the foci of a hyperbola to lie inside each curve on an extension of the line segment we called “a”, at a distance  $c$  from the center, such that  $c^2 = a^2 + b^2$ . This should work well for you, since it looks just like the original Pythagorean Theorem. Notice that  $b$  is the distance between the vertex and the asymptote.  $2b$  is the length of the **conjugate axis** of the hyperbola. Although we are drawing  $b$  along the vertex here, the conjugate axis is considered to be a line in the middle between the two vertices.

For each point on the hyperbola the distance from that point to the far focus minus the distance from the point to the closest focus is a constant. By selecting one of the points at the center of the hyperbola, like  $(a,0)$  we can see that this constant is  $2a$ : The distance from the point  $(a,0)$  to the focus  $(-c,0)$  is  $c + a$ . The distance from  $(a,0)$  to the focus  $(c,0)$  is  $c - a$ . Subtracting, we get  $(c + a) - (c - a) = 2a$ .

## Shifting Hyperbolas

Without modification, the equations  $\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1$  and  $\frac{y^2}{a^2} - \frac{x^2}{b^2} = 1$  will always produce hyperbolas centered at the origin. You can shift them just like other conic sections.

$$\frac{(x-h)^2}{a^2} - \frac{(y-k)^2}{b^2} = 1 \quad \text{and} \quad \frac{(y-k)^2}{a^2} - \frac{(x-h)^2}{b^2} = 1 \quad \text{are centered at } (h,k).$$

The asymptotes are also easy to shift. Just subtract  $h$  from  $x$  and  $k$  from  $y$ :

$$y - k = \pm \frac{b}{a}(x - h) \quad \text{and} \quad y - k = \pm \frac{a}{b}(x - h).$$

## Deriving the Equation of a Hyperbola (Optional)

We can prove that for each point on the hyperbola the distance from any point  $(x,y)$  to one focus minus the distance from that point to the other focus is a constant. For convenience, pick a point  $(x,y)$  on the hyperbola  $(x/a)^2 - (y/b)^2 = 1$  so that  $x$  and  $y$  are both positive [a point in the upper part of the right branch of the hyperbola]. The distance between this point and the left focus  $(-c,0)$  is  $\sqrt{(c+x)^2 + y^2}$ . The distance from the point to the right focus is  $\sqrt{(c-x)^2 + y^2}$ . Therefore,  $\sqrt{(c+x)^2 + y^2} - \sqrt{(c-x)^2 + y^2}$  is a constant, which is  $2a$ .

Since  $\sqrt{(c+x)^2 + y^2} - \sqrt{(c-x)^2 + y^2} = 2a$ , we write  $\sqrt{(c+x)^2 + y^2} = 2a + \sqrt{(c-x)^2 + y^2}$  to make it easier to square both sides:

$$(c+x)^2 + y^2 = 4a^2 + 4a\sqrt{(c-x)^2 + y^2} + (c-x)^2 + y^2. \quad \text{This expands to:}$$

$$c^2 + 2cx + x^2 + y^2 = 4a^2 + 4a\sqrt{(c-x)^2 + y^2} + c^2 - 2cx + x^2 + y^2$$

Removing the redundant terms and rearranging to make it easier to square again, we get

$$4cx - 4a^2 = 4a\sqrt{(c-x)^2 + y^2} \quad \text{or} \quad cx - a^2 = a\sqrt{(c-x)^2 + y^2}$$

Squaring both sides gives:  $c^2x^2 - 2a^2cx + a^4 = a^2(c^2 - 2cx + x^2 + y^2)$ , so

$$c^2x^2 - 2a^2cx + a^4 = a^2c^2 - 2a^2cx + a^2x^2 + a^2y^2 \quad \text{or} \quad c^2x^2 + a^4 = a^2c^2 + a^2x^2 + a^2y^2$$

We want to get rid of  $c$ , since this is not part of the equation for the hyperbola. Substituting  $c^2 = a^2 + b^2$  leads us to conclude that:

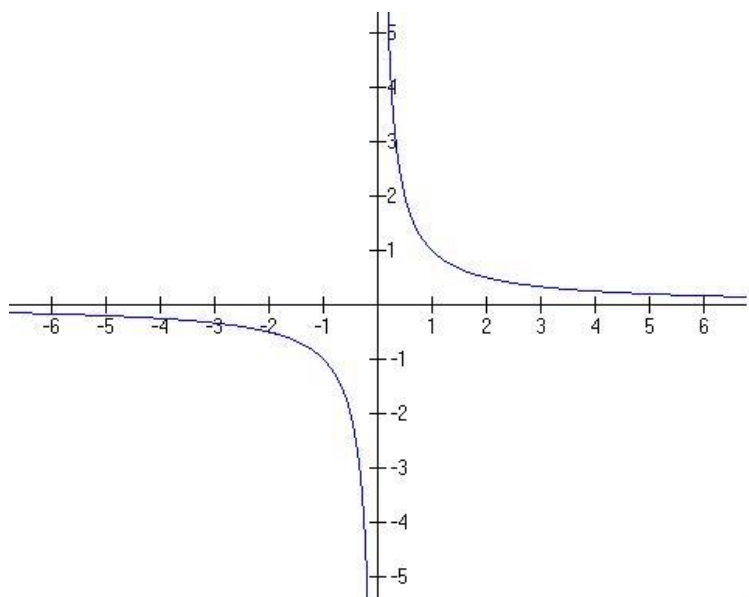
$$a^2x^2 + b^2x^2 + a^4 = a^4 + a^2b^2 + a^2x^2 + a^2y^2, \quad \text{so} \quad b^2x^2 = a^2b^2 + a^2y^2 \quad \text{or} \quad b^2x^2 - a^2y^2 = a^2b^2$$

Now all we have to do is divide by  $a^2b^2$  to get the equation of the original hyperbola

$$(x/a)^2 - (y/b)^2 = 1.$$

## Reciprocal Functions

Surprisingly, the equation  $y = \frac{1}{x}$  also describes a hyperbola. If you rewrite this equation as  $yx = 1$ , you can see that as  $x$  gets bigger  $y$  gets smaller, and vice versa. There is a point where  $x$  and  $y$  are equal. At this point the hyperbola makes its closest approach to the origin and both  $x$  and  $y$  are equal to  $+\sqrt{1}$ , or  $-\sqrt{1}$ . Using this equation, it is not possible to get the curve to approach the origin more closely than a distance of  $\sqrt{2}$ . If you change the equation to  $y = 3/x$ , the shortest possible distance between the curve and the origin is  $\sqrt{6}$ , since both  $x$  and  $y$  are  $+\sqrt{3}$  or  $-\sqrt{3}$  here. For this type of hyperbola, defined by  $y = n/x$ , the distance “ $a$ ” is equal to  $\sqrt{2n}$ . The asymptotes are the lines  $y = 0$  and  $x = 0$ , since either  $y$  or  $x$  can be very very small but never reach 0. These lines are the  $x$ -axis and the  $y$ -axis respectively.



## Sequences and Series

### Arithmetic Sequences and Series

Recursive formula:  $a_n = a_{n-1} + d$

The nth term:  $a_n = a_1 + (n - 1)d$  or  $a_n = a_0 + nd$

The sum of the first n terms:  $S_n = \frac{n}{2}(a_1 + a_n)$

### Geometric Sequences and Series

Recursive formula:  $a_n = a_{n-1} r$

The nth term:  $a_n = a_1 r^{n-1}$

The sum of the first n terms:  $S_n = \frac{a(1-r^n)}{1-r}$

The sum of an infinite geometric series with  $|r| < 1$ :  $S_n = \frac{a}{1-r}$

## Arithmetic Sequences

There are two types of sequences that you will encounter most often. The first one of these is the **arithmetic sequence**. It progresses by adding the same constant amount to each previous term. An arithmetic sequence looks like this:

3, 5, 7, 9, 11, ...

If you look carefully, you can see that the next number in this sequence should be 13, because the numbers are increasing by 2. Here 2 is called the **common difference**,  $d$ . For a sequence like 1, 4, 7, 10, 13, ..., the common difference  $d$  is 3. To determine  $d$ , take any term and subtract the previous term. Keep in mind that  $d$  could be a negative number, in which case the numbers of the sequence keep getting smaller.  $d$  can also be a fraction or a decimal, as in: 9, 8.9, 8.8, 8.7, 8.6, ....

If you have to create your own sequence you should show at least five terms of it so that the pattern of your sequence is relatively clear to someone else. You may think you are showing an arithmetic sequence, but if you provide too few numbers someone else could come up with a different kind of pattern that also fits those same numbers.

By convention, we use the letter  $a$  with a subscript to identify the terms in a sequence. The first term is  $a_1$ , the second term is  $a_2$ , and so on. An arbitrary term somewhere along the line is called  $a_n$ . Just before  $a_n$  you would find  $a_{n-1}$ , and just after it  $a_{n+1}$ :

$a_1, a_2, a_3, a_4, a_5, \dots, a_{n-1}, a_n, a_{n+1}, \dots$

To find the next term in an arithmetic sequence, take the last available term and add the common difference  $d$ . Keep doing that. In computer science, you would accomplish this by creating a loop in your program that the computer keeps following over and over until some specified condition is met. This is called recursion. The recursive formula for finding each successive term just says in algebra what we just said in words: to find any arbitrary term  $a_n$ , take the previous term ( $a_{n-1}$ ) and add  $d$ :

$$a_n = a_{n-1} + d$$

If you also know the first term,  $a_1$ , you can generate the sequence from the recursive formula:

$a_1, a_1 + d, a_2 + d, a_3 + d, a_4 + d, a_5 + d, \dots$

### Example

Find a recursive formula for the arithmetic sequence 3, 5, 7, 9, 11, ....

This sequence can also be written as 3,  $3 + 2$ ,  $5 + 2$ ,  $7 + 2$ ,  $9 + 2$ , etc. Each term is created by adding 2 to the previous term. The recursive formula is  $a_n = a_{n-1} + 2$ , and  $a_1 = 3$ . This says: "Start at 3 and find each next term by adding 2 to the previous term."

Now, suppose that we need to find the 100<sup>th</sup> term in the sequence 3, 5, 7, 9, 11, .... Eek, that looks like a lot of work! Fortunately though there is a regular pattern here, so we should be able to predict what a particular term will be. First, let's look again at how the terms are made. To get the second term in the series we would start with 3 and then add 2, the common difference. For the third term we can start with 3 and add 2 twice to get 7. For the fourth term, start with 3 and add 2 three times to get 9:

$$a_1 = 3$$

$$a_2 = 3 + 2$$

$$a_3 = 3 + 2 \cdot 2$$

$$a_4 = 3 + 3 \cdot 2$$

$$a_5 = 3 + 4 \cdot 2$$

and so on. We can write 3, 5, 7, 9, 11, ... as 3, 3 + 2, 3 + 4, 3 + 6, 3 + 9, ... or in a general way as:

$$a_1, a_1 + d, a_1 + 2d, a_1 + 3d, a_1 + 4d, \dots$$

$a_1$ , the first term, is just an arbitrary starting number, and multiples of the common difference are added to it. Notice that each time the common difference is multiplied by a number that is one less than the number of the term. For the second term, the common difference has been added once, for the third term the common difference has been added twice, and so on. For the  $n^{\text{th}}$  term the common difference has been added  $n - 1$  times. We can find  $a_n$ , the  $n^{\text{th}}$  term of an arithmetic sequence or series by using the **explicit formula**:

$$a_n = a_1 + (n - 1)d$$

For the sequence 3, 5, 7, 9, 11, ..., this formula becomes  $a_n = 3 + (n - 1) \cdot 2$ . Use the distributive property to simplify the part on the right to  $3 + 2n - 2$ , which is  $2n + 1$ .  $a_n = 2n + 1$  provides a simple formula for finding any term of the sequence. The first term,  $a_1$ , is  $2(1) + 1$ , which is 3. For the fifth term,  $n$  is 5 so we get  $a_5 = 2(5) + 1$  which is 11. Now it isn't so hard to find the 100<sup>th</sup> term:  $a_{100} = 2(100) + 1 = 201$ .

We can find a simpler explicit formula if we start with  $a_0$ , which is an imagined term that would come before  $a_1$ . For 3, 5, 7, 9, 11, ...,  $a_0$  would be 1. Now the sequence looks like this:

$$a_0 + d, a_0 + 2d, a_0 + 3d, a_0 + 4d, \dots$$

To get the first term, the common difference is added once, to get the second term it is added twice, for the third term it is added three times, and for the  $n^{\text{th}}$  term it is added  $n$  times:

$$a_n = a_0 + nd$$

This looks simpler and provides the answer you are looking for faster, but you must start by finding  $a_0$ .

Example

Find an explicit formula for the sequence 100, 96, 92, 88, 84, .... Use your simplified formula to find the 10<sup>th</sup> term.

Use the general explicit formula for an arithmetic sequence, and fill in what you know. The common difference here is 4, but it is negative since the numbers are going down.

$$\begin{array}{ll} a_n = a_1 + (n - 1)d & \text{or: } a_n = a_0 + nd \\ a_n = 100 + (n - 1)(-4) & a_n = 104 + n(-4) \\ a_n = 100 - 4n + 4 & a_n = 104 - 4n \\ a_n = 104 - 4n & \end{array}$$

$$a_{10} = 104 - 40 = 64$$

Although  $a_n = a_0 + nd$  looks fairly simple to memorize, there could be a problem when you're writing an exam and running out of time. If you can't remember the formula for  $a_n$ , you can still find the answer by using a shifted sequence. This method treats a sequence like a function that can be shifted. Suppose we have the sequence 5, 7, 9, 11, .... This goes up by 2's, just like 2, 4, 6, 8, 10, .... Here the base, or simplest form of the sequence is  $a_n = 2n$ . When  $n$  is 1, the formula generates 2, when  $n$  is 2 the formula generates 4, and so on. Since the first term of the actual sequence is 5, we need to shift the base sequence by adding 3:  $a_n = 2n + 3$ .

For 100, 96, 92, 88, 84, ..., with a common difference of -4, the base sequence is -4, -8, -12, -16, -20, .... That is the sequence  $-4n$ . Since the first term of the actual sequence is 100, the base sequence has been shifted up by 104. The explicit formula must be  $a_n = -4n + 104$ , just as we saw earlier.

Make your own simple arithmetic sequence, and see if you can find the 10<sup>th</sup> term by using the explicit formula. Check your work by also finding the 4<sup>th</sup> term using your formula.

Example

Find the common difference  $d$  for an arithmetic sequence if the first term is 13 and the 4<sup>th</sup> term is 34. Then find the second and third terms.

We can fill in the terms we know:

$$13, \_, \_, 34$$



At first you may wonder how you could find the missing terms with so little information. However, if you stop and think about it you should realize that to get from the number 13 to the number 34 the common difference  $d$  has been added three times, since 34 is the fourth term. The difference between 34 and 13 is 21. Three times the common difference is 21, so  $d$  must be 7. You can get this same result by using the general formula  $a_n = a_1 + (n - 1)d$  and filling in what you know:  $a_4 = a_1 + (4 - 1)d$  or  $34 = 13 + 3d$ .  $21 = 3d$  so  $d = 7$ . The second and third terms are 20 and 27.

### Example

The following is an arithmetic sequence. Find the missing terms.

..., 5, \_\_, \_\_, 3.5, ...,

Here the common difference must be negative because the numbers are getting smaller. Since the total difference between the two numbers is -1.5, the common difference must be -1.5 divided by 3, or -0.5. The sequence is: 5.5, 5, 4.5, 4, 3.5, 3,....

For an arithmetic sequence, an unknown term that is exactly halfway between two terms can be found by taking the average (arithmetic mean) of the two known terms. Try it out and see if you can understand why that is so. If an odd number of terms are missing, you can easily find the middle one of those missing terms because it will be the average of the known terms on either side:

3, \_\_, \_\_, \_\_, 15

The central missing term is 9, because it is the average of 3 and 15:  $\frac{3 + 15}{2} = 9$ . Once you know this term, the remaining ones are also easily found by taking the average.

If the number of missing terms is even you can use the common difference to find them as we saw earlier. Nick's Rule uses the method of averages to give you another way to determine the first missing term, thereby reducing the remaining missing terms to an odd number that is easy to handle. Nick's Rule is named after a student who insisted that this should be possible, and he turned out to be right.

Nick's Rule for finding the first one of  $m$  missing terms:

$a_1$ , \_\_, \_\_, ..., \_\_,  $a_n$

$$a_2 = \frac{ma_1 + a_n}{m+1}$$

You can check that this is correct by using the formula for  $a_n$ :  $a_n = a_1 + (n-1)d$ . The number of missing terms,  $m$ , is  $n-2$ , because  $a_1$  and  $a_n$  are known.

$$a_2 = \frac{(n-2)a_1 + a_1 + (n-1)d}{n-1}$$

$$a_2 = \frac{na_1 - 2a_1 + a_1 + (n-1)d}{n-1}$$

$$a_2 = \frac{na_1 - a_1 + (n-1)d}{n-1}$$

$$a_2 = \frac{a_1(n-1) + (n-1)d}{n-1}$$

$$a_2 = a_1 + d$$

## Geometric Sequences

The second type of sequence that we need to consider is called *geometric*. This ordered progression is created by multiplying repeatedly by the same number. We start with the first term  $a_1$ , and then multiply by a number that will be the common ratio,  $r$ . The sequence looks like this:  $a_1, a_1r, a_1r^2, a_1r^3, \dots$ . An example would be:

5, 10, 20, 40, 80, ....

Here  $a_1$  is 5. The common ratio is easily found by taking a random term and dividing it by the previous term. We see that  $r$  is 2 in this geometric sequence. To find the next term, take the previous term, 80, and multiply it by 2 to get 160. Keep going. The recursive formula for a geometric sequence is:

$$a_n = a_{n-1} r$$

To generate the sequence, you must also specify  $a_1$ .

### Example

Find the recursive formula for the sequence 3, 15, 75, 375, 1875, ....

It is fairly obvious that this is a geometric sequence rather than an arithmetic one, because the terms get big very quickly. You can look at the first two terms to figure out that the common ratio should be 5, but if the problem doesn't state what kind of sequence you are dealing with you might want to verify that each term is in fact five times larger than the previous one. The recursive formula for this sequence is  $a_n = 5a_{n-1}$ ,  $a_1 = 3$ .

The common ratio  $r$  in a geometric sequence is often a fraction, which causes the numbers to keep getting smaller. It can also be a negative number, which causes the numbers in the sequence to alternate between positive and negative.

Each term in a geometric sequence is made by multiplying  $a_1$  by  $r$  one more time. You can write 5, 10, 20, 40, 80, .... as  $5, 5 \cdot 2, 5 \cdot 2 \cdot 2, 5 \cdot 2 \cdot 2 \cdot 2, 5 \cdot 2 \cdot 2 \cdot 2 \cdot 2, \dots$  or in general as

$$a_1, a_1r, a_1r^2, a_1r^3, a_1r^4, \dots$$

The fourth term in the sequence 5, 10, 20, 40 is  $a_1r^3$ , or 5 times  $2^3$ , which is 40. To find the  $n^{\text{th}}$  term of a geometric sequence without knowing the previous term, the logical explicit formula would be:

$$a_n = a_1r^{n-1}$$

For the sequence 5, 10, 20, 40, 80, ...,  $a_1$  is 5, and the common ratio is 2. To find the  $n^{\text{th}}$  term in the sequence, we can just fill in the explicit formula:

$$a_n = a_1r^{n-1}$$

$$a_n = 5 \cdot 2^{n-1}$$

You can use parentheses to write that as  $5(2)^{n-1}$ . That doesn't create confusion because the exponent always has priority over the multiplication, so it will be clear that only the number 2 is raised to the power  $n - 1$ .

Now again suppose that you are writing an exam, and you just can't remember the explicit formula for a geometric sequence,  $a_n = a_1r^{n-1}$ . Here you can also use the simple trick of taking a base sequence and shifting it. It will take a little more skill because we have to deal with exponents, but it can definitely be done. The common ratio is 2, so your base sequence would be  $2^n$ , as in 2, 4, 8, 16, 32, .... But the real sequence starts with 5, so now what? Well, geometric sequences involve multiplication. Just think of what you would need to multiply 2 by to get 5. After a bit of puzzling you can figure that out:  $2 \cdot \frac{5}{2} = 5$ . Just take your base sequence,  $2^n$ , and multiply it by  $\frac{5}{2}$ :  $a_n = \frac{5}{2}(2)^n$ . But wait, isn't that a different answer than the one we got

before? Well, not really, because we are taking  $2^n$ , multiplying it by 5, and dividing it by 2.  $\frac{2^n}{2} = \frac{2^n}{2^1} = 2^{n-1}$ , so the answer is still  $5(2)^{n-1}$ .

### Example

Determine the explicit formula for the geometric sequence 3, 1.5, 0.75, 0.375, 0.1875, ..., and use it to find the 10<sup>th</sup> term.

To find the common ratio  $r$ , divide a random term by the previous term.  $1.5 \div 3 = \frac{1}{2}$ . The common ratio is  $\frac{1}{2}$ . The first term is 3, so that is  $a_1$ . Now just plug that into the formula:

$$a_n = a_1 r^{n-1}$$

$$a_n = 3 \left( \frac{1}{2} \right)^{n-1}$$

$$a_{10} = 3 \left( \frac{1}{2} \right)^{10-1} = 3 \left( \frac{1}{2} \right)^9 = 3 \left( \frac{1^9}{2^9} \right) = 3 \left( \frac{1}{2^9} \right) = 3 \left( \frac{1}{512} \right) = \frac{3}{512} \text{ or } 0.005859375.$$

If you can't remember the formula, you can create the base geometric sequence that has the common ratio one-half:  $\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{16}, \frac{1}{32}, \dots$ . This is the sequence  $\left( \frac{1}{2} \right)^n$ . Now you just need to multiply by something so the first term will be 3. That would be 6:  $6 \left( \frac{1}{2} \right)^n$ . Think carefully about what the exponents mean to see that this is the same expression as  $3 \left( \frac{1}{2} \right)^{n-1}$ . You can also see that the 10<sup>th</sup> term would be  $6 \left( \frac{1}{2} \right)^{10}$ , which is  $6 \left( \frac{1^{10}}{2^{10}} \right) = 3 \cdot 2 \left( \frac{1}{2^{10}} \right) = 3 \left( \frac{2}{2^{10}} \right) = 3 \left( \frac{1}{2^9} \right)$ .

Make up a simple geometric sequence and find the 10<sup>th</sup> term. Check your work by also finding the 4<sup>th</sup> term using your explicit formula.

### Example

Find the missing terms of the geometric sequence:  $\underline{\hspace{1cm}}, 1, \underline{\hspace{1cm}}, \underline{\hspace{1cm}}, \frac{1}{8}$ .

We can solve this if we can figure out the common ratio. Notice that 1 is not the first term in the sequence, but we know that each term is created by multiplying the previous term by  $r$ , the common ratio. Therefore,  $\frac{1}{8} = 1 \cdot r \cdot r \cdot r$ . This means that  $r^3 = \frac{1}{8}$ , so  $r$  is the cube root of  $\frac{1}{8}$ .

$$\sqrt[3]{\frac{1}{8}} = \frac{\sqrt[3]{1}}{\sqrt[3]{8}} = \frac{1}{2}. \text{ The sequence is } 2, 1, \frac{1}{2}, \frac{1}{4}, \frac{1}{8}.$$

For any geometric sequence you can find an unknown term that is exactly between two known terms by taking the geometric mean of the two known terms. The geometric mean is a special kind of average. If you take any two numbers  $a$  and  $b$  and use them as the sides of a rectangle, the rectangle will have area  $ab$ . Now consider a square with that same area  $ab$ . The sides of the square will have length  $\sqrt{ab}$ , which is the geometric mean of the two numbers  $a$  and  $b$ . For example, the geometric mean of 1 and 9 is  $\sqrt{1 \cdot 9} = 3$ .

Why does this work for a geometric sequence? Well, let's find the fifth term in the sequence 3, 15, 45, 135, 405, 1215, 3645. We'll pretend that we know only the third term, 45, and the seventh term, 3645:  $\sqrt{45 \cdot 3645} = 405$ . That isn't magic, as you can see by looking at the structure of the terms. The third term is  $a_1 r^2$ , and the 7<sup>th</sup> term is  $a_1 r^6$ . When you take the geometric mean of these two terms, you get  $\sqrt{a_1 r^2 \cdot a_1 r^6} = \sqrt{(a_1)^2 r^8} = a_1 r^4$ , the fifth term.

### Example

The 5<sup>th</sup> term of a geometric sequence is 324, and the 7<sup>th</sup> term is 2916. What is the first term?

We can find  $r$  by saying that  $a_7$  is just  $a_5 \cdot r \cdot r$ . That means that  $2916 = 324r^2$ .

$$r^2 = \frac{2916}{324} = 9$$

$r = 3$ . Substitute that into the explicit formula  $a_n = a_1 r^{n-1}$ , along with one of the terms that we already know ( $a_5$  or  $a_7$ ):

$$a_5 = a_1 r^{5-1}$$

$$324 = a_1 \cdot 3^4$$

This will tell you that  $a_1 = 4$ . Check your answer by substituting the values for  $a_1$  and  $r$  into the other possible equation,  $2916 = a_1 r^{7-1}$ .

Nick's Rule for finding the first missing term of the geometric sequence  $a_1, \_, \_, \dots, \_, a_n$  is  $a_2 = \sqrt[n+1]{a_1^m a_n}$ . You can check that this is correct by using the formula  $a_n = a_1 r^{n-1}$ . The number of missing terms,  $m$ , is still  $n - 2$ .

$$a_2 = \sqrt[n-2+1]{a_1^{n-2} \cdot a_1 \cdot r^{n-1}}$$

$$a_2 = \sqrt[n-1]{a_1^{n-1} \cdot r^{n-1}}$$

$$a_2 = a_1 r$$

## Arithmetic Series

An arithmetic series looks like an arithmetic sequence with plus signs between the terms:

$$3 + 5 + 7 + 9 + 11 + \dots$$

Because series have + signs they can be added up. Once you start learning about series, you will probably come across the summation sign  $\Sigma$ . This is the Greek letter S. It is really just a fancy S, and there is no reason to feel intimidated by it. This symbol allows you to quickly write a series without having to make all of those little + signs. The notation uses n for “a variable number”. The starting value for n appears underneath the  $\Sigma$  sign, and the ending value is written above it. [Annoyingly, my text editor shows the starting and ending values for n next to the sign.]

$\Sigma_1^{10} n$  simply means that the starting value of n is 1, the ending value is 10, and all those n’s need to be added up.  $\Sigma_1^{10} n = 1 + 2 + 3 + 4 + 5 + 6 + 7 + 8 + 9 + 10$ .

To write the series  $3 + 5 + 7 + 9 + 11 + \dots$  we have to get a little fancier. Using the formula  $a_n = a_1 + (n-1)d$ , or just some common sense, we write  $\Sigma_1^{10} 3 + 2(n-1)$ . Look at this notation carefully and translate it bit by bit. When  $n=1$  we get 3, when  $n=2$  we get 5, and so on. Of course mathematicians can’t just leave things alone; they have to simplify  $\Sigma_1^{10} 3 + 2(n-1)$  to  $\Sigma_1^{10} 3 + 2n - 2$  which is the same as  $\Sigma_1^{10} 2n + 1$ . Again translate it: when  $n=1$  we get 3, etc. just like we did before.

Infinite series often have an infinite sum, so you will be asked to find partial sums instead. A common question would be to find the sum of the first 20 terms of the series  $3 + 5 + 7 + 9 + 11 + \dots$  (The sum of the first 20 terms is usually indicated by writing  $S_{20}$ ). That would be a lot of work, so you might want to figure out a general way of doing this for any arithmetic series. Fortunately it is surprisingly easy.

The method for adding up an arithmetic series was discovered by a young student who was told to add up the first 100 counting numbers. He could have done that the hard way, but he was smart and realized that these numbers come in pairs. The first number plus the last number add to 101. The second number, 2, and the second-last number, 99, also add to 101. This holds all the way, so that the two middle numbers, 50 and 51 again add up to 101. All we really have to do is say that there are 50 pairs of numbers, each with a sum of 101. The sum is 5050. That sure beats adding all those numbers one by one! The formula for this trick would be  $\frac{n}{2}(1 + n)$ .

Try it out; it actually works whether  $n$  is even or odd. The sum of the first 11 counting numbers is 5.5 times 12 = 66.

Now what would happen with numbers that are not consecutive, but still have regular spacing between them like in our arithmetic series? Let's add up the first 6 terms of  $3 + 5 + 7 + 9 + 11 + \dots$ . The 6<sup>th</sup> term of this series is 13. Again we can add these numbers in pairs, so we get  $3 + 13$ , and  $5 + 11$ , and  $7 + 9$ . Because there are six numbers there are 3 pairs, and the sum is 3 times 16, or  $\frac{6}{2}(3 + 13)$ . This shows us that the general formula for the sum of the first  $n$  terms of a series should be:

$$S_n = \frac{n}{2}(a_1 + a_n)$$

Here  $a_n$  is the last one of the terms that you want to add. [You should try that out with some examples so you can see that this formula always gives the correct answer for the sum.](#)

To get back to our original question, we can now find the sum of the first 20 terms of the series  $3 + 5 + 7 + 9 + 11 + \dots$  quite easily. Using our formula, we get  $S_{20} = \frac{n}{2}(a_1 + a_n)$  or  $\frac{20}{2}(3 + a_{20})$ . The 20<sup>th</sup> term,  $a_{20}$ , can be found from our previous formula  $a_n = a_1 + (n - 1)d$ . The 20<sup>th</sup> term is the first term, 3, plus 19 times the common difference of 2. That makes this term 41, and the sum will be  $10(3 + 41) = 440$ .

## Geometric Series

$5 + 10 + 20 + 40 + \dots$

This is a geometric series. The general form of a geometric series is  $a_1 + a_1r + a_1r^2 + a_1r^3 + a_1r^4 + \dots$ , where  $a_1$  is the first term. Subscripts tend to be a bit awkward when there are a lot of them, so we're just going to write that as  $a + ar + ar^2 + ar^3 + ar^4 + \dots$ , where we all agree that  $a$  represents the first term.

We have a natural tendency to order numbers from smallest to largest, so many geometric series that we might want to write would have an infinitely large sum. However, if the absolute value of the common ratio  $r$  is less than one, the terms of the series keep getting smaller and smaller, and the series actually has a definite sum. It may seem strange that you can add infinitely many numbers and get a definite answer, but it can really be done. You may wonder how mathematicians were able to accomplish this amazing feat. The trick involves finding a formula for the sum of the first  $n$  terms of a geometric series, and then getting rid of most of those terms through a simple subtraction. Watch closely, because it is a lot like magic:

First we say that the sum of the first  $n$  terms,  $S_n$ , is:

$$S_n = a + ar + ar^2 + ar^3 + \dots + ar^{n-1}$$

Then we multiply both sides by  $r$ , to get:

$$r S_n = ar + ar^2 + ar^3 + \dots + ar^n$$

Even though both these expressions have an infinite number of terms, most of them cancel when you subtract them:

$$S_n - r S_n = a - ar^n$$

Because we are interested in the general sum  $S_n$ , we factor the left side and rewrite this as:

$$S_n(1 - r) = a - ar^n, \quad \text{so } S_n = \frac{a(1-r^n)}{1-r}$$

### Example

A finite geometric series with a common ratio of 3 has a sum of 728. The first term is 2. How many terms are there?

If you look at the sum formula, you see that we know all of the variables except for  $n$ .

$$S_n = \frac{a_1(1-r^n)}{1-r}$$

$$728 = \frac{2(1-3^n)}{1-3}$$

$$728 = \frac{2(1-3^n)}{-2}$$

$$728 = -(1 - 3^n)$$

$$728 = -1 + 3^n$$

This tells us that  $3^n = 729$ . Because  $n$  is an exponent, we need something that will get it down for us. A log function will do that:  $n \log 3 = \log 729$ . The natural log function can also be used:  $n \ln 3 = \ln 729$ . Either way, your calculator will tell you that  $n = 6$ . There are 6 terms in the series.

The formula  $S_n = \frac{a(1-r^n)}{1-r}$  works if you are adding up the first 6, or 10 or 100 terms of a series, but it is more interesting to see what happens when you are adding up infinitely many terms. In fact, things involving infinity are often rather interesting. As  $n$  gets larger and larger, so will



$r^n$  if the absolute value of  $r$  is bigger than 1. In that case the sum of the series is infinite. However, if  $|r|$  is smaller than 1,  $r^n$  gets smaller and smaller as  $n$  gets larger. For example, if  $r = \frac{1}{2}$ , then  $r^2 = 0.25$ ,  $r^3 = 0.125$ , and  $r^{10} = 0.000975625$ . You can see where that is heading. We say that the *limit* of  $r^n$  as  $n$  goes to infinity is zero. A limit is often defined as a value that we can get really, really close to, but never quite reach.

Somewhere in the process of us adding up infinitely many terms, the  $r^n$  part of the formula magically disappears, and we are left with  $S_n = \frac{a}{1-r}$ . To see that this is correct, simply divide  $a$  by  $(1 - r)$  using long division. The result is  $a + ar + ar^2 + ar^3 + \dots$

This formula makes mathematicians very happy, because we can now find the sum of the series. Notice that when  $r$  is a negative number smaller than 1, the terms of the series still alternate between negative and positive as they always do when  $r$  is negative. However, when  $r$  is small like that, the terms themselves will get closer and closer to zero, so there actually is a definite sum.

The subject of infinite series is very closely related to one of Zeno's paradoxes. Zeno was a philosopher who lived in ancient Greece. He was interested in the question of infinite divisibility. He developed several paradoxes that help people think about this fascinating question. To see one of these paradoxes, pick a clear space in your home with a wall at the end of it. We will use a straight hallway as an example. Start at one end, and walk halfway to the other end. Estimate the halfway point between where you are and the end of the hallway, and walk there. You have now walked a distance of  $\frac{1}{2}$  the hallway +  $\frac{1}{4}$  of the hallway. The remaining distance is  $\frac{1}{4}$  of the hallway. Again go half of this distance, so that there is only  $\frac{1}{8}$  left. Keep repeating this procedure. Although you approach the end of the hallway very quickly, you can never actually get to the wall, since there is always some infinitely small distance left. Does that infinitely small distance ever become zero? If so when does that happen? If you walk down the hallway in such a weird way, you imagine an infinitely small distance between where you are and the end wall. If you were actually successful in creating this infinitely small distance, everything else would become infinitely large by comparison. Now your hallway is infinitely long, so you can never get to the end of it! 🤖

The total distance from one end of the hallway to the other can be represented by the sum of the infinite series  $\frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \frac{1}{16} + \frac{1}{32} + \frac{1}{64} + \dots$ . This is a geometric series with a first term of  $\frac{1}{2}$  and a common ratio of  $\frac{1}{2}$ . Use the formula for the sum of an infinite series to reassure yourself that you still have 1 hallway [more or less... 😊].

The formula for the sum of an infinite series can be used for a commonly encountered type of geometric series, which is the repeating decimal. A number like 0.99999999.... is really a series that can be written as

$$\frac{9}{10} + \frac{9}{100} + \frac{9}{1000} + \frac{9}{10000} + \dots$$

Notice that the first term,  $a$  is  $\frac{9}{10}$ , and the common ratio is  $\frac{1}{10}$ . Because  $|\frac{1}{10}|$  is less than 1, we can use the formula above to find the sum, even though the series is infinite. The limit of the sum  $S_n$  as  $n$  becomes infinite is  $S_n = \frac{a}{1-r} = \frac{9/10}{1-\frac{1}{10}} = 1$ . This does not actually prove that 0.99999999.... is equal to 1; it merely says that the limit is 1. Using this technique, we can easily convert a number with repeating decimals to a fraction. Try it out for yourself with 0.3333333....

[The alternative, simpler method is to say that  $S = 0.3333333\dots$ , so  $10S = 3.333333\dots$ . Then  $10S - S = 3.333333\dots - 0.333333\dots$ , or  $9S = 3$ , so  $S = 1/3$ . This is a good shortcut, but it doesn't really include a convincing argument that it is a valid method when there are infinitely many numbers after the decimal point.]

## Probability and Statistics

Outliers are more than 1.5 times the interquartile range below Q1 or above Q3.

Z-Score:  $z = \frac{X - \mu}{\sigma}$

Normal distribution: ~ 68% of the data lie within one standard deviation of the mean,  
 ~ 95% of the data will be within 2 standard deviations  
 ~ 99.7% of the data are within 3 standard deviations

Choosing  $r$  items from a total of  $n$ :  $\binom{n}{r} = \frac{n!}{(n-r)!r!}$

The probability of  $r$  successes in  $n$  trials is  $\binom{n}{r} p^r q^{n-r}$

Mean Absolute Deviation:  $\frac{\sum_1^N |X_i - \mu|}{N}$

Standard Deviation (population):  $\sigma = \sqrt{\frac{\sum_1^N (X_i - \mu)^2}{N}}$ , which is equal to  $\sqrt{\frac{\sum_1^N X_i^2}{N} - \mu^2}$

Standard Deviation (sample):  $s = \sqrt{\frac{\sum_1^n (x_i - \bar{x})^2}{n-1}}$

## Basic Statistics Terms

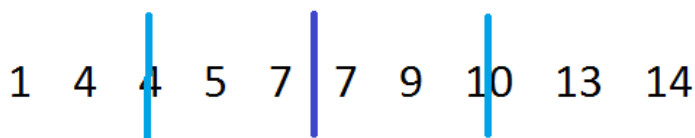
Data that we collect are often numbers. Numerical data are called **quantitative** data; they represent a quantity (an amount). If the data are not numbers we call them **qualitative**, and we sort them into categories.

If a **population** is very large, researchers may choose to collect data from a representative **sample**. The word population here originally referred to people, but now may include animals, plants, and inanimate objects. It is important to select a sample at random in such a way that each member of the population has an equal chance to be chosen as part of the sample. A **statistic** is an attribute of a sample that can be measured. The corresponding value for the population is called a **parameter**. Both capital letters and Greek letters (such as  $\mu$  and  $\sigma$ ) are commonly used for population values, while regular lowercase letters usually refer to sample values. The number of data points is usually called  $N$  for population values, and  $n$  for sample values.

The **mean** is the average of the numbers that make up the data. Simply add up the numbers and then divide by how many numbers you have. The population mean is indicated by  $\mu$  or  $\bar{X}$ , while the sample mean is  $\bar{x}$ . If you have  $n$  sample numbers, you can call them  $x_1, x_2, x_3, \dots, x_n$ , or in general,  $x_i$ .  $\sum_1^n x_i$  means the sum of all of the  $x$ 's, starting at  $x_1$  and ending at  $x_n$ . To get the mean, divide the sum by  $n$ :  $\frac{\sum_1^n x_i}{n}$ .

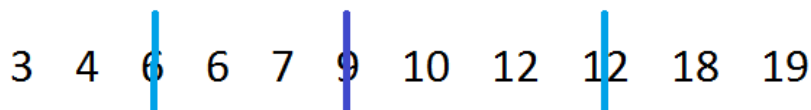
To find the **median**, order your numbers from smallest to largest. The median is the middle point of these numbers, which means that it is either the *middle number* or the *average of the two middle numbers*. For the data set 2, 4, 4, 7, 8, the median is 4 because that is the middle number. For 5, 6, 8, 200, the median is 7 because that is the average of 6 and 8.

The median divides the data in half, and when you divide those halves in half again you get the **quartiles**. This division is done the same way as for the median. If the median is between two numbers, these numbers are included in the half that you divide to get the quartiles:



In the picture above, there are an even number of data points. The median ( $Q_2$ ) is 7, the lower quartile ( $Q_1$ ) is 4 and the upper quartile ( $Q_3$ ) is 10.

If there are an odd number of data points, there is an actual middle number that is the median. Put a line through the middle number and then ignore it as you look for the quartiles:



Just remember that 50% of the data are always at or below the median, and the other 50% are at or above the median. 25% of the data are located in each quartile.

**Outliers** are data points that don't seem to "fit" with the remaining data because their value is too large or too small. Outliers may represent an unusual event or an error in measurement. If you are asked to calculate if a particular point is an outlier, find the interquartile range (the distance between the upper and lower quartile) and multiply it by 1.5. Then add this figure to the number that represents the top quartile and subtract it from the number that represents the bottom quartile. A value outside this range may be considered an outlier.

A **boxplot** (box-and-whisker plot) shows the median, the upper and lower quartiles, and the range: <http://www.basic-mathematics.com/box-and-whiskers-plot.html>. The whiskers may extend to the highest and lowest values in the data set, or unusually high or low values may be marked separately with an asterisk.

The **range** is the difference between the highest value and the lowest value. It shows the "spread" of the data. If the range is large, that could be because there is just a single very low or very high value, or it could mean that all of the data points are very spread out.

## The Binomial Theorem

To find the terms of  $(a + b)^n$  you can use Pascal's Triangle, or  $\binom{n}{k}a^{n-k}b^k$  where  $k$  starts at 0 and goes up to  $n$ .

You can thank (or not) Isaac Newton for this theorem, which allows you to quickly expand an expression like  $(a + b)^n$ . By using successively larger values of  $n$  and observing the resulting pattern, Newton was able to find a general formula for these expansions. He then cleverly extrapolated his results to fractional values for  $n$  to get an expression for  $\sqrt{a + b}$ , which is really  $(a + b)^{1/2}$ . Oddly enough, when you use fractional values for  $n$  your answer ends up having an infinite number of terms!

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

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

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

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

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

Notice that this pattern creates a symmetrical pyramid. The exponents of  $a$  and  $b$  are completely predictable. For  $(a + b)^n$ , the first term is always  $a^n$ , which we could also write as  $a^n b^0$  (recall that any number raised to the  $0^{\text{th}}$  power is 1 – see “Exponents”). The second term always has  $a^{n-1} b^1$ , and the pattern continues with  $a^{n-2} b^2$ ,  $a^{n-3} b^3$ , and so on until we reach  $a^0 b^n$ . Also note that the exponents of  $a$  and  $b$  for each term always add up to  $n$ : for  $a^4 + 4a^3 b^1 + 6a^2 b^2 + 4a^1 b^3 + b^4$  the sum of the exponents is always 4.

The number in front of the  $ab$  parts is called the *coefficient*. If there is no number the coefficient is 1.

Instead of writing out the entire pyramid shown above, we can quickly find the coefficients from Pascal’s Triangle:

$$\begin{array}{ccccccc}
 & & & & 1 & & & \\
 & & & 1 & & 1 & & \\
 & & 1 & & 2 & & 1 & \\
 & 1 & & 3 & & 3 & & 1 \\
 1 & & 4 & & 6 & & 4 & & 1 \\
 1 & & 5 & & 10 & & 10 & & 5 & & 1 \text{ etc.}
 \end{array}$$

In Pascal’s Triangle the outside numbers are always 1, and any other number is always the sum of the two numbers directly above it.

The coefficients of the binomial expansion also correspond to the formula for choosing some items from a number of items, where the order of the chosen items does not matter. This is called *Combinations* (see “Choices and Chances” in Algebra 1.) The first coefficient corresponds to choosing 0 items from a list of  $n$ , also written as  $\binom{n}{0}$  – read “ $n$  choose 0.” So, for  $(a + b)^5$  we would find the first coefficient by counting how many ways we could choose 0 items out of a list of 5. Of course there is only 1 way to do that; we just don’t select any. For the next coefficient we use  $\binom{5}{1}$ , or 5 choose 1. There are 5 different ways to do that, and the coefficient is 5. For the third coefficient we choose 2 items out of 5,  $\binom{5}{2}$ . This is slightly more complicated. For our first item we have 5 choices, but once we have selected that item there are only 4 choices for the second item. That gives  $5 \cdot 4 = 20$  possibilities, but recall that the order of the chosen items doesn’t matter. The choice of “item 3 and item 5” is the same as the choice of “item 5 and item 3.” For every combination of two items, there is another one with those same two items reversed. The total number of possible choices must be divided by 2:  $20 \div 2 = 10$ . The third coefficient is the result of  $\binom{5}{3}$ . When we choose 3 items out of 5, we have 5 choices for the first, 4 for the second, and 3 for the third, for a total of  $5 \cdot 4 \cdot 3 = 60$  possibilities. However, many of those possibilities are the same. We have to divide that by the

number of ways in which we can arrange the chosen items. When you have 3 items, you have 3 choices for the first one, 2 for the second, and 1 for the third. There are a total of  $3 \cdot 2 \cdot 1 = 6$  different ways to order these three items. That means we have to divide the total number of possibilities by 6 to account for the fact that many of our choices are really the same thing.  $60 \div 6 = 10$ . Because of the symmetry of the coefficients we don't have to go any further. The remaining coefficients are 5 and 1.

There is a formula to calculate how to choose  $r$  things from a list of  $n$  items, if the order doesn't matter. A shorthand notation has been developed to express this formula, because it gets very tedious to write things like  $10 \cdot 9 \cdot 8 \cdot 7 \cdot 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1$ . Instead we write  $10!$ , which is 10 *factorial*. The formula for  $\binom{5}{3}$  would be  $\frac{5!}{2!3!}$ . The 2! On the bottom makes sure that we get  $5 \cdot 4 \cdot 3$  choices, and then we have to divide by the number of ways to rearrange 3 items which is  $3 \cdot 2 \cdot 1$  or  $3!$ .

It may look complicated to calculate  $\frac{5!}{2!3!}$ , but a lot of the terms cancel out:  $\frac{5 \cdot 4 \cdot 3 \cdot 2 \cdot 1}{2 \cdot 1 \cdot 3 \cdot 2 \cdot 1}$  is the same as  $\frac{5 \cdot 4}{2 \cdot 1}$ . Always cancel the longest string of numbers, top and bottom.

If there are  $r$  items to be chosen from a total of  $n$  items, that is often written as  $nCr$ , for  $n$  Choose  $r$ , or with parentheses as  $\binom{n}{r}$ . The formula is  $\binom{n}{r} = \frac{n!}{(n-r)!r!}$ .

A note of caution here:  $r$  starts at 0, for the first term. This means that  $r$  would be 3 for the 4<sup>th</sup> term. **Remember to subtract 1 from each term number!** Also, check your work using Pascal's triangle, or a calculator. On the TI-84 calculator press MATH and navigate to PRB. Here you will find both ! and  $nCr$ . You must choose  $n$  first before selecting  $nCr$ , and then you can enter  $r$ .

## The Binomial Distribution

Mathematical discoveries that may seem very abstract are often driven by very practical needs. The normal distribution was discovered as a result of people needing to win at gambling. Once it was there, it was soon used for other purposes, but it is still best understood in terms of simple probabilities. The simplest form of probability is that of two equally possible outcomes, such as heads or tails on a coin flip.

Imagine that you have 2 coins. If you flip them both and see whether they land heads or tails, you may not get 1 head and 1 tail. It is possible to calculate the probability of a particular outcome. In this simple situation, we can consider each possibility separately. There is only

one way to get 2 heads, and only 1 way to get 2 tails, but there are 2 different ways to get 1 head and 1 tail:

HH

HT

TH

TT

The outcome of one head and one tail has a probability of 2 out of 4, or  $\frac{1}{2}$ . The outcome of 2 heads has a probability of 1 out of 4, or  $\frac{1}{4}$ , and 2 tails also has a probability of  $\frac{1}{4}$ . Note that the total probability is 1, because the result will always be one of these 3 possibilities.

You may already know that to find the probability of two events both occurring, you should multiply their individual probabilities. For two heads, you would say that the probability of the first coin coming up heads is  $\frac{1}{2}$ , and the probability of the second coin coming up heads is also  $\frac{1}{2}$ .  $\frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$ . However, you also have to multiply by the number of different ways that this event can happen. Two heads can only happen in one way, so the probability is  $1 \times \frac{1}{2} \times \frac{1}{2}$ . One head and one tail can happen in two different ways, so multiply by 2:  $2 \times \frac{1}{2} \times \frac{1}{2} = \frac{1}{2}$ .

When you have more coins to flip, the resulting probabilities are a little more complex to find by listing all of the outcomes. For 4 coins, you can reason that there definitely is only one way to get 4 heads, and one way to get 4 tails. The probability of 4 heads is  $\frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} = 1/16$ . It may look like 1 head and 3 tails is a single outcome, but there are 4 ways that this could happen: HTTT, THTT, TTHT, and TTTH. The universe carefully keeps track of this so that the outcome 1 head and 3 tails is four times as likely to happen as 4 heads. The probability of 1 head and three tails is  $4 \times \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} = 4/16$  or  $\frac{1}{4}$ . Again for 1 tail and 3 heads, there are 4 ways, and the probability of that is also  $\frac{1}{4}$ . There are actually 6 ways get 2 heads and 2 tails. You can see that by listing them individually, or by calculating it like this: 2 of the coins will be tails – how many different ways can we choose these 2 coins out of a total of 4? Well, to pick the first coin we have 4 choices, and then when we go to pick the second there are 3 choices left. There are  $4 \times 3$  or 12 ways to pick those 2 coins. But, once we have picked them some of our choices are the same. Label the four coins A, B, C, and D. Then picking A and C is the same as picking C and A. Each choice has a double, because there are two ways to arrange the two coins that we pick. That means the total of 12 must be divided by 2 to get 6 unique ways to pick 2 of the coins to be tails, or 2 to be heads. The probability of 2 heads and 2 tails is  $6 \times \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} = 6/16$  or  $3/8$ . When you do calculations like this, check to make sure that the total probability still adds up to 1. Here we have  $1/16 + 4/16 + 6/16 + 4/16 + 1/16 = 1$

The more coins you flip in one session, the more possible outcomes there are. Fortunately coin flips follow the Binomial Theorem:



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

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

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

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

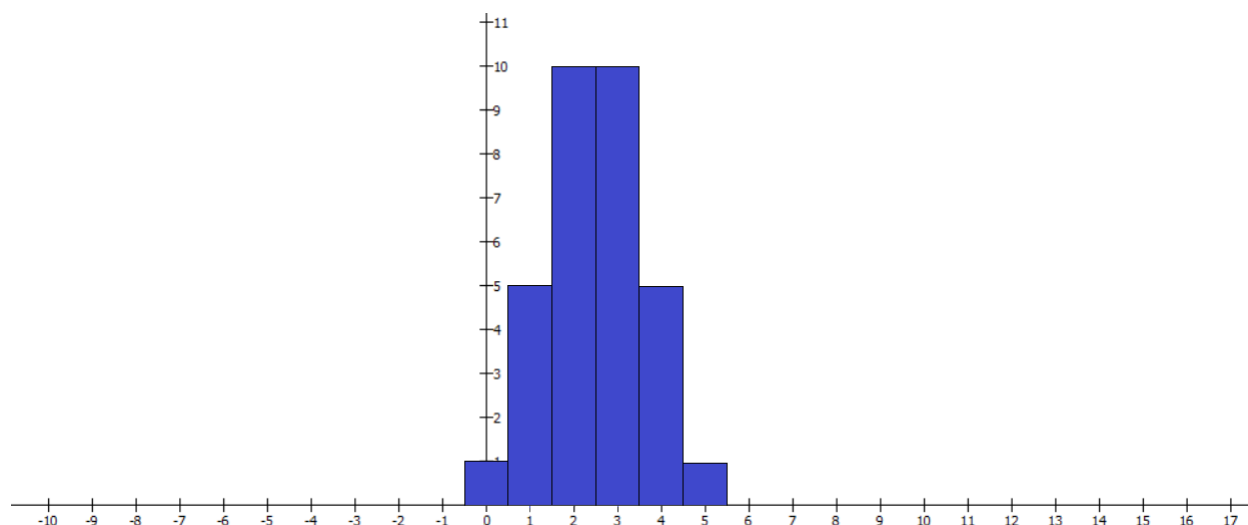
$$(a + b)^4 = a^4 + 4a^3b + 6a^2b^2 + 4ab^3 + b^4$$

When you multiply  $(a + b)(a + b)(a + b)(a + b)$ , you have to take each individual part and multiply it by the other parts. That creates the same kind of situation as the different combinations that result from flipping coins. For 4 coins it looks like this:

$$(H + T)^4 = 1HHHH + 4HHHT + 6HHTT + 4HTTT + 1TTTT$$

In general, the coefficients for  $(H + T)^n$  use the formula  $\binom{n}{r}$ , where  $r$  is the number of heads (or tails), and  $\binom{n}{r} = \frac{n!}{(n-r)!r!}$ .

Here is a bar graph showing the number of ways that 0, 1, 2, 3, 4, or 5 heads can appear when five coins are flipped:



The bars look tall because I have graphed the frequencies of the outcomes rather than their probability. The probability of each outcome is the frequency of that outcome divided by the total number of outcomes. You can calculate that from the binomial theorem, or more directly as straight probability.

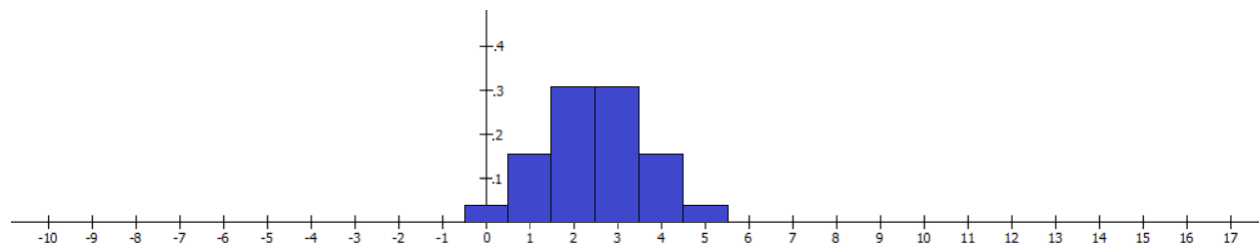
To find the probability of 1 head and 4 tails when five coins are flipped, consider that the probability of heads is  $\frac{1}{2}$ , and the probability of tails is  $\frac{1}{2}$ . So, the probability of coin A coming

up heads is  $\frac{1}{2}$ , and coins B, C, D, and E had a probability of  $\frac{1}{2}$  to come up tails. To get the probability of the overall event, we have to multiply the individual probabilities:

$$\frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} = 1/32$$

However, the same outcome can be achieved if coin B comes up heads and coin A comes up tails. As we saw earlier, there are 5 different ways to get 1 head and 4 tails. The overall probability is  $5 \times \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} = 5/32$ . In general, the probability of  $r$  heads (or  $r$  tails) is  $\binom{n}{r}$  [which is  $= \frac{n!}{(n-r)!r!}$ ], multiplied by the probability of each individual head or tail.

The probability of 0 heads is  $1 \times \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} = 1/32$ , or 0.0325. That is hard to graph, so people often magnify the scale on the y-axis. The graph below has a 10-magnification on the y-scale so you can actually see something:



The more coins there are, the more a graph like this starts looking like you could draw a smooth curve using the tops of the bars. That curve turns out to be a **normal curve**, which is a bell-shaped probability curve.

Heads or tails are two mutually exclusive outcomes, and there are many others like that. Often they are labeled “Success” and “Failure”, and the probability of each one could be more or less than 50%. Use the same rules even when the probabilities are uneven. For example, if the probability of success is 60% and the probability of failure is 40%, you can calculate the probability of 2 successes in 5 trials as follows:  $\binom{5}{2} \times 0.6 \times 0.6 \times 0.4 \times 0.4 \times 0.4$ .

If the probability of “Success” is  $p$ , and the probability of “Failure” is  $q$ , then the probability of  $r$  successes in  $n$  trials is  $\binom{n}{r} p^r q^{n-r}$

## The Normal Distribution

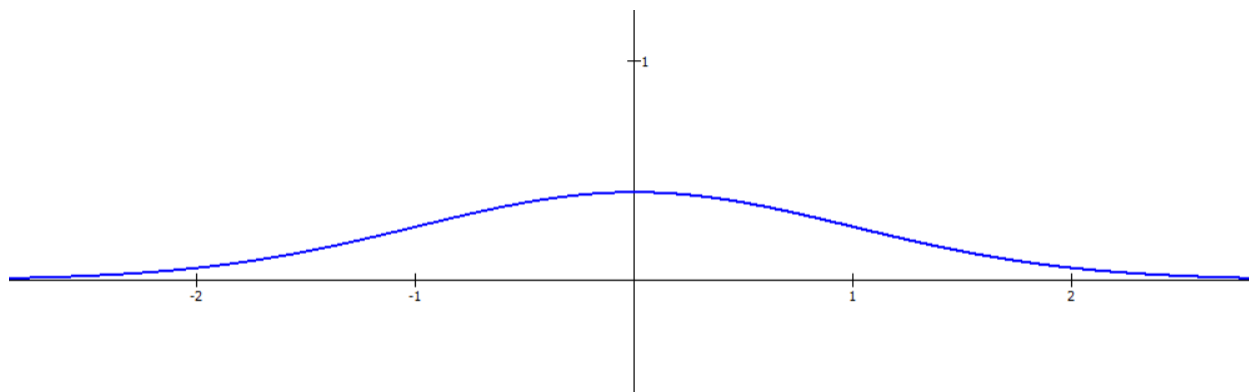
The normal curve was useful for gambling predictions, but as it turns out many natural phenomena have an approximately normal distribution. For example, if you graph the height of a large number of people you get a set of dots that tend to form a bell-shaped curve. We can model these data by drawing a normal curve, such that the area under the curve shows the probability that a particular person's height will be between two given values.

The *standard* normal curve has a standard deviation of 1, and a mean of zero. Its formula is

$$y = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

Because this looks complicated, you may tend to just glaze over it. However, you already know a lot about functions, so you can see that this is very similar to an exponential function, with a base of  $e$ . The minus sign on the exponent means that the value of the function decreases as  $x$  gets larger:  $e^{-\frac{x^2}{2}} = \frac{1}{e^{\frac{x^2}{2}}}$ . Because  $x$  is squared, the graph will be symmetrical, with decreasing

values to the right and the left. The total probability for multiple possible outcomes is always 1, so the area underneath the normal curve is 1. Normal curves are usually depicted as being quite tall, with a magnified y-scale. However, if the area under the curve has to be 1, the height has to be significantly less than 1. Here is a picture of the standard normal curve drawn from its formula without changing the scale:



The maximum value occurs when  $x = 0$ , and that value is  $\frac{1}{\sqrt{2\pi}}$ , which is just a little bit less than 0.4. The minimum value looks like it is zero, but that is not actually true. Far-out  $x$ -values are not impossible, just extremely unlikely.  $\frac{1}{e^{\frac{x^2}{2}}}$  gets smaller as  $x$  gets larger or very negative, but it

never reaches zero. The x-axis is a horizontal asymptote, and the graph never touches it. Still, the area underneath the graph is 1!

The mean, which is the average x value, is zero. That isn't always convenient. For example, if you are measuring the height of growing plant seedlings, you might find that your mean is 5 inches. When you learned about function graphs, you also learned how to shift those graphs right or left. To change the mean from 0 to 5, you want to subtract 5 directly from x:

$$y = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-5)^2}{2}}$$

Ooh, I altered a really complicated function. Yes, that is allowed. In general, to shift the curve to have a mean of  $\mu$ , the population mean, you should subtract  $\mu$  directly from x:

$$y = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2}}$$

## Standard Deviation

A graph or boxplot shows how “spread out” the data are, but we also need a precise mathematical way of expressing that as a number. The range is somewhat helpful, but it only gives information about the highest and lowest value.

Let's consider a really simple example. That's always a good idea when you are trying to figure out how to do something. Here is a set of data:

3   9   12

The first thing you might want to do when figuring out the spread of the data is to see how far away each value is from the mean. In this case the mean,  $\mu$ , is 8. We'll call the first data point  $x_1$ , the second one  $x_2$ , and the third  $x_3$ . In general, we'll use  $x_i$  to refer to a non-specific data point. We could take each data point and measure its deviation from the mean like this:  $x_i - \mu$ . That gives us three values:  $3 - 8$ ,  $9 - 8$ , and  $12 - 8$ . To get the average deviation we would have to add these values and divide by 3:  $(-5 + 1 + 4)/3$ . But oops, that is 0. And unfortunately that is no accident. The deviations that are below the mean will always cancel out the deviations above the mean. There are two ways to avoid this problem. The simplest way is to measure the positive distance between each data point and the mean by taking the absolute value,  $|x_i - \mu|$ . Then find the mean (the average) of these values, which is the **mean absolute deviation**. For this simple set of data the mean absolute deviation is  $(5 + 1 + 4)/3 = 3\frac{1}{3}$ .

The other way to compensate for the fact that the negative deviations cancel out the positive ones is to use squares:  $(X_i - \mu)^2$ . You can add these squares and average them to get the **variance**. Then you need to take the square root of this average because otherwise your units won't match up with the original units of your data. This square root of the average of the squared differences is the **standard deviation**. The Greek letter  $\sigma$ ,  $\sigma$ , is used for the standard deviation of a population, while a plain letter  $s$  is used for the standard deviation of a sample. Here we are just using a made-up set of values, which is the entire population rather than a sample of it. The formula is:

$$\sigma = \sqrt{\frac{\sum_i^N (X_i - \mu)^2}{N}}, \text{ where } N \text{ is the number of values.}$$

$\sum_i^N (X_i - \mu)^2$  is the sum of all of the squares of the differences from the mean:  $(3 - 8)^2 + (9 - 8)^2 + (12 - 8)^2$ . That works out to  $25 + 1 + 16 = 42$ . To get the average of the squares divide by 3, which gives you 14. The standard deviation is  $\sqrt{14}$ , which is about 3.74.

Alternative formula for the standard deviation:

$$\sigma = \sqrt{\frac{\sum_i^N X_i^2}{N} - \mu^2}$$

Is this really the same formula? To figure that out I used the simple set of 3 data above, and the fact that  $(a - b)^2 = (a - b)(a - b) = a^2 - 2ab + b^2$ .

$$(3 - 8)^2 + (9 - 8)^2 + (12 - 8)^2 = 3^2 - 2 \cdot 3 \cdot 8 + 8^2 + 9^2 - 2 \cdot 9 \cdot 8 + 8^2 + 12^2 - 2 \cdot 12 \cdot 8 + 8^2$$

$$\text{Rearranging that, I got } 3^2 + 9^2 + 12^2 - 2 \cdot 8(3 + 9 + 12) + 3 \cdot 8^2$$

Now divide everything by  $N$ , the number of data points, which is 3 in this case:

$$\frac{3^2 + 9^2 + 12^2 - 2 \cdot 8 \cdot (3 + 9 + 12) + 3 \cdot 8^2}{3}$$

Each term, which is separated from the other terms by a plus or minus sign, has to be divided by 3. The best way to do that is like this:

$$\frac{3^2 + 9^2 + 12^2}{3} - \frac{2 \cdot 8 \cdot (3 + 9 + 12)}{3} + \frac{3 \cdot 8^2}{3}$$

Notice that the first part is the sum of all the squares of the data, divided by  $N$ . For the middle part,  $\frac{2 \cdot 8 \cdot (3 + 9 + 12)}{3}$ , we can get rid of the 3 by dividing  $\frac{(3 + 9 + 12)}{3}$ , which we already know is the mean, 8. That leaves:

$$\frac{3^2+9^2+12^2}{3} - 2 \cdot 8 \cdot 8 + 8^2$$

which is the same as  $\frac{3^2+9^2+12^2}{3} - 2 \cdot 8^2 + 8^2$ , or  $\frac{3^2+9^2+12^2}{3} - 8^2$ . Then take the square root, and it's just like filling in the formula:

$$\sigma = \sqrt{\frac{\sum_i^N X_i^2}{N} - \mu^2} = \sqrt{\frac{3^2+9^2+12^2}{3} - 8^2}$$

If you are examining data from a sample that is randomly taken from a large population, you might expect that these sample data would show less variability than what exists in the general population. For example, if the population has just a few very tall people, they probably won't be part of the sample. To correct for that, when you are calculating the standard deviation of a

sample, you divide by  $n - 1$  rather than by  $n$ :  $s = \sqrt{\frac{\sum_i^n (x_i - \mu)^2}{n-1}}$ . The division by a smaller

number increases the value you get for the standard deviation, so that it is closer to the actual standard deviation of the population. As the number of values in the sample increases, subtracting 1 has a relatively smaller effect.

These calculations are best done with a graphing calculator or online calculator.

Most normal distributions don't have a standard deviation of 1. The standard deviation is more complex to change, but it is done by adjusting the variance,  $\sigma^2$ . That didn't show in our previous formula for the normal curve because its value was 1. The full equation looks like this:

$$y = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \text{ which is the same as } y = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

## Z-Scores

For any normal distribution, approximately 68% of the data lie within one standard deviation of the mean, about 95% of the data will be within 2 standard deviations, and by the time you get 3 standard deviations away from the mean in either direction you will have included approximately 99.7% of all of the results. This is true regardless of the actual size of the standard deviation, or the value of the mean. Mathematicians have already calculated areas under normal distribution curves in great detail, and created tables of the results. So, if you

know that your data has an approximately normal distribution, you can find what percentage of those data lie above or below a certain value. All you do is take that value and find how many standard deviations away from the mean it is.

For example, if the mean height of your plant seedlings is 5 inches, with a standard deviation of 1.2 inches, you can find how many of those seedlings are likely to be taller than 7 ½ inches. First we find how many standard deviations from the mean 7.5 inches would be.  $7.5 - 5 = 2.5$ , so that is 2.5 inches taller than the mean. Just looking at it you can see that this represents just over 2 standard deviations, but to get the actual value you need to divide 2.5 by 1.2. The z-score is  $\frac{X - \mu}{\sigma}$  or  $\frac{7.5 - 5}{1.2} \approx 2.08$ . If you look that up in a table you get .9812, which means that 98.12% of the seedlings are *shorter* than 7.5 inches, since the z-score gives the area of the curve below the given value. This means that 1.88% of the seedlings are expected to be taller than 7.5 inches. Tables are a bit of a pain, and online calculators do the job much faster.

Your TI-84 will also find z-scores. Use the normalcdf( function found under DISTR, located above the VARS key. You must tell your calculator that you would like the area under the curve from 2.08 to positive infinity, since the normal curve really doesn't end. There is no  $\infty$  key, but the curve approaches zero very quickly. Enter normalcdf(2.08,1000), which should return approximately 0.0188. You could even enter normalcdf(2.08,10^99), but it doesn't really make a difference.

### Example

The Koi fish in a large pond have an average length of 12 inches, with a standard deviation of 1.5 inches. What percentage of the fish are smaller than 8 inches?

Fish that are smaller than 8 inches are more than 4 inches below the mean. Since the standard deviation is 1.5 inches, 4 inches represents  $4/1.5$  or about 2.6667 standard deviations. The z-score is:

$$z = \frac{X - \mu}{\sigma}$$

$$z = \frac{8 - 12}{1.5}$$

$$z \approx -2.6667$$

This z-score is negative because we are looking at a value below the mean. Converting that to a percentage, we see that the area under the normal curve from negative infinity to -2.6667 represents approximately 0.38% of the population. If there are 100 fish in the pond, that would

work out to 0.38 fish. Be careful about rounding that to zero, as it is still quite possible that there is actually a very small fish in the pond.

We can also start with a percentage, and find an associated value. Lets look at the lengths of the longest 20% of the fish in the example above. In this case we need the z-score associated with that percentage. Because we always count from left to right, we are really talking about finding a value at the 80% mark. If you are using a z-score table, you would look for the value in the body of the table that is closest to 0.8000, and then find the associated z-score. On my calculator, I go to DISTR (above the VARS key), and select invNorm(. Entering a value of 0.8 returns a z-score of .8416. This z-score is positive because it is 0.8416 standard deviations above the mean. Once you have a z-score you can find a value:

$$z = \frac{X - \mu}{\sigma}$$

$$0.8416 = \frac{X - 12}{1.5}$$

$$1.2624 = X - 12$$

$$X = 13.2624$$

The longest 20% of the fish are about 13 ¼ inches or longer.

## Margin of Error

Even if we select a perfectly random sample, it may not accurately represent the population it was taken from. We can calculate a **margin of error** by using the following formulas:

$$\text{Margin of Error} = Z \frac{s}{\sqrt{n}} \text{ or } Z \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}}$$

The first formula is for numerical population parameters. Here the sample standard deviation  $s$  is used to estimate the standard deviation for the population,  $\sigma$ . That works if the sample size is sufficiently large. Just use whatever standard deviation value is proved in your problem. If there is more variability in the population for the characteristic we are measuring, there is more



chance of error. The size of the sample is indicated by  $n$ . Notice that the margin of error will get smaller as the sample size gets larger.  $Z$  is the z-score associated with the **confidence interval**. If we want 99% of all possible samples to include the actual population parameter, within the margin of error, we have to use  $Z = 2.576$ . More commonly people will allow for 5% of the samples to be outside the limits given by the margin of error, by using  $Z = 1.96$ . This means that there is a 1 in 20 chance that the true value is not in the range indicated by your sample. For example, if a researcher finds that the average weight of a population of frogs is 200 grams plus or minus 10 grams, it is quite possible that the actual average weight of these frogs is 185 grams. He just happened to catch more fat frogs. However, we can say that it is more likely that the average weight is actually between 190 and 210 grams.

The second formula is used if you are looking for a proportion, such as the percentage of people who will vote for a certain presidential candidate.  $\hat{p}$  is the number of people in your sample who say they will vote for the candidate, divided by the total number of people in the sample. Polls commonly use a 95% confidence interval.

## Systems of Equations

If you have an equation with two unknowns, there is no way to get a specific numerical answer for one of the unknowns. However, if you have two equations containing the same two unknowns, you can usually find the values of both the unknowns. We can solve a system of equations by substitution, or by adding (or subtracting) the two equations - see Algebra 1 topics.

If you are given three equations containing three unknowns, you can still use the same strategies. [Graphically the solution represents the intersection of three planes, since we can draw a plane using an equation like  $z = x + 2y + 3$ . These problems have been carefully selected for you so that the intersection is always a single point.] Use two of the equations to eliminate one of the unknowns. Addition is usually the most efficient method here. Then use another combination of two equations and again eliminate the same unknown. This should give you two different statements containing two unknowns, which can now be solved. If the unknowns are  $x$ ,  $y$  and  $z$ , it is usually expected that you will eliminate  $x$  to get two equations containing  $y$  and  $z$ . Then you will use these two equations to obtain a value for  $z$ . The remaining variables are found by substituting the known values into the equations.

$$1) \quad 2x + y + 3z = 1$$

$$2) \quad 4x + 3y + 5z = 1$$

$$3) \quad 6x + 5y + 5z = -3$$

Eliminate  $x$  from equations 1 and 2 by multiplying equation 1 by  $-2$ , and using addition. Be very careful to multiply the **entire** equation by  $-2$ :

$$\begin{array}{r} -4x - 2y - 6z = -2 \\ \underline{4x + 3y + 5z = 1} \\ y - z = -1 \end{array}$$

Eliminate  $x$  from equations 2) and 3) by multiplying the second equation by 3 and the third equation 3 by  $-2$ :

$$\begin{array}{r} 12x + 9y + 15z = 3 \\ \underline{-12x - 10y - 10z = 6} \\ -y + 5z = 9 \end{array}$$

Now solve the two equations with two unknowns:

$$y - z = -1$$

$$\underline{-y + 5z = 9}$$

$$4z = 8$$

$$z = 2$$

Substitution then gives  $y = 1$  and  $x = -3$ . **Always check your answers!**

To create your own problems, simply pick some (small) values for  $x$ ,  $y$  and  $z$ , and create three equations. See if you can get your original numbers back when you solve the problem.

A graphing calculator or computer can solve systems of equations and help you figure out where you made those pesky little mistakes. Here is an online one:

<http://easycalculation.com/matrix/matrix-algebra.php>.

To use your TI-84 calculator press  $2^{\text{nd}}$  -> matrix-> edit, and enter  $3 \times 4$  as the size of the matrix. Enter only the coefficients of  $x$ ,  $y$  and  $z$ , and the number after the equals sign. For the problem

above, the matrix would look like this:  $\begin{bmatrix} 2 & 1 & 3 & 1 \\ 4 & 3 & 5 & 1 \\ 6 & 5 & 5 & -3 \end{bmatrix}$ . Press  $2^{\text{nd}}$  -> quit to leave the screen.

Return to the matrix menu and select MATH. The option you want is  $\text{rref}()$ , which is so far down that it is faster to scroll up. Now you have to enter the matrix you edited (usually matrix A), so press  $2^{\text{nd}}$  -> Matrix and select the proper matrix. When you press enter the solution appears in the following form:

$$\begin{bmatrix} 1 & 0 & 0 & -3 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 2 \end{bmatrix}.$$

This says that

$$1x + 0y + 0z = -3$$

$$0x + 1y + 0z = 1$$

$$0x + 0y + 1z = 2.$$

Systems of equations can also be solved if one or more of the equations are not linear. For example, we could solve

$$y = x^2$$

$$y = x$$

The solution would represent the point(s) where the curve and the line intersect. Because we are dealing with a curve and a line there are often two solution points.

You can solve this system of equations by simple substitution: if  $y = x^2$  and  $y = x$ , then  $x = x^2$ . Be careful when solving this, since you may be tempted to divide both sides of the equation by  $x$ . Before you do that, consider if  $x$  could be 0! In this case that will give you one solution. For the other one you may divide by  $x$  to get  $1 = x$ .

If you use subtraction you will get  $x^2 - x = 0$ . This might suggest to you that you can factor:  $x(x - 1) = 0$ , so either  $x = 0$  or  $x - 1 = 0$ .

For  $x = 0$ ,  $y = 0$  and for  $x = 1$ ,  $y = 1$ . The solutions are  $(0,0)$  and  $(1,1)$ .

A graph of both equations will show the intersect points.

## Linear Programming

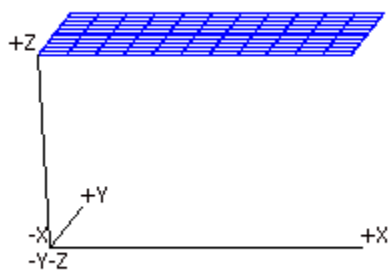
Linear programming problems involve maximizing or minimizing some quantity that depends on more than one variable. While real-world problems of this type are often complex and best solved by a computer program, the ones in your course will be simple enough to solve by hand. The most common type of problem involves maximizing profit where this profit depends on two variables. Usually both  $x$  and  $y$  must be greater than or equal to zero, and there are various other *constraints* that limit the values of the variables in some way.

To help us understand what it really going on here, we will look at a simple problem.

Joe's little factory produces white and neon-yellow ping pong balls. The maximum number of white balls that can be made is 8,000 ( $0 \leq x \leq 8,000$ ), and the largest number of yellow balls that can be produced is 4,000 ( $0 \leq y \leq 4,000$ ). Joe's father is very wealthy and owns a chain of sports supply stores. Each month Joe's father buys all of Joe's ping pong balls and pays him whatever it costs to make them, plus \$5,000. Find the maximum profit.

In this odd but very simple arrangement the profit  $P$  is always 5,000, and this is represented by the equation  $P = 5,000$

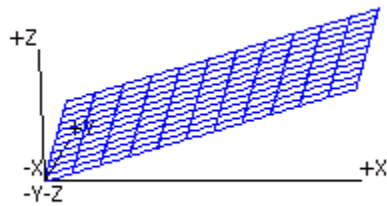
Because we are working with 3 variables in these problems, we need a 3-dimensional graph to really represent them properly. We use a  $z$ -axis that is perpendicular to the  $x$ - $y$  plane, and graph  $P$  on the  $z$ -axis. Since  $P$  is always 5,000 in this problem, the graph is a level plane that sits at a distance of 5,000 above the  $x$ - $y$  plane.



This graph was created with MathGV, using the  $p$ ,  $x$  and  $y$  values in thousands. The plane representing  $P$  appears at a distance of +5,000 along the  $z$ -axis. The boundaries of the plane are  $x = 0$ ,  $x = 8,000$ ,  $y = 0$  and  $y = 4,000$ .

For this particular problem there is no minimum or maximum value for  $P$ . The plane is level and  $P$  is the same everywhere.

One day Joe and his father have a fight, and Joe decides to sell his ping pong balls to a rival store instead. After looking at the numbers, Joe realizes that he can make a profit of 30 cents on the white ping pong balls, and 25 cents on the yellow ones. The equation that represents the total profit is now  $P = .30x + .25y$ . Find the values for  $x$  and  $y$  that will give the maximum profit, assuming that  $0 \leq x \leq 8,000$  and  $0 \leq y \leq 4,000$ .



The profit  $P$  can still be represented by a plane because this is a linear equation in  $x$  and  $y$ , but the plane is tilted. Again the boundaries of the plane are  $x = 0$ ,  $x = 8,000$ ,  $y = 0$  and  $y = 4,000$ . The minimum value of  $P$  is at the lowest corner of the plane, which corresponds to  $x = 0$  and  $y = 0$ . That really makes sense, since the lowest profit will occur when no ping pong balls are produced. The largest value of  $P$  is at the top corner of the plane, which corresponds to  $x = 8,000$  and  $y = 4,000$  on the graph. This means that  $x$  should be 8,000 and  $y$  should be 4,000 for maximum profit. That makes sense too since producing the maximum number of ping pong balls would make the most profit for Joe.

Grab yourself a plane so you can try this out. The top of a box is most convenient. No matter how you tilt the plane, if there is a single highest point it is always at one of the corners. The corners of the plane are easy to find mathematically, since they are located where the boundary lines intersect. In our simple case, the top corner of the plane is sitting above the intersection of the lines  $x = 8,000$  and  $y = 4,000$ . Plug these values into the equation for  $P$  to get 3400.

In general, this works for non-rectangular planes too. All we need to do is find the boundary lines and see where they intersect, so we can find all the corners of the plane above them. Then we have to test out the different values of  $P$  at those points to see where  $P$  is the greatest.

## Matrices

Matrices were already used by the Chinese sometime before the year 200. They were developed to solve systems of linear equations. You are already familiar with simple systems like

$$\begin{aligned} 5x + 3y &= 11 \\ x + 4y &= 9 \end{aligned}$$

When you have more complicated systems of linear equations, they can be solved more efficiently using matrices. Computer programmers find matrices handy to store and manipulate data. Another use for matrices is to create secret codes that are just about impossible to break.

Matrices are described by the number of rows and columns they have. The matrix  $\begin{bmatrix} 2 & 0 & 5 \\ -3 & 1 & 7 \end{bmatrix}$  has two rows and three columns, so it is a 2 x 3 (read as 2 by 3) matrix. In general, an  $m \times n$  matrix has  $m$  rows and  $n$  columns.

If two matrices have the same dimensions, you can add or subtract them by adding or subtracting their entries.

## Solving Systems of Linear Equations with Matrices

Let's see how we can use matrices to solve the simple system of equations mentioned above:

$$\begin{aligned} 5x + 3y &= 11 \\ x + 4y &= 9 \end{aligned}$$

The matrix associated with this system is called a **coefficient matrix** because it is made up of the coefficients of the variables. The coefficient matrix is  $\begin{bmatrix} 5 & 3 \\ 1 & 4 \end{bmatrix}$ . The entire system can be put into an **augmented matrix**, like this:  $\left[ \begin{array}{cc|c} 5 & 3 & 11 \\ 1 & 4 & 9 \end{array} \right]$

Matrices may be manipulated to solve a system of equations. Valid manipulations are

1. Exchanging two rows
2. Multiplying a row by a constant
3. Changing a row by adding or subtracting a multiple of another row

You can perform any of these manipulations in any order until you get the desired result.

For example, for the matrix  $\begin{bmatrix} 5 & 3 & | & 11 \\ 1 & 4 & | & 9 \end{bmatrix}$ , we can exchange the two rows to get  $\begin{bmatrix} 1 & 4 & | & 9 \\ 5 & 3 & | & 11 \end{bmatrix}$ .

Then we can change the bottom row by subtracting 5 times the top row from it. That gives us  $\begin{bmatrix} 1 & 4 & | & 9 \\ 0 & -17 & | & -34 \end{bmatrix}$ . Notice that the top row remains unchanged.

Now divide the bottom row by -17, which is the same as multiplying by  $-\frac{1}{17}$ , to get  $\begin{bmatrix} 1 & 4 & | & 9 \\ 0 & 1 & | & 2 \end{bmatrix}$ .

When you change this augmented matrix back into a system of equations, you can find the solution:

$$x + 4y = 9$$

$$0x + 1y = 2$$

Because  $y = 2$ ,  $x$  must be equal to 1. Notice that we are able to obtain the solution easily now because the coefficient of  $x$  in the second row is zero, and the coefficient of  $y$  is 1. The first row conveniently has a coefficient of 1 for  $x$ . However, there is still a calculation involved.

The **Reduced Row Echelon Form** allows you to read the solution to a system of equations directly from the resulting matrix. You can ask your graphing calculator to get it for you (rref in the "MATRIX" section under "MATH"), or find it by hand. To see how to do it manually we will look at an example:

$$2x + 3y = 13$$

$$x + 2y = 8$$

Ignore the variables, and enter the numbers into an augmented matrix:

$$\begin{bmatrix} 2 & 3 & | & 13 \\ 1 & 2 & | & 8 \end{bmatrix}$$

The line always takes the place of the equals sign. (Your graphing calculator will not need a line to solve the system, so you would just enter the data into a 2 x 3 matrix.) We want to start by subtracting two times the second row from the first to get

$$\begin{bmatrix} 0 & -1 & | & -3 \\ 1 & 2 & | & 8 \end{bmatrix}$$

Then we reverse the rows:

$$\begin{bmatrix} 1 & 2 & | & 8 \\ 0 & -1 & | & -3 \end{bmatrix}$$



Next we would like to have a zero as the second entry in the first row, so we add two times the second row to the first:

$$\left[ \begin{array}{cc|c} 1 & 0 & 2 \\ 0 & -1 & -3 \end{array} \right]$$

Multiply the second row by -1 to get a positive leading entry:

$$\left[ \begin{array}{cc|c} 1 & 0 & 2 \\ 0 & 1 & 3 \end{array} \right]$$

Now convert this augmented matrix back to a system of equations to quickly see the solution:

$$1x + 0y = 2$$

$$0x + 1y = 3$$

If you have a system of three linear equations, you can use the same methods. The reduced row echelon form will look something like this:

$$\left[ \begin{array}{ccc|c} 1 & 0 & 0 & 3 \\ 0 & 1 & 0 & -7 \\ 0 & 0 & 1 & 4 \end{array} \right]$$

For each of the variables  $x$ ,  $y$ , and  $z$  you can read the value directly from the resulting matrix.

## Matrix Multiplication

You can multiply a matrix by a constant, just by multiplying each entry in the matrix by that constant.

To multiply two matrices, you **multiply the rows of the first matrix by the columns of the second matrix**. That seems rather arbitrary and confusing, but we can look at some history to see the reason for it. Back in the 19<sup>th</sup> century the mathematician Arthur Cayley was interested in changing points  $(x, y)$  to new points  $(x', y')$  in an orderly way. These changes are called transformations, and they work like this:

$$x' = ax$$

$$y' = dy$$

This uses information from the original  $x$  coordinate to transform it to a new  $x$  coordinate, and information from the original  $y$  coordinate to transform it to a new  $y$  coordinate.  $a$  and  $d$  represent constants that accomplish the desired transformation. We can also use information

from both the original x and y coordinate to determine the new x coordinate, and do the same for the new y coordinate:

$$x' = ax + by$$

$$y' = cx + dy$$

Here a, b, c and d are constants that accomplish this slightly more complicated transformation. Cayley used matrices as a shorthand notation for these transformations (you can imagine that he would get tired of writing the whole thing out completely each time). He just used the matrix  $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$  to describe the transformation. Now we have a systematic way to transform points (x, y) or vectors  $\begin{bmatrix} x \\ y \end{bmatrix}$  into other points or vectors in a predictable and often useful way.

Let's look closely at some specific examples. Set up a coordinate system and draw an arrow from the point (0,0) to the point (3,4). By the Pythagorean Theorem, the length of this arrow will be 5 units. Next, we will reflect the arrow over a line. Here the line is looked at as a mirror, and each new point is the mirror image of the original point. If you reflect over the line  $y = x$ , what happens is that each x-coordinate turns into a y-coordinate and vice versa. Try it out. Our original arrow from (0,0) to (3,4) changes to an arrow from (0,0) to (4,3). We can write that as:

$$x' = 0x + 1y$$

$$y' = 1x + 0y$$

The new x-coordinate is the old y-coordinate, and vice versa. We could write the transformation we just did as  $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ .

Now let's take the original arrow and rotate it clockwise by 90 degrees. The tip of the arrow is at (3,4), and after a 90 degree rotation the tip is at (4,-3), while the origin remains at (0,0). Notice that the original y coordinate of the tip has become the x coordinate, and the x coordinate has been multiplied by -1 and becomes the y coordinate. This transformation looks like this:

$$x' = 0x + 1y$$

$$y' = -1x + 0y$$

The point (0,0) is unchanged, and (3,4) is transformed into (4,-3)

If we apply this same transformation to our new arrow, it rotates another 90 degrees clockwise:

$$x'' = 0x' + 1y'$$

$$y'' = -1x' + 0y'$$

Substituting for  $x'$  and  $y'$ , we could write this as one single transformation:

$$x'' = 0(0x + 1y) + 1(-1x + 0y)$$

$$y'' = -1(0x + 1y) + 0(-1x + 0y)$$

which gives

$$x'' = 0x + 0y - 1x + 0y$$

$$y'' = 0x - 1y + 0x + 0y$$

so,

$$x'' = -1x + 0y$$

$$y'' = 0x - 1y$$

This rotates our original arrow 180 degrees to put its tip at (-3,-4).

Looking at the general case, two successive transformations can be written as

$$x' = ax + by$$

$$y' = cx + dy$$

$$x'' = ex' + fy'$$

$$y'' = gx' + hy'$$

where  $e$ ,  $f$ ,  $g$ , and  $h$  are the set of numbers that accomplish the second transformation. The

shorthand for this second transformation would be  $\begin{bmatrix} e & f \\ g & h \end{bmatrix}$ .

If we take these two transformations together, we see that

$$x'' = e(ax + by) + f(cx + dy)$$

$$y'' = g(ax + by) + h(cx + dy)$$

(I just used copy and paste to substitute for  $x'$  and  $y'$ ). Multiplying that out and rearranging to a new matrix form, we get:

$$x'' = (ae + cf)x + (be + df)y$$

$$y'' = (ag + ch)x + (bg + dh)y$$

The product of the two transformations can be written as the new transformation matrix

$$\begin{bmatrix} ae + cf & be + df \\ ag + ch & bg + dh \end{bmatrix}$$

This leads us to the idea that we could multiply the two transformation matrices to get

$$\begin{bmatrix} e & f \\ g & h \end{bmatrix} \cdot \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} ae + cf & be + df \\ ag + ch & bg + dh \end{bmatrix}$$

To get the first entry in the first column, multiply the first row by the first column.

To get the second entry in the first column, multiply the second row by the first column.

To get the first entry in the next column, multiply the first row by the second column.

To get the second entry in the next column, multiply the second row by the second column.

Because of the way matrix multiplication is done it is not always possible to multiply two matrices. The number of columns in the first matrix must match the number of rows in the second matrix.

Notice that the way this is set up is that when you multiply one matrix by a second, you put the second one in front. This is called pre-multiplying. Matrix multiplication is not commutative, which means that you don't get the same result if you do it the other way around.

Going back to our specific example of rotating an arrow, we can multiply the two matrices like this:

$$\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} = \begin{bmatrix} 0 \cdot 0 + 1 \cdot -1 & 0 \cdot 1 + 1 \cdot 0 \\ -1 \cdot 0 + 0 \cdot -1 & -1 \cdot 1 + 0 \cdot 0 \end{bmatrix} = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}$$

This is the same transformation matrix we obtained above.

## Determinants

The first point to note about determinants is that they only apply to square matrices. When there is a square coefficient matrix in a system of linear equations that system has the same number of statements as the number of unknowns. This is usually the ideal situation for getting a unique solution for the system. If there are fewer statements than unknowns the system will not have a unique solution. If there are more statements than unknowns the system is called *overdetermined*, and the additional statements may cause the system to be inconsistent so that there is no solution.

Even if it is associated with a square matrix, a system of linear equations may not have a unique solution. Consider the system

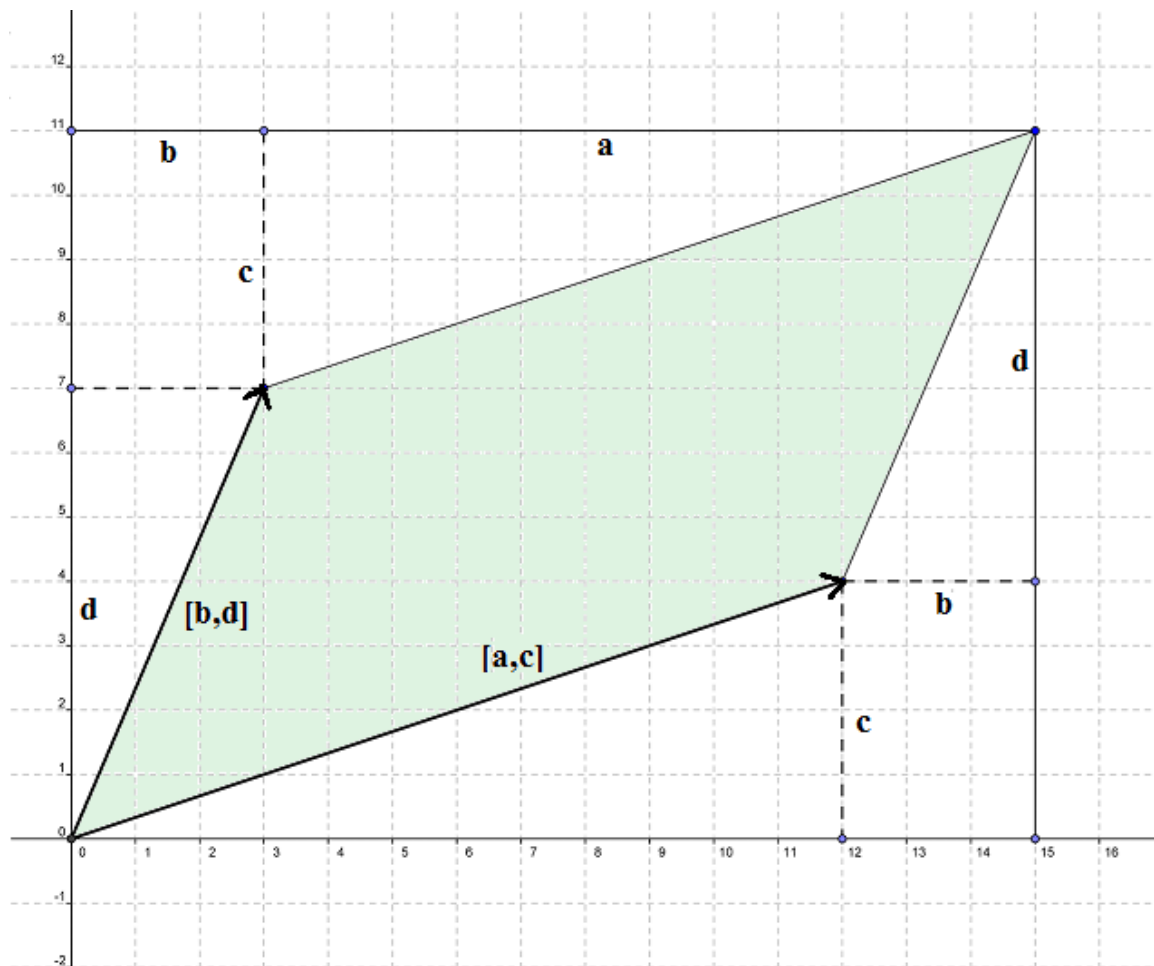
$$2x + 3y = 10$$

$$4x + 6y = 20$$

It has a square coefficient matrix associated with it,  $\begin{bmatrix} 2 & 3 \\ 4 & 6 \end{bmatrix}$ , and if you try to solve it you get  $0 = 0$ . For any  $x$  you select, there is a corresponding value of  $y$  that makes this system of equations true. The fact that there is no unique solution can be easily determined beforehand. I have simply taken the first equation and multiplied it by 2 to get the second. This makes 2 and 3 nicely balanced with 4 and 6. In fact, 2 times 6 = 3 times 4, or  $2 \times 6 - 3 \times 4 = 0$ . The numbers that come after the equals sign don't really matter if we are just looking to see whether the system has a *unique* solution. If the 10 and the 20 balance in the same way as the left-hand parts of the equations, then the system has an infinite number of solutions. If they don't balance in that way, there are no solutions at all. We can see if there is a unique solution or not just by looking at the coefficient matrix  $\begin{bmatrix} 2 & 3 \\ 4 & 6 \end{bmatrix}$ . In general, for any matrix  $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$ , if  $ad - bc = 0$  then the corresponding system of equations does not have a unique solution.  $ad - bc$  is called the **determinant** of this 2 by 2 matrix.

Actually, determinants make sense even in unexpected ways. If you look at a random matrix  $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$  as being composed of two vectors,  $\begin{bmatrix} a \\ c \end{bmatrix}$  and  $\begin{bmatrix} b \\ d \end{bmatrix}$ , the absolute value of the determinant turns out to be the area of the parallelogram defined by these two vectors.

This image shows two vectors  $\begin{bmatrix} a \\ c \end{bmatrix}$  and  $\begin{bmatrix} b \\ d \end{bmatrix}$ . The area of the parallelogram they define can be calculated:



The total area of the large rectangle is  $(a + b)(c + d)$ . From this we subtract the two small rectangles that have area  $bc$ , and also the four triangles. Two of the triangles have area  $\frac{1}{2} ac$ , and the other two have area  $\frac{1}{2} bd$ . The area of the parallelogram is  $(a + b)(c + d) - 2bc - ac - bd$ , which is  $ac + ad + bc + bd - 2bc - ac - bd = ad - bc$ . Then add an absolute value sign because area is always positive even when the vectors have negative components:  $|ad - bc|$ .

If the determinant is zero, this area is zero because the vectors are parallel. Just draw the vectors  $\begin{bmatrix} 2 \\ 4 \end{bmatrix}$  and  $\begin{bmatrix} 3 \\ 6 \end{bmatrix}$  from the example above so you can see what is happening.

Another interesting fact is that the linear transformation described by the matrix  $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$  will transform a figure (by moving its defining points) into a figure with an area given by the original area times the absolute value of the determinant of the matrix.

Even more amazing is that both of the above facts also apply to the determinants of  $3 \times 3$  matrices if we consider volume rather than area. The determinant of an  $n \times n$  matrix that is even larger represents the volume of a figure in  $n$ -dimensional space.

---

The determinants of larger matrices are more difficult to find, but fortunately we have calculators that can do that for us. If you don't have a graphing calculator handy you can still find the determinant manually.

### Example

Find the determinant of  $\begin{bmatrix} 2 & 0 & 4 \\ 3 & 1 & 2 \\ 5 & 4 & 6 \end{bmatrix}$ .

The determinant of a matrix is usually indicated by placing lines on either side of the numbers,

like this:  $\begin{vmatrix} 2 & 0 & 4 \\ 3 & 1 & 2 \\ 5 & 4 & 6 \end{vmatrix}$ .

First, select a row or column of your square matrix, preferably one with one or more zeros in it. Usually people select the first row or column so they don't get mixed up, and for this example I will use the first row. Next, look at the first entry in your row, which is 2 in this case. This entry has the general form  $a_{11}$ , where the subscripts mean the first row and the first column. Add these subscripts and note that the result is an even number. Next, cross out all of the entries in both the row and the column that this entry belongs in. That leaves you with  $\begin{vmatrix} 1 & 2 \\ 4 & 6 \end{vmatrix}$ . The determinant of this smaller matrix is called the **minor** of 2. Multiply 2 by its minor,  $2 \cdot \begin{vmatrix} 1 & 2 \\ 4 & 6 \end{vmatrix}$ . The next entry in the row is 0, which is  $a_{12}$ , indicating that it is in the first row and the second column. When you add these two subscripts you get an odd number, so we will *subtract* the product of this entry and its remaining determinant. After crossing out, we look for the determinant of  $\begin{vmatrix} 3 & 2 \\ 5 & 6 \end{vmatrix}$ . So far our calculation reads:

$$2 \cdot \begin{vmatrix} 1 & 2 \\ 4 & 6 \end{vmatrix} - 0 \cdot \begin{vmatrix} 3 & 2 \\ 5 & 6 \end{vmatrix} \dots$$

The last entry in the row is 4, which is entry  $a_{13}$ . Adding its subscripts gives an even number so we *add* the product of this entry and its remaining determinant  $\begin{vmatrix} 3 & 1 \\ 5 & 4 \end{vmatrix}$ , to get:

$2 \cdot \begin{vmatrix} 1 & 2 \\ 4 & 6 \end{vmatrix} - 0 \cdot \begin{vmatrix} 3 & 2 \\ 5 & 6 \end{vmatrix} + 4 \cdot \begin{vmatrix} 3 & 1 \\ 5 & 4 \end{vmatrix}$  which is equal to  $2 \cdot -2 - 0 + 4 \cdot 7 = 24$ . (Choosing the row with the 0 allowed me to skip calculating one of the determinants).

In general, we find the determinant of a square  $n \times n$  matrix  $A$  as follows:  $M_{ij}$  is the smaller matrix obtained by deleting the row and column containing entry  $a_{ij}$  (where  $a_{ij}$  is the entry in the  $i$ th row and  $j$ th column). The **cofactor** of entry  $a_{ij}$  is  $A_{ij}$ , which is equal to  $(-1)^{i+j} \cdot \det(M_{ij})$ . Choose a row or column and multiply each successive entry by its cofactor.

If we choose row  $i$ , then  $\det(A) = a_{i1}A_{i1} + a_{i2}A_{i2} + a_{i3}A_{i3} + \dots + a_{in}A_{in}$

If we choose column  $j$ , then  $\det(A) = a_{1j}A_{1j} + a_{2j}A_{2j} + a_{3j}A_{3j} + \dots + a_{nj}A_{nj}$

Note that we could use this procedure with any size square ( $n \times n$ ) matrix, but when a minor is larger than  $2 \times 2$  we have to find its determinant by using its entries multiplied by their cofactors, etc, etc. until the minors are small enough. This can get very tedious. 🤖

If a matrix has one row or column consisting of all zeros, its determinant must be zero. You can see this by using this row or column to calculate the determinant, or by thinking about this matrix as the coefficient matrix of a linear system that does not have a unique solution. For example, the system

$$x + 3y + 4z = 12$$

$$3x - y - 2z = 9$$

will not have a unique solution. It has an associated square coefficient

matrix  $\begin{bmatrix} 1 & 3 & 4 \\ 3 & -1 & -2 \\ 0 & 0 & 0 \end{bmatrix}$  with a determinant of 0.

## Cramer's Rule

We have seen that a system of linear equations can be solved by manipulating rows of its augmented matrix. Some people, like me, find that when they do this they inevitably make at least one small mistake and their solution doesn't work out. Gabriel Cramer lived in the 18<sup>th</sup> century, so he didn't have a graphing calculator. He came up with a clever way to solve systems of linear equations by just using determinants. Cramer's rule uses a simple method, which is appreciated by a lot of students who are not allowed to use calculators for their exams. It also turned out to be very useful for solving other problems in advanced mathematics.



To use Cramer's rule, first find the determinant of the coefficient matrix associated with the system. Cramer's rule works only if your system has a unique solution (the determinant of the coefficient matrix is not 0). As an example, we will solve the system

$$x + 2y + z = 5$$

$$2x + 2y + z = 6$$

$$x + 2y + 3z = 9$$

Although there are only three unknowns here, I want to emphasize that Cramer's Rule works for larger systems too. (Because larger systems are much more tedious to solve with this method you will hopefully never see them in homework or on exams.) We will call the

coefficient matrix of the system "A". The determinant of A,  $\begin{vmatrix} 1 & 2 & 1 \\ 2 & 2 & 1 \\ 1 & 2 & 3 \end{vmatrix}$ , works out to -4.

Now we will solve for each unknown in turn. To find the value of x, replace the first column of the coefficient matrix by the column on the right side of the equals sign,  $\begin{bmatrix} 5 \\ 6 \\ 9 \end{bmatrix}$ , so you get

$\begin{bmatrix} 5 & 2 & 1 \\ 6 & 2 & 1 \\ 9 & 2 & 3 \end{bmatrix}$ . We will call this new matrix  $A_x$  because we replaced the coefficients of x. It is also often called  $A_1$ , because the variables used for larger systems are  $x_1, x_2, x_3, x_4$ , etc. instead of just x, y and z. Now find the determinant of this new matrix, which is also -4. To get x divide the determinant of the new matrix by the determinant of the original matrix:

$$x = \frac{\det(A_x)}{\det(A)} = \frac{-4}{-4} = 1$$

To find the next unknown, y, we replace the coefficients of y by  $\begin{bmatrix} 5 \\ 6 \\ 9 \end{bmatrix}$ , to get the new matrix  $A_y$ ,

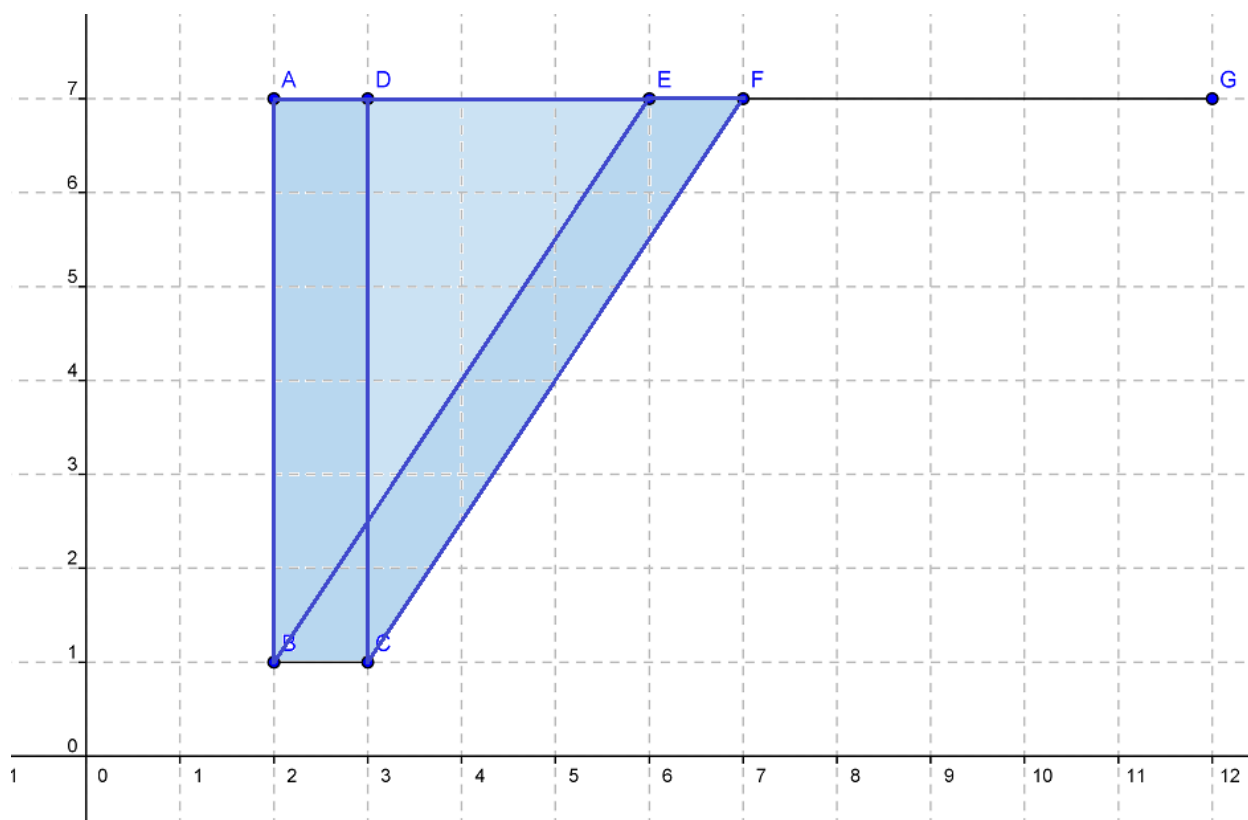
which looks like this:  $\begin{bmatrix} 1 & 5 & 1 \\ 2 & 6 & 1 \\ 1 & 9 & 3 \end{bmatrix}$ .  $y = \frac{\det(A_y)}{\det(A)} = -4$ .

The last unknown, z, is equal to  $\frac{\det(A_z)}{\det(A)}$ , where  $A_z$  is obtained by replacing the coefficients of z by  $\begin{bmatrix} 5 \\ 6 \\ 9 \end{bmatrix}$ . The determinant of  $A_z$ ,  $\begin{bmatrix} 1 & 2 & 5 \\ 2 & 2 & 6 \\ 1 & 2 & 9 \end{bmatrix}$  is -8.  $z = \frac{-8}{-4} = 2$

---

### A Closer Look at Cramer's Rule

This explanation uses a principle that you may not be familiar with, so we will consider it separately. Imagine that you have a stack of ten square crackers. If you move each cracker very slightly to one side, you will have a slanted stack of crackers that is the same height as the original stack. It is relatively easy to imagine that the volume of each stack is the same, because the amount of food is unchanged. However, it may be more difficult to see that the area of the sides is also unchanged. Consider the following image:



The area of strip ABCD is the same as the area of strip BCEF. To see that this is true, you can visualize the colored area as being made up of triangle ABE and strip BCEF. However, you can also see it as a combination of strip ABCD and triangle DCF. Triangle ABE and triangle DCF are the same size, so the two strips must have the same area.

Cramer's Rule can be derived using the principles of matrix and linear algebra, but it can also be shown to be true geometrically. Here we will consider a geometric proof for the simple case of a linear system with two equations, but the geometry also works in three dimensions and even beyond. Here is a random system of two linear equations:

$$2x + 4y = 10$$

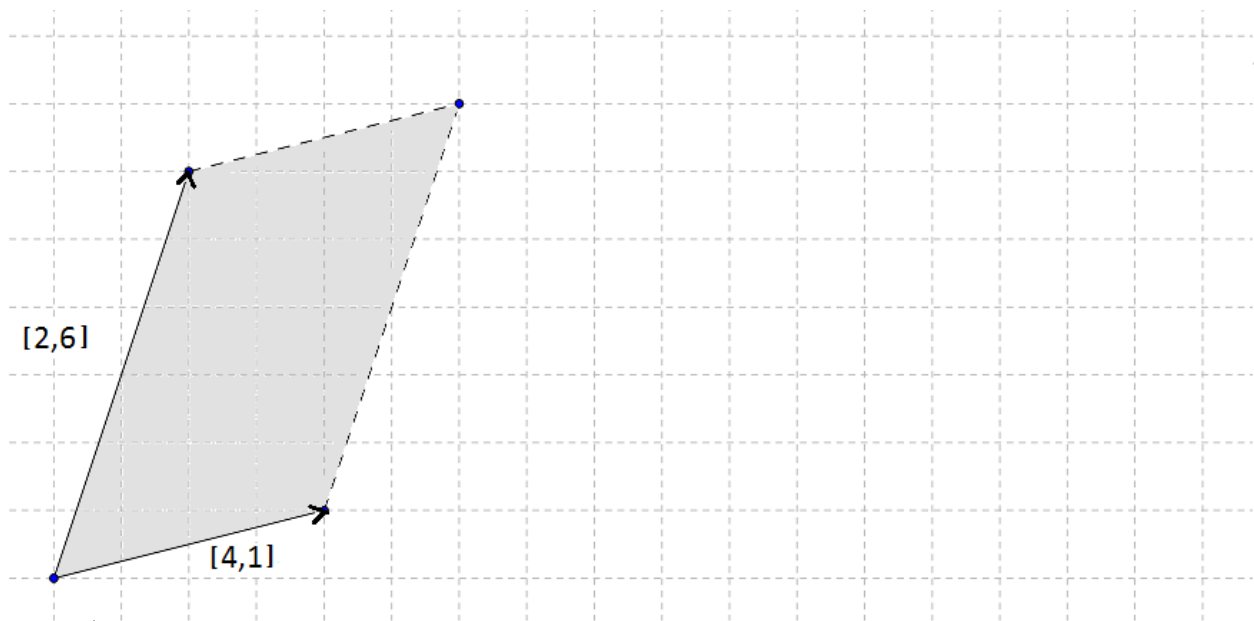
$$6x + 1y = 13.5$$

The determinant of the coefficient matrix involves two vectors,  $\begin{bmatrix} 2 \\ 6 \end{bmatrix}$  and  $\begin{bmatrix} 4 \\ 1 \end{bmatrix}$ . The system could be considered as a **linear combination** of these two vectors:

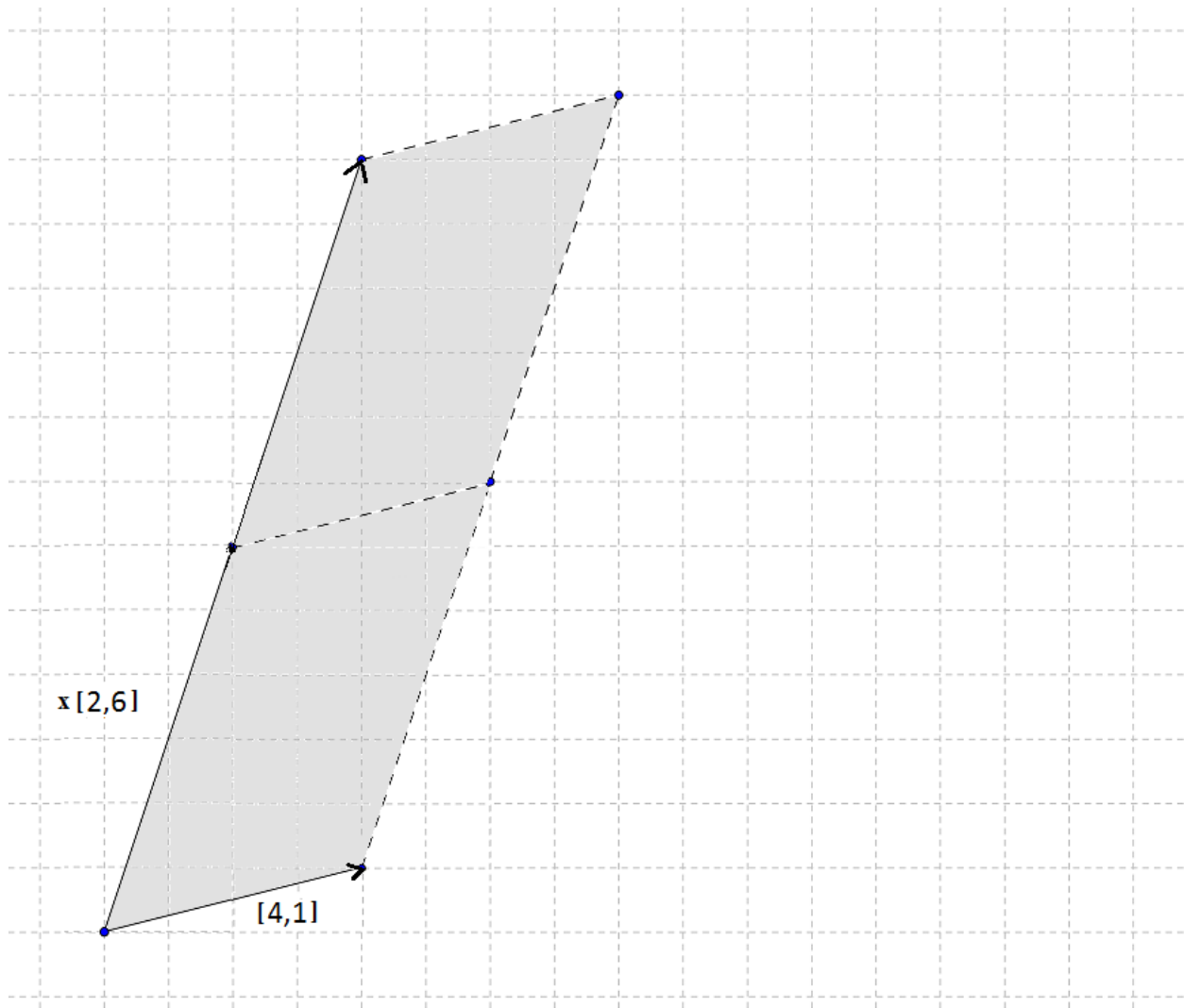
$$x \begin{bmatrix} 2 \\ 6 \end{bmatrix} + y \begin{bmatrix} 4 \\ 1 \end{bmatrix} = \begin{bmatrix} 10 \\ 13.5 \end{bmatrix}.$$

Here the vectors are multiplied by  $x$  and  $y$ , which are just numbers.

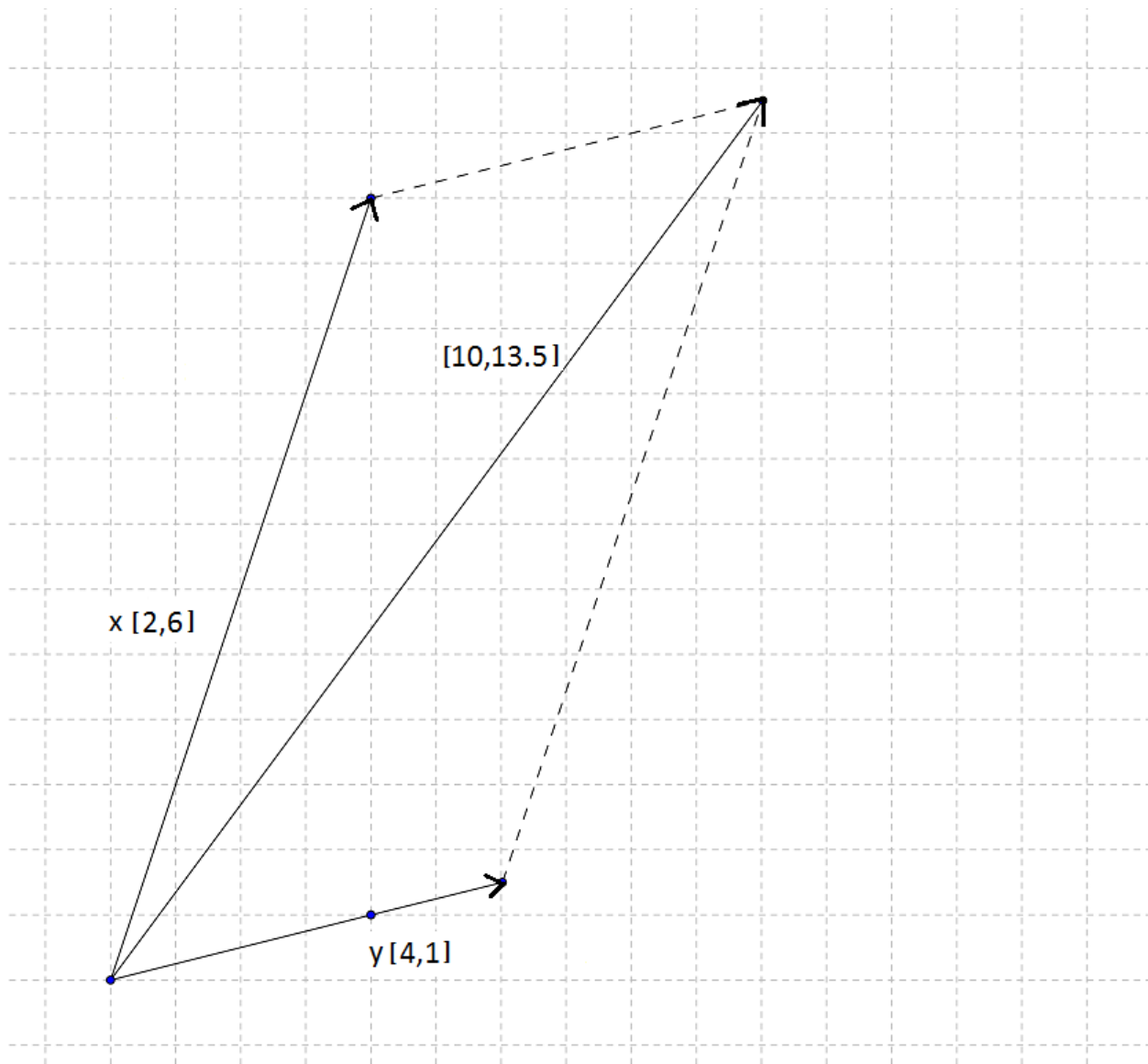
As we saw before, the absolute value of the determinant of the coefficient matrix,  $\begin{vmatrix} 2 & 4 \\ 6 & 1 \end{vmatrix}$  is the area of the parallelogram determined by  $\begin{bmatrix} 2 \\ 6 \end{bmatrix}$  and  $\begin{bmatrix} 4 \\ 1 \end{bmatrix}$ :



The area of the parallelogram determined by  $x \begin{bmatrix} 2 \\ 6 \end{bmatrix}$  and  $\begin{bmatrix} 4 \\ 1 \end{bmatrix}$  has to be  $x$  times larger since one of the sides has been multiplied by  $x$ . ( $x$  is actually 2, and you can see for yourself that the area is in fact twice as big.)

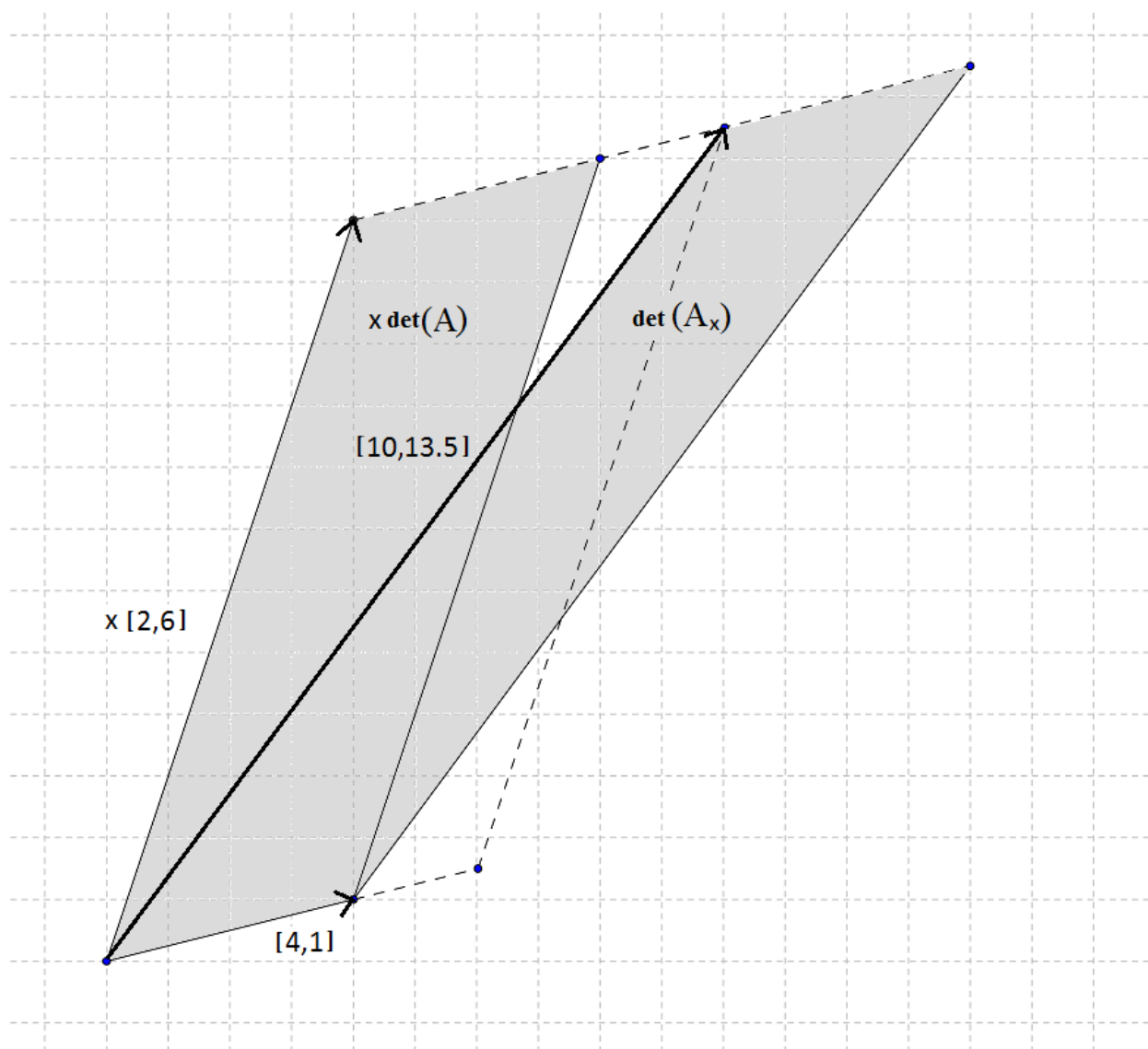


This area represents  $x \cdot \det(A)$ . Now we need to draw the vector  $\begin{bmatrix} 10 \\ 13.5 \end{bmatrix}$ , which is equal to  $x \begin{bmatrix} 2 \\ 6 \end{bmatrix} + y \begin{bmatrix} 4 \\ 1 \end{bmatrix}$ . Adding two vectors is simple, because a vector is just a  $2 \times 1$  matrix. Matrices are added by simply adding their entries. For example, if you add the vectors  $\begin{bmatrix} 3 \\ 11 \end{bmatrix}$  and  $\begin{bmatrix} 2 \\ 4 \end{bmatrix}$ , you would get the vector  $\begin{bmatrix} 5 \\ 15 \end{bmatrix}$ . You could also think of this as the vector  $\begin{bmatrix} 3 \\ 11 \end{bmatrix}$  taking you 3 units to the right and 11 units up. The vector  $\begin{bmatrix} 2 \\ 4 \end{bmatrix}$  would then take you two additional units to the right and another 4 units up. The next image shows the vector  $\begin{bmatrix} 10 \\ 13.5 \end{bmatrix}$ , created by adding  $x \begin{bmatrix} 2 \\ 6 \end{bmatrix}$  and  $y \begin{bmatrix} 4 \\ 1 \end{bmatrix}$ :



You can see from this picture that it should be possible to find  $x$  and  $y$  geometrically, but it actually takes a clever trick to do it properly.

The matrix  $A_x$  is  $\begin{bmatrix} 10 & 4 \\ 13.5 & 1 \end{bmatrix}$ , because we replaced the first column by the numbers found to the right of the equals sign. The determinant of  $A_x$  represents the area of the parallelogram defined by  $\begin{bmatrix} 10 \\ 13.5 \end{bmatrix}$  and  $\begin{bmatrix} 4 \\ 1 \end{bmatrix}$ . This area is marked “ $\det(A_x)$ ” in the image below.



Here you can see that  $x \cdot \det(A) = \det(A_x)$ . Therefore,  $x = \frac{\det(A_x)}{\det(A)}$ , and a similar argument would show that  $y = \frac{\det(A_y)}{\det(A)}$ .

## Identity Matrices

If we think of the coordinates  $(x,y)$  as the vector  $\begin{bmatrix} x \\ y \end{bmatrix}$  (or the  $2 \times 1$  matrix  $\begin{bmatrix} x \\ y \end{bmatrix}$ ), then the new vector  $\begin{bmatrix} x' \\ y' \end{bmatrix}$  that results from a transformation would be obtained by multiplying the original

vector by the transformation matrix:

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} ax + by \\ cx + dy \end{bmatrix}$$

Again, notice the order in which this multiplication occurs.

If we follow the first transformation with a second one, the second vector  $\begin{bmatrix} x'' \\ y'' \end{bmatrix}$  is obtained by multiplying  $\begin{bmatrix} e & f \\ g & h \end{bmatrix} \cdot \begin{bmatrix} a & b \\ c & d \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix}$ .

Matrix multiplication is associative, so you can either multiply the first two matrices first, or the last two first. Either way you get

$$\begin{bmatrix} x'' \\ y'' \end{bmatrix} = \begin{bmatrix} (ae + cf)x + (be + df)y \\ (ag + ch)x + (bg + dh)y \end{bmatrix}$$

We can also multiply a vector  $\begin{bmatrix} x \\ y \end{bmatrix}$  by the **identity matrix**,  $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ . This does not change the original vector at all:  $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 1x + 0y \\ 0x + 1y \end{bmatrix} = \begin{bmatrix} x \\ y \end{bmatrix}$ .

In fact, any suitably sized matrix may be pre-multiplied, and also post-multiplied if possible, by an identity matrix, and it will not change. Matrices are usually represented by a capital letter, and the identity matrix is represented by the letter I. We can write that  $A \cdot I = A$ , and  $I \cdot A = A$

Larger identity matrices also exist, like  $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ . Here we are adding an additional dimension. We can pre-multiply the vector  $\begin{bmatrix} x \\ y \\ z \end{bmatrix}$  by this larger identity matrix, and it will not change. Identity matrices are always square, which means that they have the same number of rows as columns.

## Inverse Matrices

Many square matrices have an “inverse”,  $A^{-1}$ , such that  $A^{-1}A = I$ , and also  $AA^{-1} = I$ .

Inverse matrices can be used to solve a system of equations. Suppose we have a system of 2 equations like this:

$$2x + 3y = 13$$

$$x + 2y = 8$$

We can represent this system by a matrix equation as follows:

$$\begin{bmatrix} 2 & 3 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 13 \\ 8 \end{bmatrix}$$

If we were to pre-multiply both sides of the equation by the inverse of  $\begin{bmatrix} 2 & 3 \\ 1 & 2 \end{bmatrix}$ , we would get:

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = (\text{the inverse of } \begin{bmatrix} 2 & 3 \\ 1 & 2 \end{bmatrix}) \begin{bmatrix} 13 \\ 8 \end{bmatrix}$$

The system of equations then turns into

$$1x + 0y = \dots$$

$$0x + 1y = \dots$$

and the values for  $x$  and  $y$  can simply be read directly from the equations. This gives us a relatively easy way to solve a system of equations.

Note that not every matrix has an inverse. In particular, if your system of equations does not have a solution you can be sure that the associated matrix does not have an inverse. That makes sense, because if there was an inverse matrix we should be able to use it to get a solution. Matrices that do not have an inverse are called **singular** matrices.

There is an interesting way to find the inverse of a matrix. Since we know that  $A^{-1}A = I$ , we can write the matrix  $A$  next to the matrix  $I$ , and manipulate both of them at the same time until we manage to turn  $A$  into  $I$ . At that point, we know that we have multiplied  $A$  by  $A^{-1}$  through our manipulations, and we have done the same to  $I$ . Since  $A^{-1}I = A^{-1}$ , the inverse matrix will automatically appear where  $I$  used to be. In our example, we would write:

$$\left[ \begin{array}{cc|cc} 2 & 3 & 1 & 0 \\ 1 & 2 & 0 & 1 \end{array} \right]$$

Multiply the second row by 2 and subtract it from the first row:

$$\left[ \begin{array}{cc|cc} 0 & -1 & 1 & -2 \\ 1 & 2 & 0 & 1 \end{array} \right]$$

Reverse the top and bottom rows:

$$\left[ \begin{array}{cc|cc} 1 & 2 & 0 & 1 \\ 0 & -1 & 1 & -2 \end{array} \right]$$



Multiply the bottom row by 2 and add it to the top row:

$$\left[ \begin{array}{cc|cc} 1 & 0 & 2 & -3 \\ 0 & -1 & 1 & -2 \end{array} \right]$$

Multiply the bottom row by -1:

$$\left[ \begin{array}{cc|cc} 1 & 0 & 2 & -3 \\ 0 & 1 & -1 & 2 \end{array} \right]$$

The inverse matrix we want is now on the right side of the separation line in the augmented matrix.

Pre-multiply both sides of the equation by the inverse matrix:

$$\begin{bmatrix} 2 & -3 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} 2 & 3 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 2 & -3 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} 13 \\ 8 \end{bmatrix}$$

On the right side multiplying the matrix by its inverse gives the identity matrix. The left side works out to:  $\begin{bmatrix} 2 & -3 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} 13 \\ 8 \end{bmatrix} = \begin{bmatrix} 26 - 24 \\ -13 + 16 \end{bmatrix} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$ .

The net result is  $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$ , which says that

$$1x + 0y = 2$$

$$0x + 1y = 3$$

Inverse matrices are easy to obtain using a graphing calculator, however these calculators will also solve a system of equations for you directly if you just put it into an augmented matrix and ask for the reduced-row-echelon form.

When the determinant of a matrix is 0, it does not have an inverse. It is impossible to manipulate such a matrix to change it into an identity matrix. If a system has no unique solution it makes sense that the coefficient matrix should not have an inverse, otherwise we should be able to solve it by using that inverse. A system of linear equations does not have a unique solution if the determinant of its coefficient matrix is 0.

If the determinant is not zero, and the numbers on the left side of the equation are all zero, then the only solution is a zero value for all the variables.

As an example, let's look at

$$2x - 4y = 0$$

$$3x + 2y = 0$$

The coefficient matrix of this system, which we will call  $A$ , does not have a zero determinant. That means that there is an inverse. As a result, you can solve the system using  $A^{-1}$  as follows:

$$A^{-1}A \begin{bmatrix} x \\ y \end{bmatrix} = A^{-1} \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \text{ so } I \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \text{ or } \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

This says that  $1x + 0y = 0$

$$0x + 1y = 0$$

and you can see that  $x = 0, y = 0$  is the only solution for this system. Being able to find the determinant lets you figure this out.

As we saw previously, a system of equations can just be solved without using an inverse matrix. A more interesting application of inverse matrices is in encryption. The data you want to encrypt is placed into a matrix  $A$ , which is then pre-multiplied by some random matrix  $B$ . The result is a new matrix  $BA$ . The data are now hopelessly difficult to recover for someone who does not know what matrix was used for the multiplication. However, you provide the person you are sending the data to with the inverse of matrix  $B$ . Then they can pre-multiply the matrix they receive by  $B^{-1}$  and recover the original data. The calculation is  $B^{-1}BA = IA = A$ . A very simple example would be encrypting the word “base” by using  $a = 1, b = 2$  etc. We can place this into a matrix as  $\begin{bmatrix} 2 & 19 \\ 1 & 5 \end{bmatrix}$ . Pre-multiply that by the non-singular matrix  $\begin{bmatrix} 2 & 3 \\ 1 & 2 \end{bmatrix}$ , to get  $\begin{bmatrix} 2 & 3 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 2 & 19 \\ 1 & 5 \end{bmatrix}$ , which is  $\begin{bmatrix} 7 & 53 \\ 4 & 29 \end{bmatrix}$ . This is the encrypted matrix that you would send. Now pre-multiply the encrypted data by the inverse of  $\begin{bmatrix} 2 & 3 \\ 1 & 2 \end{bmatrix}$ , which we saw earlier is  $\begin{bmatrix} 2 & -3 \\ -1 & 2 \end{bmatrix}$ :

$$\begin{bmatrix} 2 & -3 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} 7 & 53 \\ 4 & 29 \end{bmatrix} = \begin{bmatrix} 14 - 12 & 106 - 87 \\ -7 + 8 & -53 + 58 \end{bmatrix} = \begin{bmatrix} 2 & 19 \\ 1 & 5 \end{bmatrix}. \text{ The word “base” has been recovered.}$$

(See <http://aix1.uottawa.ca/~jkhoury/cryptography.htm>).

## Solving Systems with an Infinite Number of Solutions

Let’s go back to the system we considered in the last section:

$$2x + 3y = 10$$

$$4x + 6y = 20$$

Since the determinant of the associated matrix is 0, there is no unique solution. If we multiply the first row by 2 and subtract it from the second row, we get

$$2x + 3y = 10$$

$$0x + 0y = 0$$

If the square coefficient matrix of a system can be reduced to one that has a row of zeros, its determinant must be 0 since the system can't have a unique solution. There are infinitely many solutions, but "Infinitely many solutions" does not seem like a good answer to the matrix people. After all, you can't just put in any solution here, such as  $x = 2$  and  $y = 10$ . There may be infinitely many solutions but they are still specific ones. Therefore you may be asked to specify what kind of solutions you are talking about. Pick an answer for  $x$ , preferably the standard one which is  $\alpha$ , and express the other variables in terms of  $\alpha$  if possible. If that is not possible, use  $\beta$  for the next variable that you cannot express in terms of  $\alpha$ . Then use  $\gamma$  if necessary, and so on. In this particular case we should select  $\alpha$  for  $x$ , and then solve for  $y$ :  $2\alpha + 3y = 10$ .  $y = (10 - 2\alpha)/3$  so  $y = -2/3 \alpha + 10/3$ . We can write our answer as a set of points  $(\alpha, -2/3\alpha + 10/3)$ , or as a vector

$$\begin{bmatrix} \alpha \\ -2/3 \alpha + 10/3 \end{bmatrix}.$$

The previous system of linear equations can be written as a matrix times a vector that equals another vector, like this:

$$\begin{bmatrix} 2 & 3 \\ 4 & 6 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 10 \\ 20 \end{bmatrix}$$

The unknown vector can be expressed as  $\begin{bmatrix} x \\ y \end{bmatrix}$  or for larger systems,  $\begin{bmatrix} x \\ y \\ z \end{bmatrix}$ . Vectors however are not

limited to three dimensions. We can multiply an appropriately sized matrix by an unknown

vector  $\mathbf{x}$ , expressed like this:  $\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ \dots \end{bmatrix}$ . Consider the following example:

$$-x_1 + x_2 - x_3 + 3x_4 = 0$$

$$3x_1 + x_2 - x_3 - x_4 = 0$$

$$2x_1 - x_2 - 2x_3 - x_4 = 0$$

This is an underdetermined system, and the associated matrix is not square (it would be square if

we added a row of zeros, and in that case the determinant would be 0). We already know there is no unique solution. This system can be manipulated until it simplifies to:

$$x_1 - x_4 = 0$$

$$x_2 + x_4 = 0$$

$$x_3 - x_4 = 0$$

To write the solution we set  $x_1 = \alpha$ . This leads us to the conclusion that  $x_4 = \alpha$ ,  $x_2 = -\alpha$ , and  $x_3 = \alpha$ . We can express this solution as a vector:

$$\begin{bmatrix} \alpha \\ -\alpha \\ \alpha \\ \alpha \end{bmatrix}.$$

## Statistics

### The Normal Curve and Z-Scores

The standard normal curve has a mean of zero, a standard deviation of 1, and the area underneath it is equal to 1. The function that draws the standard normal curve is considered a probability density function, because the area under the curve represents the probability that the random variable  $x$  will be greater than or less than a particular value, or in between two stated values. The equation that produces the graph of the standard normal curve is:

$$y = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}, \text{ where } x \text{ is a normal random variable.}$$

The reason that I am providing you with this rather complex equation is that you should probably see for yourself what secret this curve is hiding. Just use a graphing program to draw it, and you will find that it actually is much shorter and wider than its usual pictures suggest. It would make sense that it is not very tall, because the area under the curve must be equal to 1.

# Trigonometry

## Summary

$$\tan x = \frac{\sin x}{\cos x} \quad \cotan x = \frac{\cos x}{\sin x} \quad \sec x = \frac{1}{\cos x} \quad \csc x = \frac{1}{\sin x}$$

Important Identities:  $\sin^2 x + \cos^2 x = 1$        $\tan^2 x + 1 = \sec^2 x$        $1 + \cot^2 x = \csc^2 x$

### Even-Odd Identities

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

### Range of Inverse Trigonometric Functions

Arcsin	$-\pi/2$ to $\pi/2$ radians ( $-90^\circ$ to $90^\circ$ )
Arccosec	$-\pi/2$ to $\pi/2$ radians
Arccos	$0$ to $\pi$ radians ( $0^\circ$ to $180^\circ$ )
Arcsec	$0$ to $\pi$ radians
Arctan	$-\pi/2$ to $\pi/2$ radians, not including the endpoints ( $-90^\circ$ to $90^\circ$ )
Arccotan	$-\pi/2$ to $\pi/2$ radians, not including $0$ , <b>OR</b> $0$ to $\pi$ radians, excluding the endpoints.

### Cofunction Formulas

$$\cos x = \sin(90^\circ - x) \quad \sin x = \cos(90^\circ - x)$$

$$\cot x = \tan(90^\circ - x) \quad \tan x = \cot(90^\circ - x)$$

$$\csc x = \sec(90^\circ - x) \quad \sec x = \csc(90^\circ - x)$$

Law of cosines:  $c^2 = a^2 + b^2 - 2ab \cos C$

Law of sines:  $\frac{\sin A}{a} = \frac{\sin B}{b} = \frac{\sin C}{c}$

$$\sin(\alpha + \beta) = \sin \alpha \cos \beta + \cos \alpha \sin \beta \quad \sin(2x) = 2 \sin x \cos x$$

$$\sin(\alpha - \beta) = \sin \alpha \cos \beta - \cos \alpha \sin \beta$$

$$\cos(\alpha + \beta) = \cos \alpha \cos \beta - \sin \alpha \sin \beta$$

$$\cos(\alpha - \beta) = \cos \alpha \cos \beta + \sin \alpha \sin \beta$$

$$\begin{aligned} \cos(2x) &= \cos^2 x - \sin^2 x & \text{or} & \quad \cos(2x) = \cos^2 x - (1 - \cos^2 x) \\ &= 1 - \sin^2 x - \sin^2 x & & \quad = 2 \cos^2 x - 1 \\ &= 1 - 2\sin^2 x \end{aligned}$$

$$\sin^2 x = \frac{1 - \cos(2x)}{2} \quad \sin\left(\frac{x}{2}\right) = \pm \sqrt{\frac{1 - \cos x}{2}} \quad \sin^2 x = 1 - \cos^2 x = (1 + \cos x)(1 - \cos x)$$

$$\cos^2 x = \frac{1 + \cos(2x)}{2} \quad \cos\left(\frac{x}{2}\right) = \pm \sqrt{\frac{1 + \cos x}{2}}$$

$$\tan(x + y) = \frac{\tan x + \tan y}{1 - \tan x \tan y}$$

$$\tan\left(\frac{x}{2}\right) = \frac{\sin x}{1 + \cos x} \quad \text{or} \quad \tan\left(\frac{x}{2}\right) = \frac{1 - \cos x}{\sin x}$$

**Heron's Formula** for the area of a triangle:  $A = \sqrt{s(s-a)(s-b)(s-c)}$ , where  $s$  is the semiperimeter of the triangle.  $s = \frac{a+b+c}{2}$

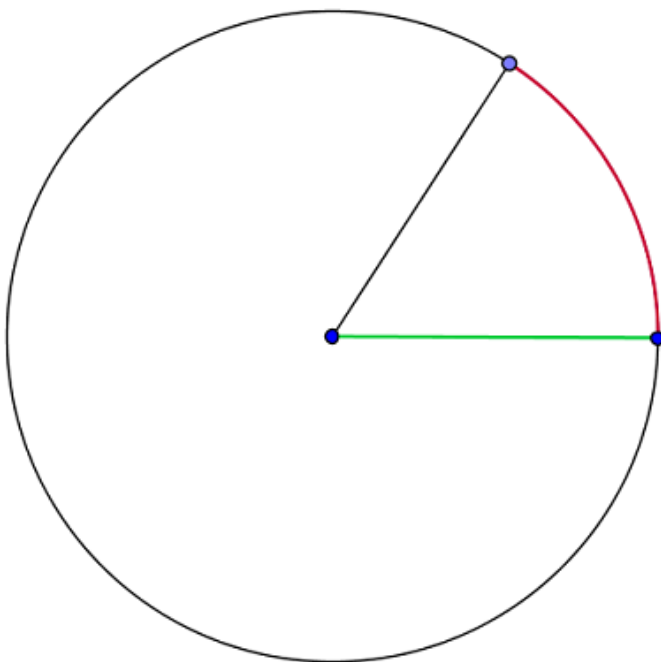
To model a change with a sine (or cosine) function, it is helpful to change the period to 1 first by multiplying  $x$  by  $2\pi$ . Then lengthen the period by dividing by what you need, like 365 days.

(Additional trigonometry resources: [www.themathpage.com](http://www.themathpage.com), and [www.intmath.com](http://www.intmath.com) )

If you play around with circles long enough you will eventually discover the idea of angles. First there is a whole circle. Then there is a half-circle, a quarter-circle, and so on. Angles are a way to divide up a circle into measured parts. Dividing a circle into 360 degrees was a really good idea. The number 360 is naturally very “divisible”, giving nice whole number angle measurements for  $1/3$  of a circle,  $1/4$ ,  $1/5$ ,  $1/6$ ,  $1/8$ ,  $1/9$ ,  $1/10$ ,  $1/12$  and more. This is a nice simple way to understand angles, and this is how students first learn about them.

The other way to measure out parts of a circle is to relate them to the radius of the circle. This reflects our deeper understanding of the relationship between the radius of a circle, its circumference, and angles. **To really understand this yourself, you need to actually work with a**

**circle.** Get some paper, a compass, and some thin string. If you do not have a compass, trace a circular object like a plate. The diameter of your circle is the largest distance across the circle. Use a ruler to find that, and then mark the center of the circle in the middle of the diameter. Once you have drawn a large circle, use a string to measure out the radius. Mark your string and lay it around the circumference of the circle. This is an angle of one radian.

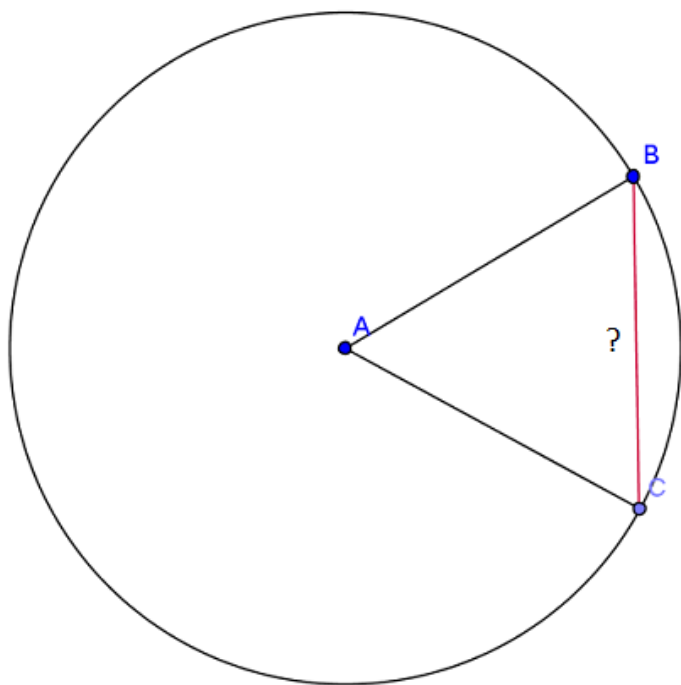


Here one radius is shown in green, and the string along the circumference is colored in red.

The red string is the same length as the radius. Because you know that the circumference of a circle is  $2\pi r$ , you can imagine that we can put  $2\pi$  copies of the red string around the edge of the circle. That is not a whole number. The circumference of the circle is 6 times the radius and a bit more. 6.283185..... radii fit around the circle, so there are that many radians in a circle. We usually just say that there are  $2\pi$  radians, and  $2\pi$  radians are equivalent to 360 degrees. That of course means that 180 degrees =  $\pi$  radians. Convert like this: 1 ~~radian~~ times  $\frac{180 \text{ degrees}}{\pi \text{ radians}}$  is about 57.3 degrees. Look at the picture of a radian again: one radian is a little smaller than a 60 degree angle.

In trigonometry you can work with either radians or degrees. **Just make sure your calculator is set to the right mode!!** Always check this if you get an unexpected answer from your calculator. The TI 84 has a MODE button; just press it and adjust between degrees and radians.

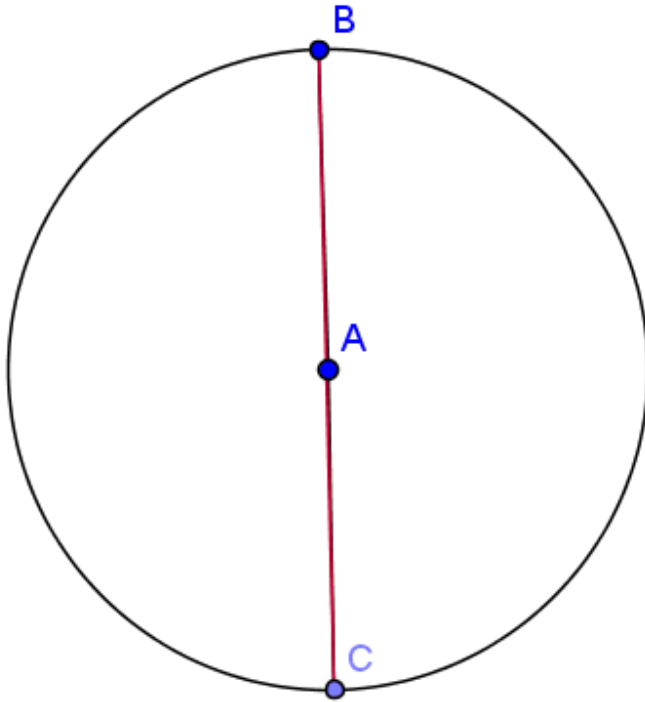
After people had played around with angles for a long time, they started to wonder about chords. These are lines that you draw from one point on the circle to another. Here is a chord (marked in red) associated with a specific angle:



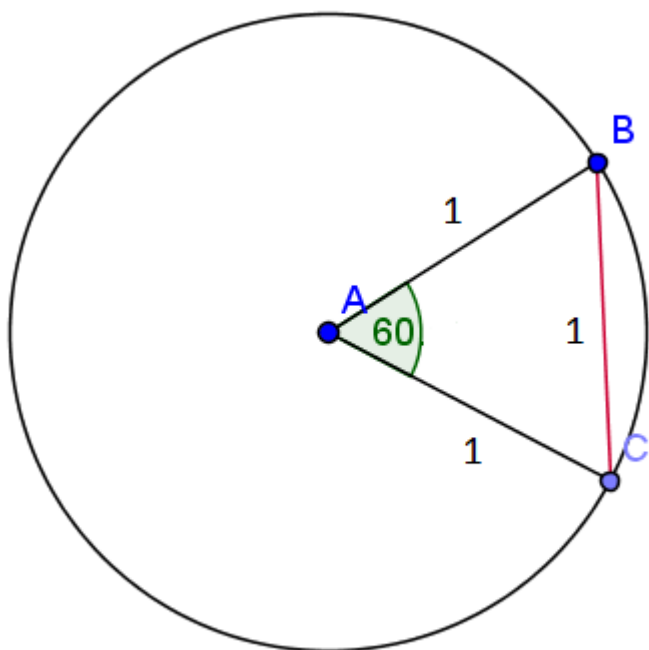
How long is the chord? Well, that depends on the angle, but of course also on the size of the circle. So, if you want to explore the relationship between the angle and the chord you should pick a fixed size for your circle. A convenient size would be to give the circle a radius of 1. “One what?” you might ask. Fortunately that doesn’t really matter. It could be 1 inch, 1 centimeter, or even 1 yard. We just say that the circle has a radius of 1 unit. Once we have decided on the size, we can look at different angles and see how long their chords are. The first one is obvious: an angle of 0 degrees has a chord of length 0.

The next one is rather simple too: an angle of 180 degrees has an associated chord of 2 times the radius, or 2 units in our case, as shown below.

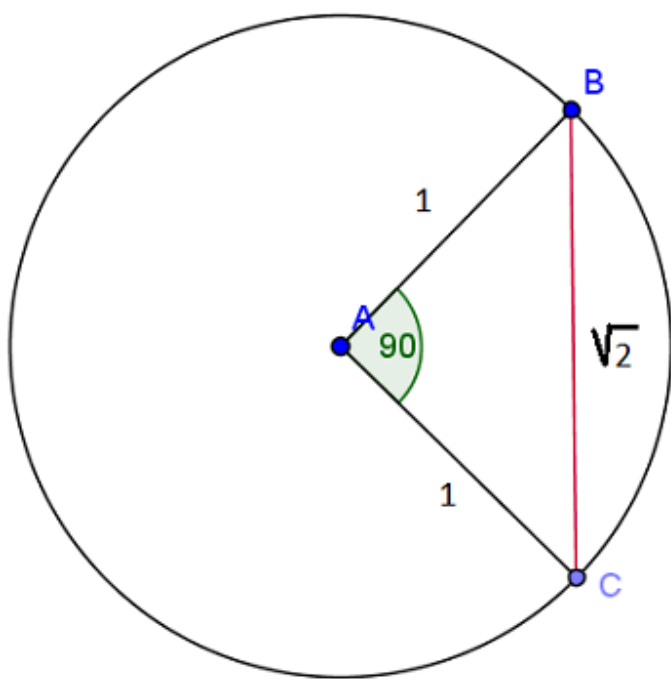




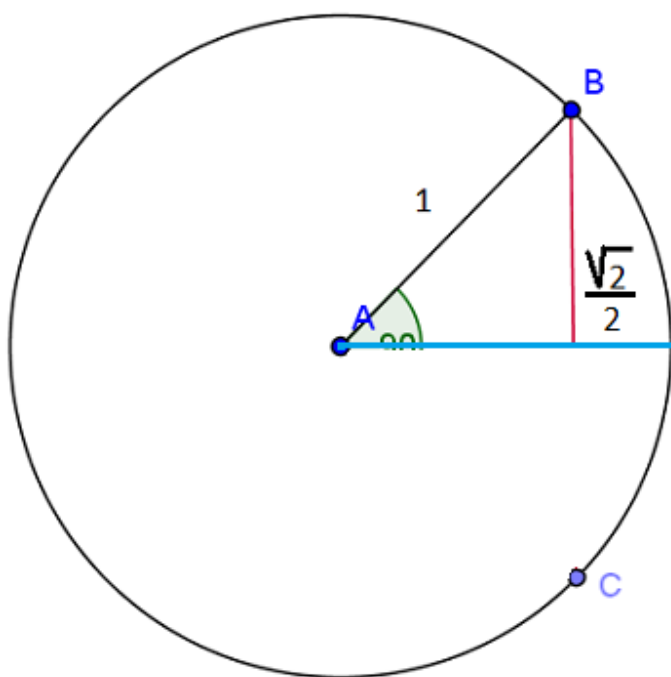
The length of the chord associated with an angle of 60 degrees is not too difficult to find if you remember your geometry. Look at the picture below to see that the angle and the chord form a triangle. The triangle must be isosceles since two of the sides are radii of the same circle. That means that the angles that are not labeled in the picture must both be equal. Because we already have a 60 degree angle, these angles have to both be 60 degrees so that the angles of the triangle will add up to 180 degrees. If all the angles are equal the triangle must be equilateral (all sides are equal). The length of the chord is 1 unit:



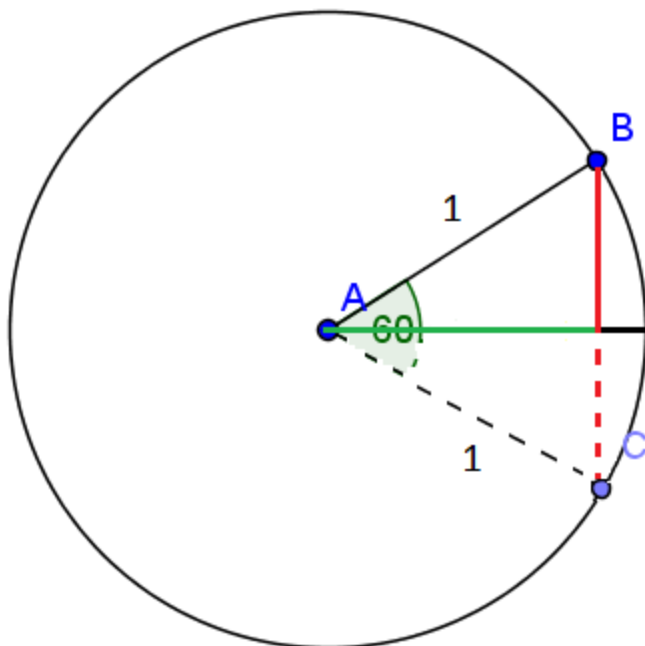
Using the Pythagorean Theorem, you can quickly find the chord length for a 90 degree angle. Since the chord is the hypotenuse for the 90 degree triangle, we find  $c$  by using  $a^2 + b^2 = c^2$  or  $1^2 + 1^2 = c^2$ :



Eventually people make things more efficient and more standardized. Notice that I drew all of these angles facing to the right, because that eventually became the standard. Also, instead of working with chords mathematicians started working with half-chords that eventually came to be called sines (apparently due to a mistake in translating the Arabic word for “half-chord”). Correspondingly, mathematicians also considered only half the angle. This picture shows the half-chord, or the sine, of 45 degrees ( $\pi/4$  radians):



Looking at things this way allows us to introduce a second measurement; the cosine:



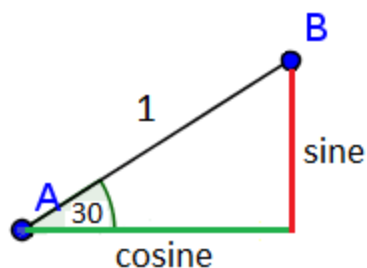
Here the solid red line shows the sine of 30 degrees, and the green line shows the cosine. The red line and the green line intersect at a right angle. We can still calculate the distances from the original chord picture using the dashed lines. The solid red line has a length of  $\frac{1}{2}$ . Using the Pythagorean Theorem, we see that the  $\text{sine}^2 + \text{cosine}^2 = 1$ , or  $(\frac{1}{2})^2 + \text{cosine}^2 = 1$ . This means that the cosine of a 30 degree angle is the square root of  $\frac{3}{4}$ , or  $\frac{\sqrt{3}}{2}$ .

Because the radius of the unit circle is 1, the relationship between the sine and the cosine is always “ $\text{sine}^2 + \text{cosine}^2 = 1$ ”. The sine and cosine depend on an angle, so we should write this important *trigonometric identity* as:

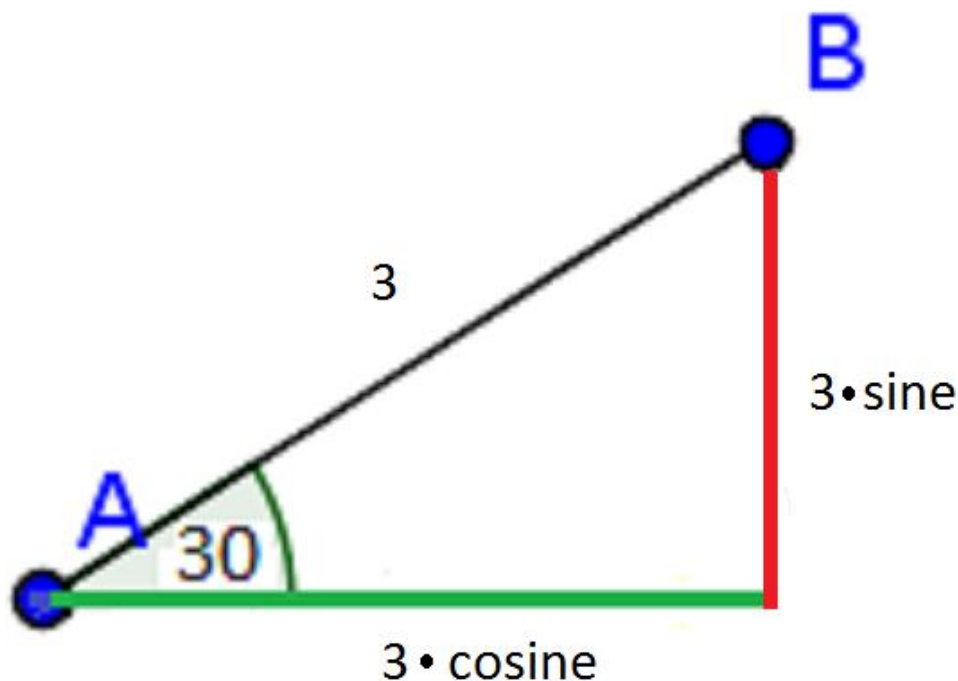
$$\text{sine}^2x + \text{cosine}^2x = 1$$

This says that  $(\text{the sine of angle } x)^2 + (\text{the cosine of angle } x)^2 = 1$ . If you know the sine of any angle you can calculate the cosine, and vice versa. The convention is to write sine as  $\sin$ , and  $\sin^2x$  rather than  $(\text{sine } x)^2$ , simply because it is easier.

It didn’t take long for someone to take the triangle made up of the sine, the cosine and the radius, and lift it right out of the unit circle:



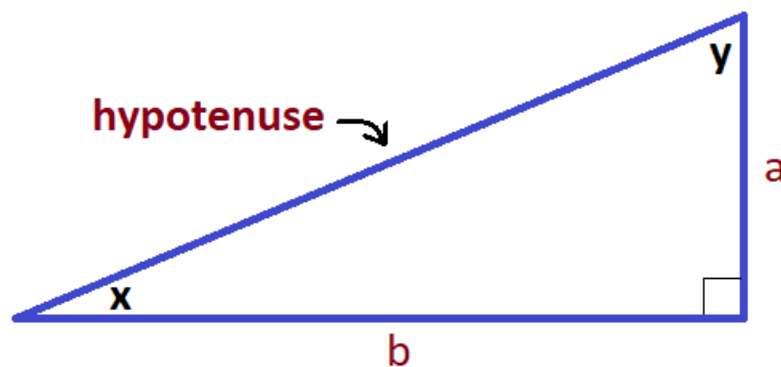
Notice that the angle at the bottom right corner of this triangle is a 90 degree angle, making this a right triangle. Knowing the sine and cosine of various angles is useful for finding the length of the sides of a *right triangle* (sometimes students will mistakenly apply the same techniques to other types of triangles, with predictably poor results). Of course not all right triangles have a hypotenuse of 1, so we may have to enlarge our picture a little:



Now we have a triangle with a hypotenuse of 3. The sides have been correspondingly enlarged to be 3 times the sine of 30 degrees, and 3 times the cosine of 30 degrees. This leads us to the idea that in any right triangle, the sine and cosine can be found by dividing the sides by the length of the hypotenuse. The sine of 30 degrees is the side opposite the angle divided by the

hypotenuse, and the cosine is the side adjacent to the angle divided by the hypotenuse. Tables listing the value of the sine and cosine for different angles were created a very long time ago, and today we can get these values from a calculator. (Windows: All Programs -> Accessories-> Calculator -> View -> Scientific.) Sine and cosine are usually abbreviated as sin and cos. We have already calculated the sine and cosine of 30 degrees, so for this triangle we can say that the red side has a length of  $3 \cdot \frac{1}{2} = 1.5$ , and the green side has a length of  $3 \cdot \frac{\sqrt{3}}{2}$ , which is about 2.6.

The sine of an angle is equal to the cosine of the complementary angle. In the image below,  $x$  and  $y$  are complementary angles ( $x + y = 90^\circ$ ):



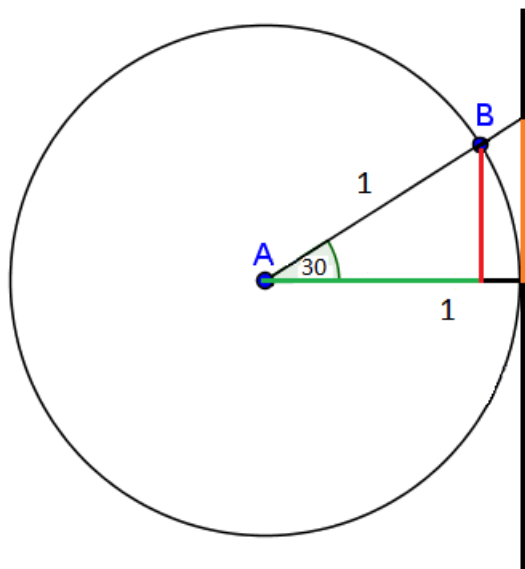
The sine of angle  $x$  is the opposite side,  $a$ , divided by the hypotenuse. The cosine of angle  $x$  is side  $b$  divided by the hypotenuse. That works the opposite way for angle  $y$ .

$$\sin x = \cos y \text{ and } \cos x = \sin y$$

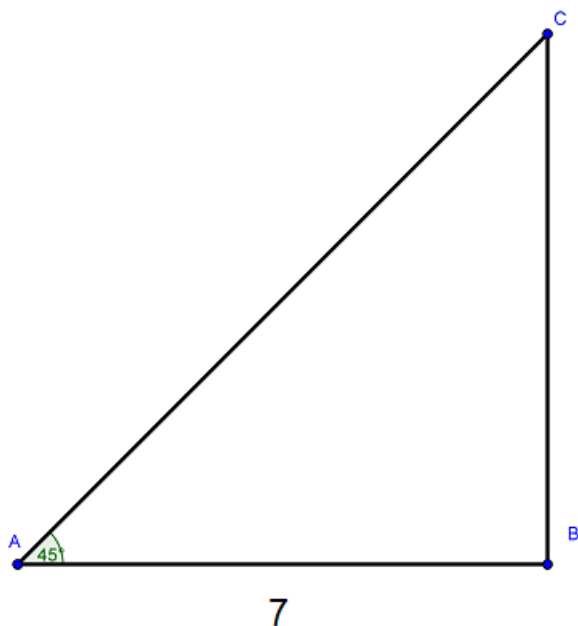
Because  $y = 90^\circ - x$  and  $x = 90^\circ - y$ , we can write

$$\sin x = \cos (90^\circ - x) \text{ and } \cos x = \sin (90^\circ - x)$$

Another measurement that came from the unit circle is the tangent. At first this was an actual tangent (a tangent is a line that touches a curve at exactly one point). In the picture below, a line has been drawn tangent to the unit circle. "The tangent" of 30 degrees is defined as the length of the orange line segment. That length is determined by extending the top line of the angle outside of the circle, to form a secant line (a secant is a line that cuts through a curve).



First consider the larger triangle with the orange line as one of its sides. The bottom side of the triangle is a radius of the circle, so its length is 1. We could say that the orange line divided by the bottom side is the tangent  $\div 1 =$  the tangent. Now consider the smaller triangle with the red side and the green side. Because its angles are the same as the angles of the larger triangle, we can say that the red side  $\div$  the green side = the tangent  $\div 1$ . That means that the sine  $\div$  the cosine = the tangent. If you are working with a right-angled triangle rather than in the unit circle, you can find the tangent by dividing the side opposite the angle by the side adjacent to the angle. This gives you the same ratio as sine/cosine. Because this is a ratio, the length of the hypotenuse doesn't matter at all. Your calculator will supply you with the value of the tangent (abbreviated as tan) of different angles, and you can use these values to figure out the lengths of the sides of various right triangles when some measurements are given.

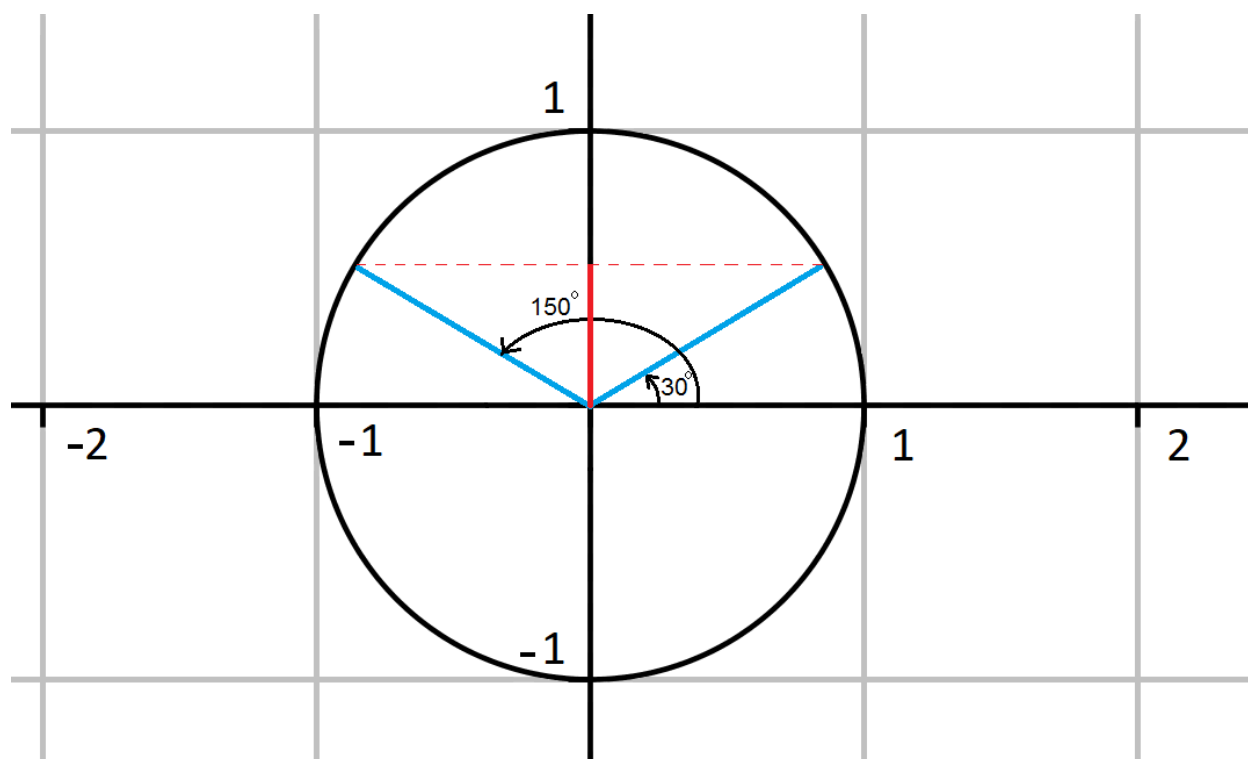


For this right triangle we can use the tangent to find the length of side BC, given that AB is 7 units, and the angle at A is 45 degrees. Don't rush to grab your calculator to find the tangent of 45 degrees. Instead, draw a 45 degree angle in a unit circle, and figure out the values of the sine and the cosine. Then you will know what the tangent is. After you find the length of side BC, you could use the Pythagorean Theorem to determine the length of AC, but you should also try using the sine or cosine. [Tangent of  $45^\circ = \frac{BC}{AB}$ , sine of  $45^\circ = \frac{BC}{AC}$ , and cosine of  $45^\circ = \frac{AB}{AC}$ ]

Right triangles are not always drawn in this same convenient position with the known angle in the bottom left corner. I used to turn the paper to adjust them, but I can't do that so easily on my computer. Here is where it might help you to use the mnemonic "SOHCAHTOA". You'll have to say that out loud a few times to remember it. It stands for Sine is Opposite ÷ Hypotenuse, Cosine is Adjacent ÷ Hypotenuse, Tangent is Opposite ÷ Adjacent. Opposite means the side opposite the angle. The adjacent side is the one touching the angle.

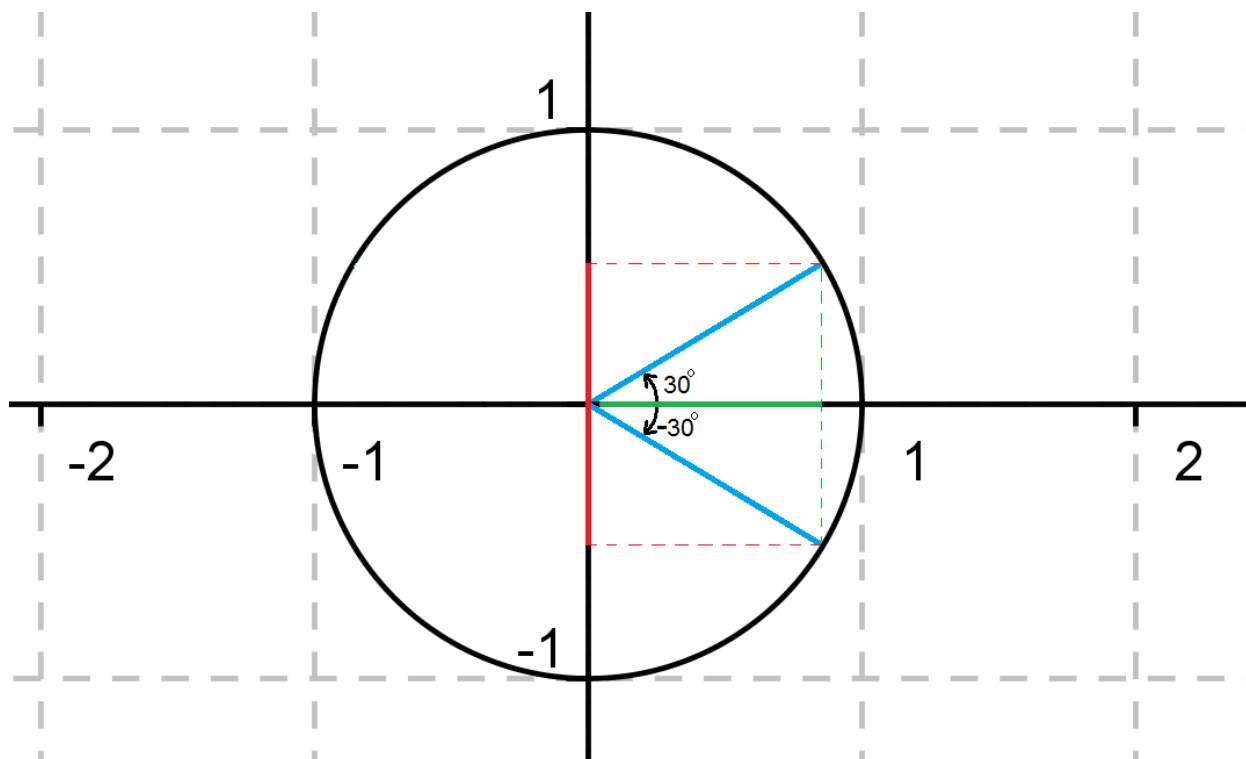


The next advance in trigonometry was the addition of coordinate axes to the unit circle. This allows us to consider the sine and cosine values of angles larger than 90 degrees. The sine is now defined as the distance along the y-axis, and the cosine is the distance along the x-axis. Because we are placing the center of the unit circle at the origin of the coordinate system, the values of the sine and cosine will vary from -1 to 1. This is an important fact to note. Sine and cosine values are never greater than 1 or smaller than -1. The tangent however can have any value from negative to positive infinity.



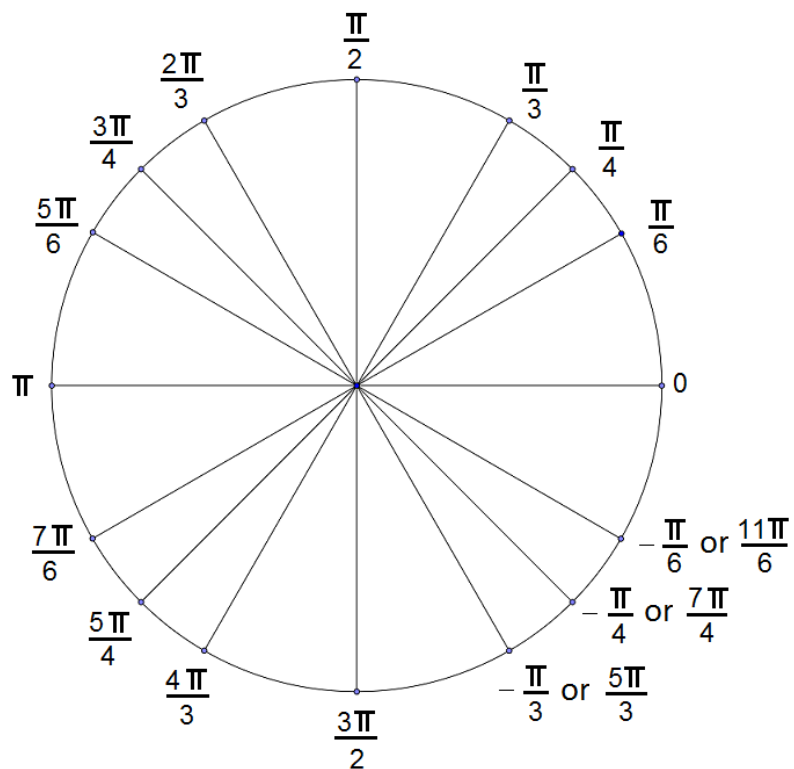
Here the sine values of a  $30^\circ$  angle and a  $150^\circ$  angle are measured along the y-axis, so you can see that the sine of both these angles is the same. Supplementary angles (two angles that add to 180 degrees) always have the same sine values. If the angle is larger than  $180^\circ$  the sine value becomes negative.

Negative angles are measured clockwise from the x-axis. Notice that the cosine of an angle is the same as that of the negative angle, while the sine of an angle is opposite to that of the negative angle:

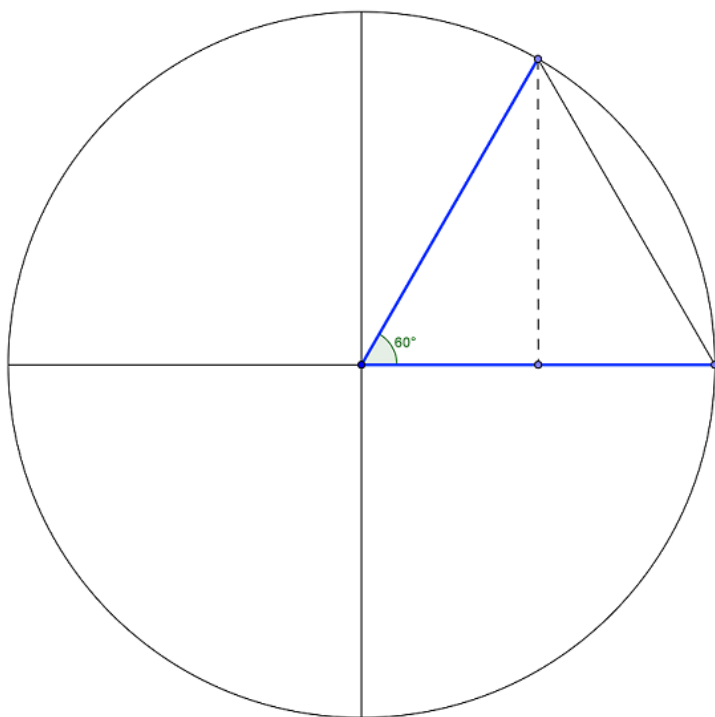


$$\cos x = \cos (-x) \quad \text{and} \quad \sin x = \sin (-x)$$

The next picture shows the important points on the unit circle, in radians. You should be able to reproduce this picture by yourself. The points in the 4<sup>th</sup> quadrant are often referred to as negative (clockwise) angle values to avoid more awkward fractions, but it is fine to use positive angle values instead. **For each of the indicated angle values you should be able to calculate the sine, cosine and tangent by hand.** Use a calculator to check if you are right (convert the radians to degrees, or set your calculator to radians). Note that as angle values get larger we just keep going around the circle. The point labeled 0 is the same as  $2\pi$ , and  $\pi/2$  is the same point as  $2.5\pi$ . There are an infinite number of ways to represent each point.



This illustration shows a sample calculation for the sine and cosine of  $\pi/3$  or  $60^\circ$ :



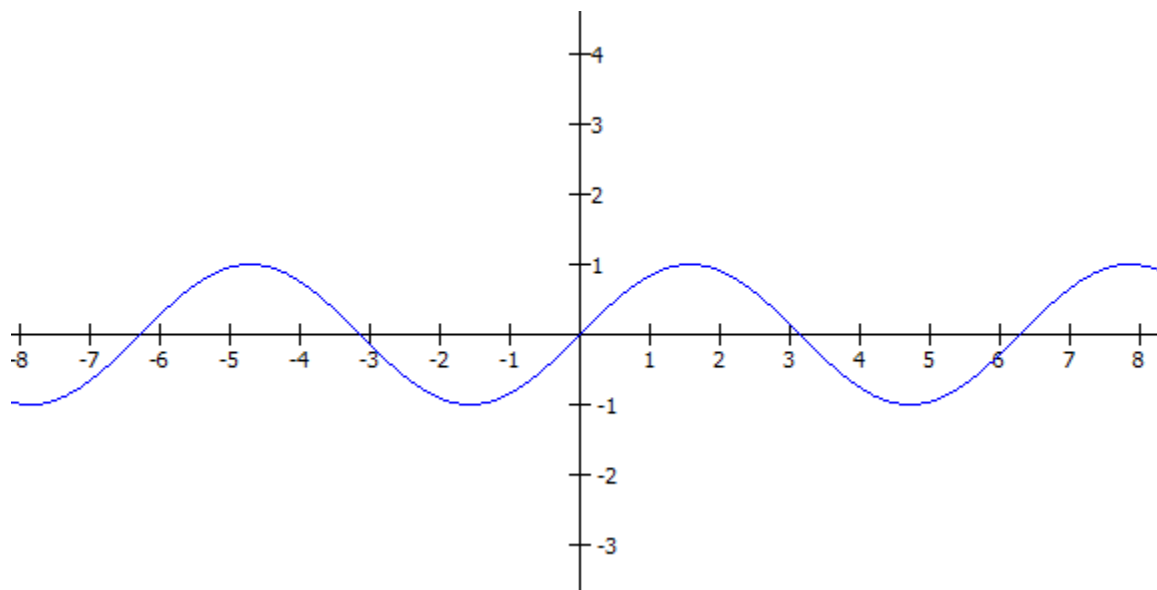
I can see that the two blue lines are both radii of the unit circle, so their length is 1. I have drawn an extra line to complete the isosceles triangle. The two remaining angles must be identical because they are the base angles of a sideways isosceles triangle. Since all the angles are 60 degrees all of the sides have the same length and the triangle is equilateral. The dashed line divides it into two identical parts. The height of the dashed line represents the sine of 60 degrees, and it intersects the x-axis at the value of the cosine, which is exactly  $1/2$ . Once we know the cosine value, we use the Pythagorean Theorem to find the height of the dashed line:

$$a^2 + b^2 = c^2, \text{ so } \left(\frac{1}{2}\right)^2 + b^2 = 1^2$$

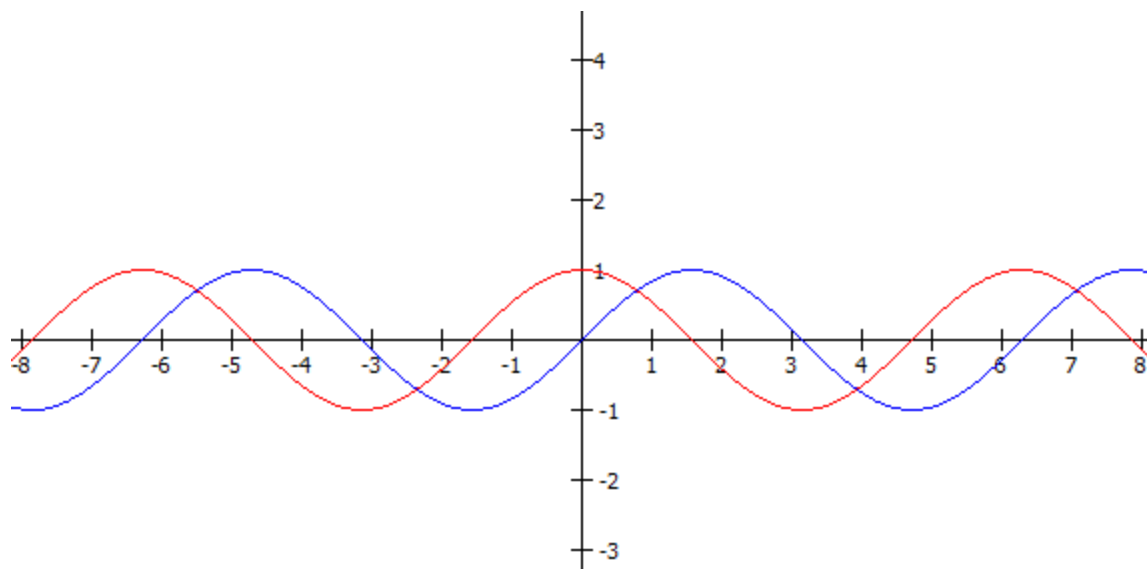
$$\frac{1}{4} + b^2 = 1, \text{ so } b^2 = \frac{3}{4} \text{ and } b = \sqrt{\frac{3}{4}} = \frac{\sqrt{3}}{\sqrt{4}} = \frac{\sqrt{3}}{2}$$

Once you have calculated all the important sine and cosine values you can create nice graphs. In fact, the unit circle itself is a graph of the cosine vs. the sine. At each point on the unit circle the x value is the cosine and the y value is the sine. The unit circle is rarely talked about as a graph because it is not a function (there are two different y values for almost every x value). Instead of graphing y against x, we can graph both the sine and the cosine against different angle values. This creates graphs that actually represent functions. Before you draw such graphs, consider that we created the unit circle setup by using radians. The values of the sine and cosine vary between 1 radius of the unit circle in the negative direction to 1 radius in the

positive direction. The angle units that go with these radius measures are radians. Using graph paper, graph the angle in radians against the value of the sine. **Do not use degrees as a scale, since it would not be meaningful to graph a sine value of 1 unit on the y-axis against a degree value of 90 units along the x-axis.** You should be able to create a fairly accurate copy of the graph below by yourself.



This graph shows that the sine value goes from 0 to 1 and back to 0 as the angle goes from 0 to  $\pi$  radians in the unit circle. As the angle continues to increase the sine value becomes negative, and then goes back to zero at an angle of  $2\pi$ . After that the same process repeats endlessly as we go around the unit circle again and again. Repeating functions like this are called *periodic* functions. The sine function has a period of  $2\pi$ . The negative values along the x-axis represent angles as we go around the unit circle in a clockwise direction rather than counterclockwise.



Here the red line shows the cosine function. You should also create your own copy of this on graph paper.

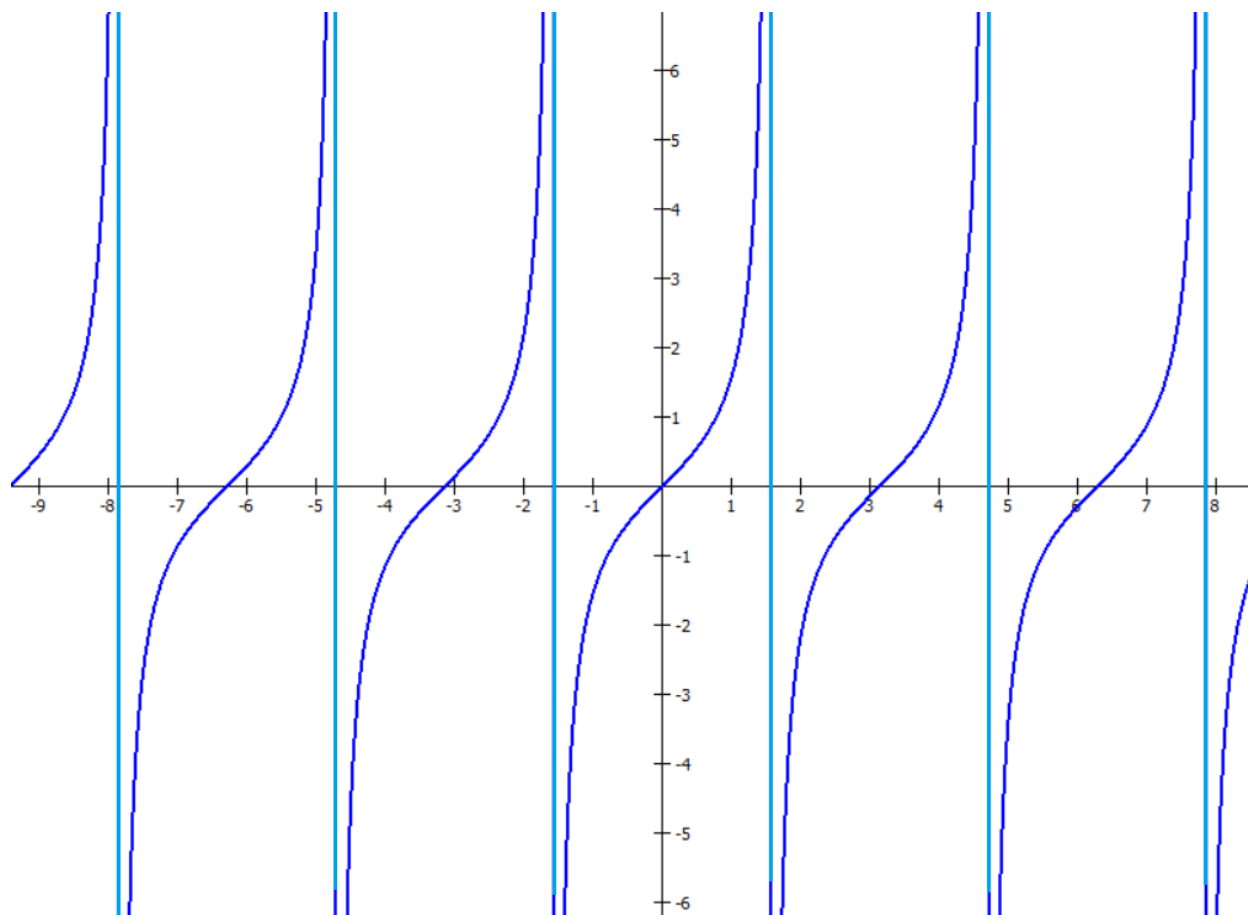
You can see that the sine and cosine functions just differ by a horizontal shift. Shift the sine function  $\frac{\pi}{2}$  units to the left to get the cosine:

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

Or shift the cosine function  $\frac{\pi}{2}$  units to the right to get the sine function:

$$\cos\left(x - \frac{\pi}{2}\right) = \sin x$$

The following interesting picture shows the tangent function. The vertical lines are vertical asymptotes, which occur at points where the cosine is 0. Since the tangent is the sine divided by the cosine, it cannot exist at these points, but it gets larger and larger in either a positive or negative direction as it approaches them. The tangent function repeats, with a period of  $\pi$ . Make sure you understand why the tangent function has the shape and the period that it does by calculating sample values of the tangent at various points.



## Inverse Trigonometric Functions

If we have an angle of  $\pi/6$  radians (30 degrees), we can figure out that the sine is  $1/2$ . The inverse sine function takes the value of the sine, and gives us the angle. The inverse sine function is called the arcsine, and it is often written, confusingly, as  $\sin^{-1}x$ . **This does not mean  $\frac{1}{\sin x}$ .**

If we know that the sine of an angle is 1, we might guess that the angle is  $\pi/2$  radians. However, it could also be  $\frac{5\pi}{2}$  radians (450°), or  $\frac{9\pi}{2}$  radians, or .... The possibilities are endless because the sine function repeats endlessly. An inverse sine function can only be created by taking a portion of the sine curve that is small enough that it does not return multiple angles for

a given value of  $\sin(x)$ . The part that is used for this purpose is the portion of the sine curve between  $-\pi/2$  and  $\pi/2$  radians ( $-90^\circ$  to  $90^\circ$ ). This may seem a bit hard to remember, but just think about what part of the unit circle you would choose if you were in charge. You have to allow for the full range of values of the sine, which is from  $-1$  to  $1$ . You could start that at  $\pi/2$  and go to  $\frac{3\pi}{2}$  radians, but chances are that you'd leave it right where it is now just to avoid all those awkward fractions in between. As a result the inverse sine function, known as the arcsine function, will always return angles between  $-\pi/2$  and  $\pi/2$  radians. Any sine value is a valid input for this function (recall that the sine is always between  $-1$  and  $1$ , so don't try to put in other numbers!).

In the same way mathematicians have defined an inverse cosine function, which returns angles between  $0$  and  $\pi$  radians. Again the input had to be between  $-1$  and  $1$ , so the most convenient interval to choose was  $0$  to  $\pi$  radians. This allows for the full range of values of the cosine, and returns one unique angle for each one.

If you look at the tangent function, you can see that over the interval  $-\pi/2$  to  $\pi/2$  radians the tangent goes from its minimum value ( $-\infty$ ) to its maximum value ( $+\infty$ ). Here again there are other intervals you could choose, but  $-\pi/2$  to  $\pi/2$  is the most convenient one. The arctangent function gives you angles between  $-\pi/2$  and  $\pi/2$  radians for any tangent value you enter.

### Range of Inverse Trigonometric Functions:

<b>Arcsin</b>	<b><math>-\pi/2</math> to <math>\pi/2</math> radians</b>	<b><math>(-90^\circ</math> to <math>90^\circ)</math></b>
<b>Arccos</b>	<b><math>0</math> to <math>\pi</math> radians</b>	<b><math>(0^\circ</math> to <math>180^\circ)</math></b>
<b>Arctan</b>	<b><math>-\pi/2</math> to <math>\pi/2</math> radians</b>	<b><math>(-90^\circ</math> to <math>90^\circ</math>, not including the endpoints)</b>

Whenever you use the inverse trigonometric functions and get a “wrong” angle, remember that the angle you want may not be in the range of the function. Look for another angle that has the same sine, cosine or tangent value!

If you take the output of a particular function and then apply the inverse of that function, you will get your original input back. For example, the inverse of  $y = 3x$  is  $y = \frac{x}{3}$ . If I take the number  $5$  and stick it into the function  $y = 3x$ , I will get  $15$ . If I then take that output and put it into the inverse function,  $y = \frac{x}{3}$ , I get my original number back. Things work the same for trigonometric functions and their inverses, within certain limits. The sine of  $\frac{\pi}{2}$  is  $1$ . When you plug  $1$  into the



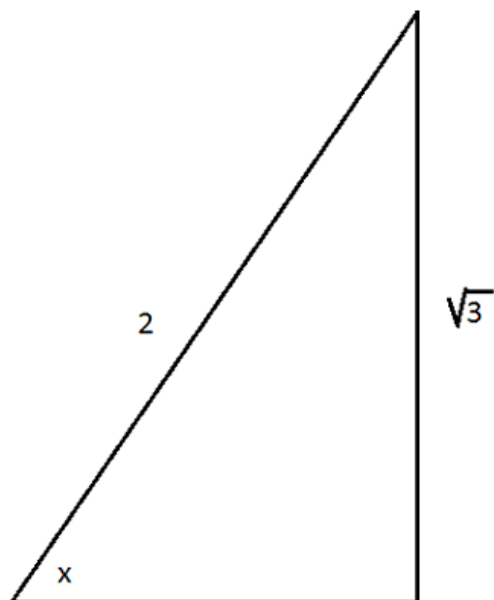
inverse sine function, you get  $\frac{\pi}{2}$  back. We can abbreviate that as

$\sin^{-1}(\sin(\frac{\pi}{2})) = \frac{\pi}{2}$ , or  $\arcsin(\sin(\frac{\pi}{2})) = \frac{\pi}{2}$ . This says that the inverse sine of the sine of  $\frac{\pi}{2}$  is  $\frac{\pi}{2}$ . In the same way,  $\sin(\sin^{-1}(1)) = 1$ . Just check it out on your calculator. Be careful not to enter a value that is outside the domain of an inverse function, or the result will be undefined. For example,  $\cos(\arcsin(2))$  is undefined because the cosine is always between -1 and 1. You cannot enter 2 as an input for the inverse cosine function. If you are outside the range of an inverse trigonometric function, you will get a different angle value than what you would expect. For example,  $\tan^{-1}(\tan(2\pi)) = 0$ , because  $2\pi$  is not within the range of the arctan function.

Sometimes you may be given a value of the arcsine, arccosine or arctangent and asked to calculate another trigonometric value for the corresponding angle. Such problems look like this:

$$\tan(\arcsin(\sqrt{3}/2)) = ?$$

To solve problems like this quickly, draw a right-angled triangle using the known values as the sides. We know that the sine can be found by dividing the side opposite to the angle by the length of the hypotenuse, so we choose  $\sqrt{3}$  for the opposite side and 2 for the hypotenuse. Now the arcsin function should return the angle  $x$  marked in the picture below:



We could just as well have set the length of the hypotenuse at 1 and the length of the opposite

side at  $\sqrt{3}/2$ , but doing things this way makes our calculations easier. We can quickly use the Pythagorean Theorem to find the length of the adjacent side, which is  $\sqrt{2^2 - \sqrt{3}^2} = \sqrt{4 - 3} = 1$ . The picture then shows us that the tangent of angle  $x$  is  $\frac{\sqrt{3}}{1}$ , since the tangent is the opposite side divided by the adjacent side.  $\tan(\arcsin(\sqrt{3}/2)) = \sqrt{3}$ .

For some practice, show that  $\sin(\arccos \frac{\sqrt{2}}{2}) = \frac{\sqrt{2}}{2}$ .

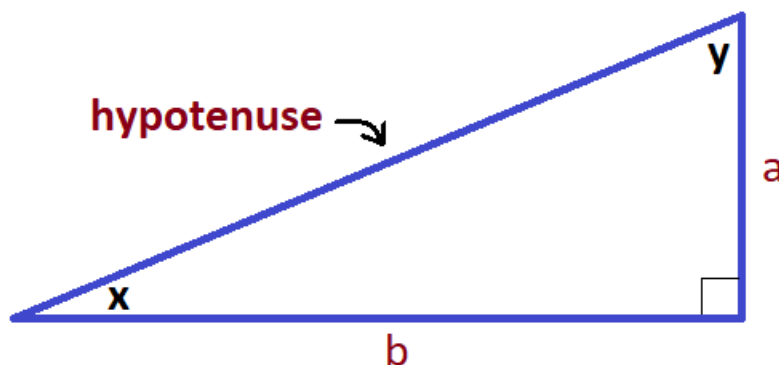
## Secant, Cosecant, and Cotangent

The inverse trigonometric functions are often written as  $\sin^{-1}$ ,  $\cos^{-1}$ , and  $\tan^{-1}$ . This does **not** mean  $1/\sin$ ,  $1/\cos$  and  $1/\tan$ , which refer to the cosecant, the secant, and the cotangent respectively. What I find the most confusing here is that  $\frac{1}{\sin x}$  is called the cosecant, while  $\frac{1}{\cos x}$  is the secant. The cotangent is easier to remember:  $\tan x = \frac{\sin x}{\cos x}$ , and  $\cot x = \frac{\cos x}{\sin x}$ . Apparently these inverse trigonometric ratios were once used to simplify calculations, but unfortunately they have failed to disappear with the advent of calculators and computers. You can do conversions using the cofunction formulas:

$$\begin{aligned} \cos x &= \sin(90^\circ - x) & \sin x &= \cos(90^\circ - x) \\ \cot x &= \tan(90^\circ - x) & \tan x &= \cot(90^\circ - x) \\ \csc x &= \sec(90^\circ - x) & \sec x &= \csc(90^\circ - x) \end{aligned}$$

Note that  $90^\circ - x$  in degrees is the same as  $\frac{\pi}{2} - x$  for radians.

These formulas work because the sine of an angle is equal to the cosine of the complementary angle. In the image below,  $x$  and  $y$  are complementary angles ( $x + y = 90^\circ$ ):



The sine of angle  $x$  is the opposite side,  $a$ , divided by the hypotenuse. The cosine of angle  $x$  is side  $b$  divided by the hypotenuse. That works the other way for angle  $y$ : The sine of angle  $y$  is side  $b$  divided by the hypotenuse, and the cosine of  $y$  is side  $a$  divided by the hypotenuse:

$$\sin x = \cos y, \text{ and } \cos x = \sin y$$

$x$  is the complementary angle of  $y$ , and since  $x + y = 90^\circ$ ,  $x = 90^\circ - y$ .

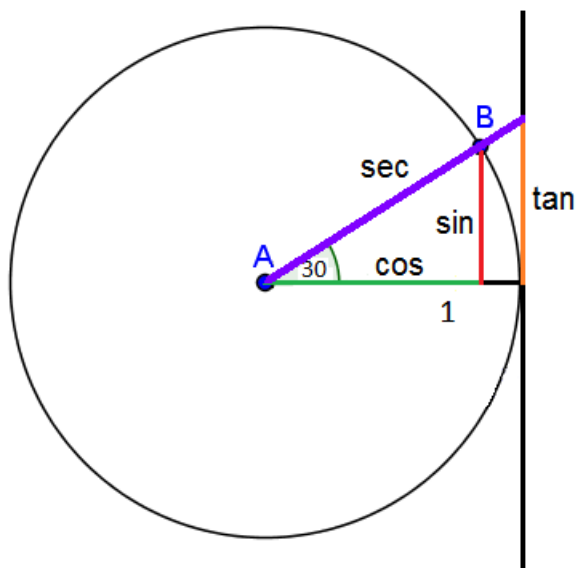
Earlier, we looked at the important trigonometric identity  $\sin^2 x + \cos^2 x = 1$ , which is just the result of applying the Pythagorean Theorem. We can create another important identity from this by dividing both sides by  $\cos^2 x$ :

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

$$\tan^2 x + 1 = \frac{1}{\cos^2 x}$$

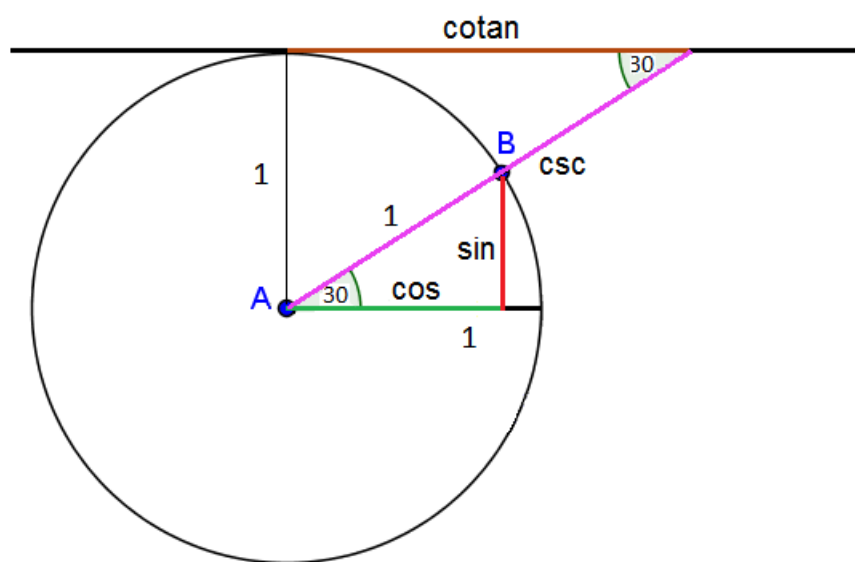
$$\tan^2 x + 1 = \sec^2 x$$

Interestingly, the secant has a specific representation in our picture that shows the sine, cosine and tangent:



The purple line is the secant. You can use similar triangles to show that  $\frac{1}{\cos}$  in the smaller triangle corresponds to  $\frac{\sec}{1}$  in the larger triangle. The hypotenuse of the smaller triangle is the radius of the unit circle, so its length is 1 unit:  $\frac{1}{\cos x} = \frac{\sec x}{1}$ . This is why the secant is defined as  $\frac{1}{\cos x}$ . If you apply the Pythagorean Theorem you can see that  $\tan^2 x + 1 = \sec^2 x$ .

Here is a picture showing the cotangent in orange and the cosecant as a pink line. The two 30 degree angles are alternate interior angles.



The cosine divided by the sine in the smaller triangle corresponds to the cotangent divided by 1 in the larger triangle. 1 divided by the sine in the small triangle corresponds to the cosecant divided by 1 in the large triangle, so the cosecant is  $\frac{1}{\cosine}$ .

A third trigonometric identity can be created by dividing both sides of  $\sin^2 x + \cos^2 x = 1$  by  $\sin^2 x$ :

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

$$1 + \cotan^2 x = \frac{1}{\sin^2 x}$$

$$1 + \cotan^2 x = \operatorname{cosec}^2 x$$

Again, this identity can be derived directly from the picture above by applying the Pythagorean Theorem.

There are of course also inverse functions for the secant, cosecant and cotangent (sigh). Luckily they have (mostly) the same range as their corresponding inverse functions:

Range of Inverse Trigonometric Functions:

Arcsine             $-\pi/2$  to  $\pi/2$  radians    ( $-90^\circ$  to  $90^\circ$ )

Arccosecant     $-\pi/2$  to  $\pi/2$  radians, not including 0 radians where the sine is zero.

Arccosine        0 to  $\pi$  radians    ( $0^\circ$  to  $180^\circ$ )

Arcsecant        0 to  $\pi$  radians, not including  $\pi/2$  where the cosine is zero.

Arctangent       $-\pi/2$  to  $\pi/2$  radians, not including the endpoints ( $-90^\circ$  to  $90^\circ$ )

Arccotangent    $-\pi/2$  to  $\pi/2$  radians, not including 0, **OR** 0 to  $\pi$  radians, excluding the endpoints.

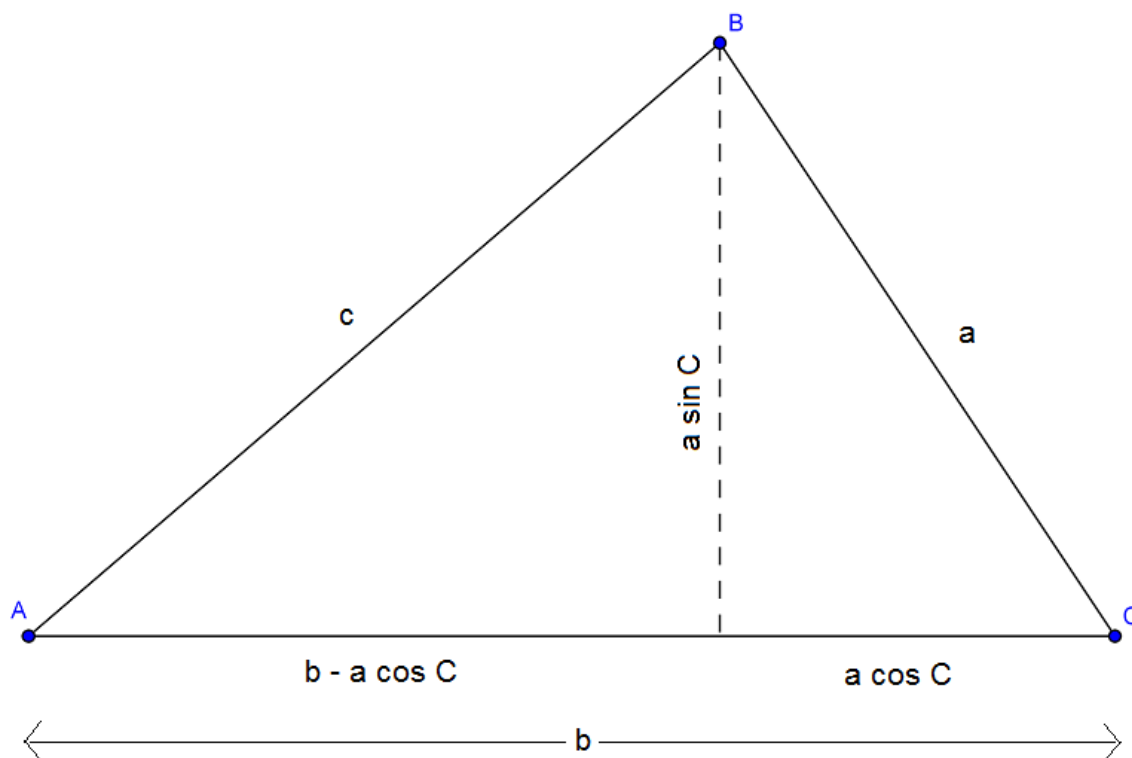
## The Law of Cosines

The Pythagorean Theorem and trigonometric ratios help you find missing measurements of right triangles. Unfortunately the Pythagorean Theorem only works for right triangles. If there

is no right angle in your triangle, you need to extend the theorem to deal with that situation. The extended Pythagorean Theorem is called the Law of Cosines.

Recall from geometry that once you draw two sides of a triangle the third side is already determined, since there is now only one way to complete it. This is expressed in the SAS Theorem: if two triangles have two sides that are the same length, and the angles between those sides are the same, the two triangles must be congruent. In the picture below, imagine that we know side  $a$ , side  $b$ , and the angle between them at point  $C$ . This means that side  $c$  has a specific fixed length, and we want to find it. We can use the Pythagorean Theorem and trigonometric ratios to do so.

You should note in the picture below that the convention is to label the sides so that they correspond to the angle opposite them. This makes sense, since the length of the side is dependent on the size of the opposite angle and vice versa.



Side  $c$  is the side of the triangle opposite point  $C$ . We are going to find an expression for the length of side  $c$ . The smart thing to do is to draw an altitude of the triangle (dotted line) so that  $c$  becomes the hypotenuse of a right triangle. Then we can just use the Pythagorean Theorem to find the length of  $c$ . All we need is the length of the other two sides, and we can use

trigonometric ratios to find those. Consider the angle at point C. The sine of angle C is the altitude divided by a, so the dotted line must have length  $a \sin C$ . By the same reasoning you can see that the segment labeled  $a \cos C$  divided by a actually represents the cosine of angle C. We can divide side b into two segments:  $a \cos C$ , and  $b$  minus  $a \cos C$ . Applying the Pythagorean Theorem we get:

$$\begin{aligned} c^2 &= (b - a \cos C)^2 + (a \sin C)^2 \\ &= b^2 - 2ab \cos C + (a \cos C)^2 + (a \sin C)^2 \\ &= b^2 - 2ab \cos C + a^2(\cos C)^2 + a^2(\sin C)^2 \end{aligned}$$

$(\cos C)^2$  is normally written as  $\cos^2 C$  and  $(\sin C)^2$  is written as  $\sin^2 C$ :

$$c^2 = b^2 - 2ab \cos C + a^2(\cos^2 C + \sin^2 C)$$

Since  $\cos^2 C + \sin^2 C = 1$ , we can simplify this to

$$c^2 = a^2 + b^2 - 2ab \cos C$$

This is the Law of Cosines. The unknown side is c, and the rest of the formula refers to the other two sides and the angle between them. When angle C is  $90^\circ$ ,  $\cos C$  is 0 and the Law of Cosines becomes the Pythagorean Theorem.

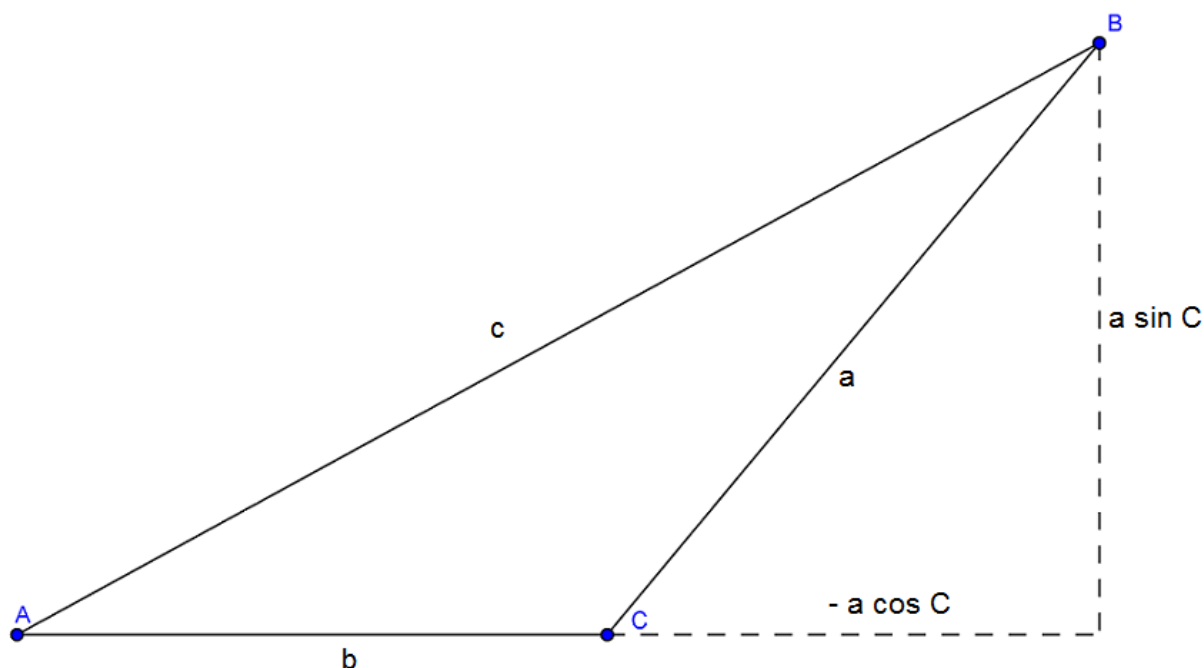
If you really want to understand the Law of Cosines, you should draw your own copy of the triangle and find the formula for determining the length of side a or b when the other two sides are known. The answers you should get are:

$$a^2 = b^2 + c^2 - 2bc \cos A$$

$$b^2 = a^2 + c^2 - 2ac \cos B$$

Notice that in each case it is the angle between the two known sides that appears in the formula (SAS congruence). When all of the sides of a triangle are known, the triangle is also completely determined (SSS congruence). Using the Law of Cosines, we can determine the angles of a triangle if we know the lengths of all three sides.

Here is a different view of the Law of Cosines, showing that it works just as well for an obtuse triangle:



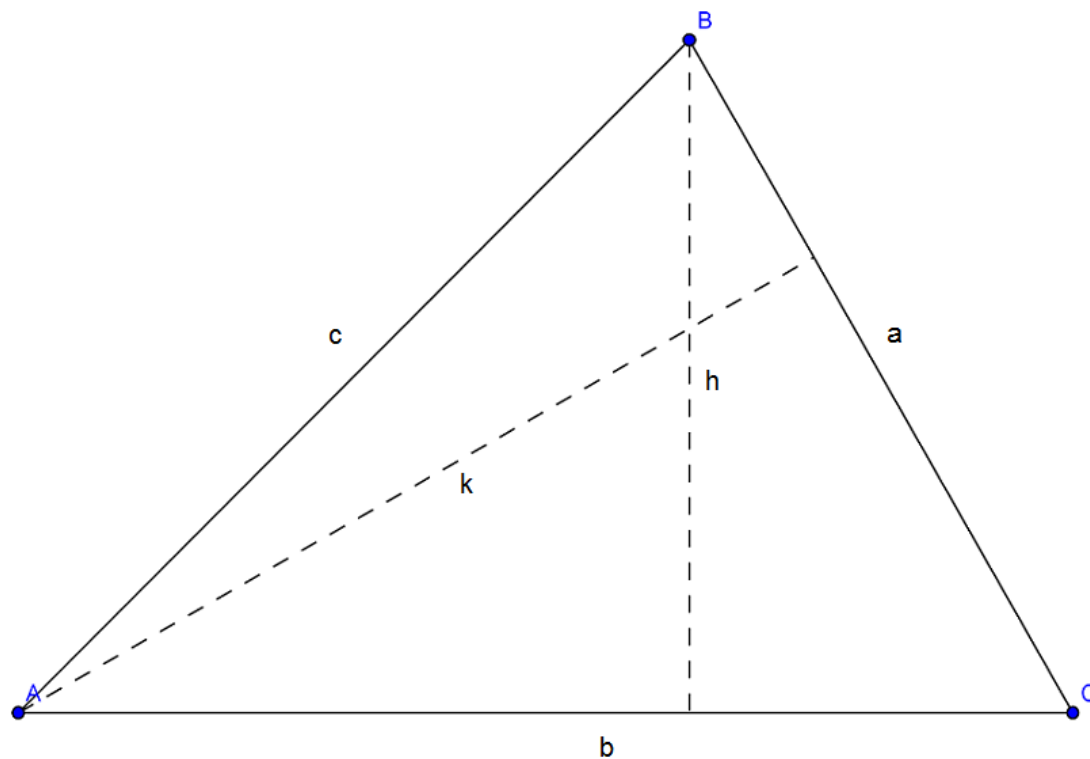
For the angle supplementary to angle C, the sine is equal to  $\sin C$ , and the cosine is equal to  $-\cos C$ .

$$\begin{aligned}
 c^2 &= (a \sin C)^2 + (b + -a \cos C)^2 \\
 &= a^2 \sin^2 C + b^2 - 2ab \cos C + a^2 \cos^2 C \\
 &= a^2 (\sin^2 C + \cos^2 C) + b^2 - 2ab \cos C \\
 &= a^2 + b^2 - 2ab \cos C
 \end{aligned}$$

## The Law of Sines

The Law of Sines provides another convenient way to deal with triangles that don't contain a right angle.





$$\sin A = \frac{h}{c}, \text{ so } h = c \sin A$$

$$\sin B = \frac{k}{c}, \text{ so } k = c \sin B$$

$$\sin C = \frac{h}{a}, \text{ so } h = a \sin C$$

$$\sin C = \frac{k}{b}, \text{ so } k = b \sin C$$

$$c \sin A = h = a \sin C$$

$$c \sin B = k = b \sin C$$

$$c \sin A = a \sin C$$

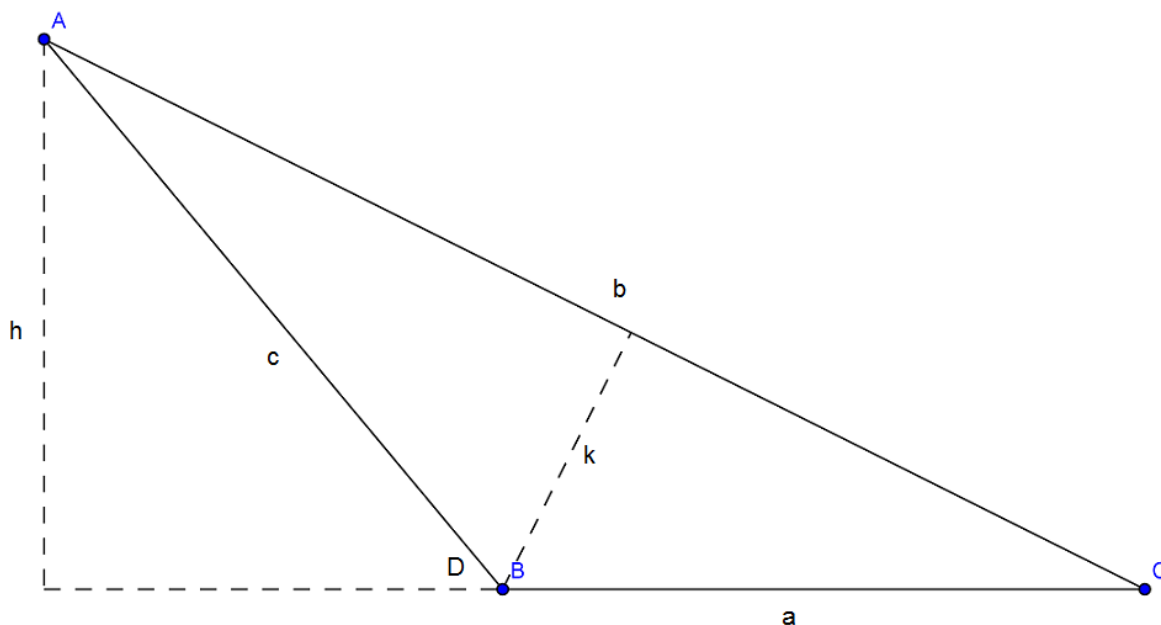
$$c \sin B = b \sin C$$

$$\frac{\sin A}{a} = \frac{\sin C}{c}$$

$$\frac{\sin B}{b} = \frac{\sin C}{c}$$

$$\frac{\sin A}{a} = \frac{\sin B}{b} = \frac{\sin C}{c}$$

The Law of Sines also works for obtuse triangles. Here the angle labeled D is supplementary to angle B, so the sine value of angle D is the same as that of angle B:



$$\sin A = \frac{k}{c}, \text{ so } k = c \sin A$$

$$\sin B = \frac{h}{c} \text{ since } \sin B = \sin D, \text{ so } h = c \sin B$$

$$\sin C = \frac{k}{a}, \text{ so } k = a \sin C$$

$$\sin C = \frac{h}{b}, \text{ so } h = b \sin C$$

$$c \sin A = a \sin C$$

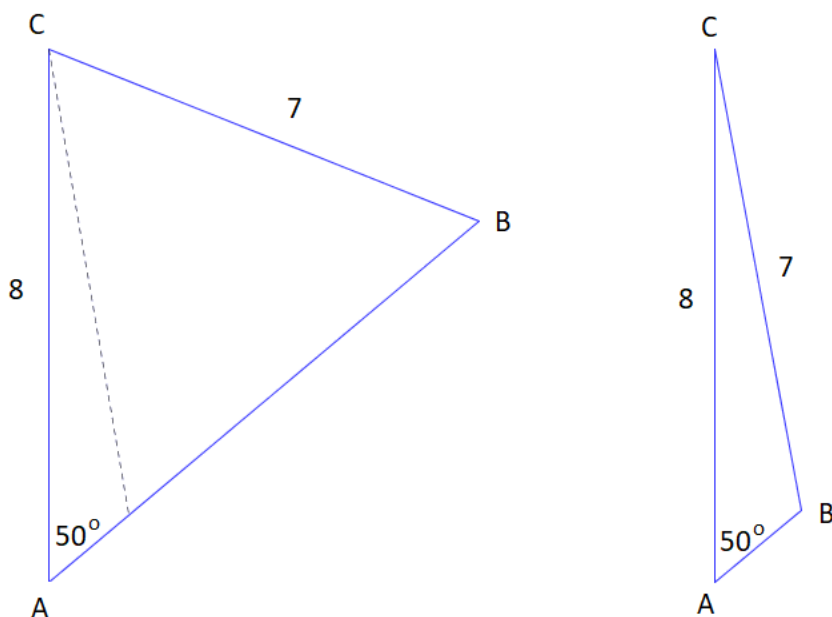
$$c \sin B = b \sin C$$

$$\frac{\sin A}{a} = \frac{\sin C}{c}$$

$$\frac{\sin B}{b} = \frac{\sin C}{c}$$

Although the Law of Sines seems straightforward, you may have some difficulties when you use it to determine the size of an obtuse angle in a triangle. It still works for obtuse triangles, but the problem is that angles in the second quadrant share sine values with angles in the first quadrant. The inverse sine function is set to only give you angles from  $-90$  degrees to  $+90$  degrees. Therefore you will not be able to get an answer of an obtuse angle using the law of sines with the inverse sine function. You'll get an angle that has the desired sine value, but it is not the right one! By now you may not remember much geometry, but there is (fortunately for teachers) no ASS congruence. If one angle and two sides of a triangle are known, and the angle is not between the two known sides, the triangle is not necessarily determined. If the known angle is acute, there can be two possibilities for the actual shape of the triangle. One of those

possibilities corresponds to the sine value for the obtuse angle, while the other one has an acute angle with the same sine value. Shown below are two triangles with sides of 7 and 8, and a known angle of  $50^\circ$  (rounded values). To distinguish between these two possibilities, your problem may specify either angle C or the length of the remaining side. If you're not careful with the Law of Sines while trying to find the measure of angle B, you will get the left triangle when the actual triangle may be the one on the right.



To avoid this problem, look in advance for the largest angle in the triangle. **The largest angle is always opposite the longest side.** Then solve for the smaller angles first. Because a triangle can only have one obtuse angle, it is always possible to leave it till last.

## Trigonometric Identities

There are an awful lot of these things, and some teachers expect you to memorize most of the significant ones. If you are lucky you get to use a formula sheet for tests, so you can get away with just memorizing a few. Why memorize any at all, you may wonder. Well, knowing just a

few of the crucial ones will help you get your work done a lot faster. We have already seen three major identities:

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

This is the most important trig identity, and it can be found easily by applying the Pythagorean Theorem to the unit circle. It says that (the sine of angle  $x$ )<sup>2</sup> plus (the cosine of angle  $x$ )<sup>2</sup> must equal 1.

Dividing both sides by  $\cos^2 x$ , we get

**2.**  $\tan^2 x + 1 = \sec^2 x$

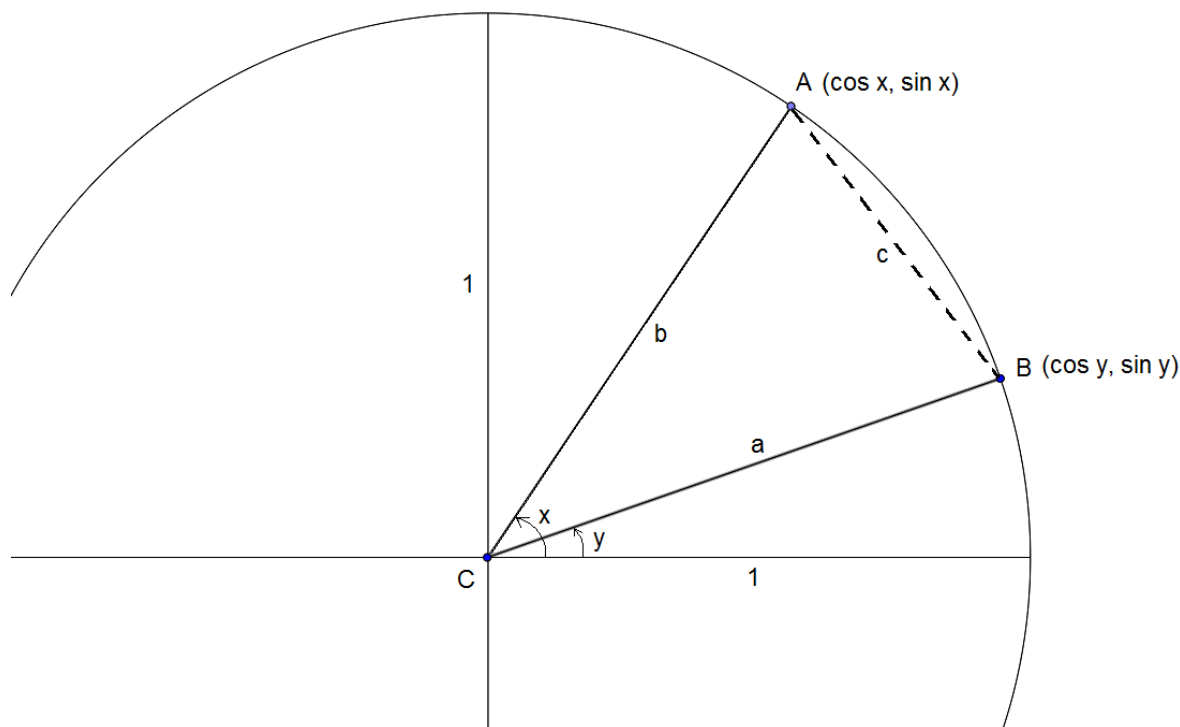
We can also divide by  $\sin^2 x$  to get

**3.**  $1 + \cot^2 x = \csc^2 x$

Now we will use the Law of Cosines to derive another important identity. This trick, probably discovered after some trial and error, provides an alternate expression for the cosine of angle  $x$  plus angle  $y$ . Because the cosine function is not a straight line, you cannot find the cosine of say 50 degrees by taking the sum of the cosine of 30 degrees and the cosine of 20 degrees. That last part may seem a bit obvious now, but take note of it because it is easy to forget when you're up to your ears in sines and cosines of various unknown angles.

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

Unfortunately, this identity must be derived from the cosine of angle  $(x - y)$ . The picture below shows angle  $x$  and angle  $y$  drawn in the unit circle.



Point A is the terminal point of angle  $x$ , so its coordinates are  $(\cos x, \sin x)$ , and point B is the terminal point of angle  $y$  with coordinates  $(\cos y, \sin y)$ . If you consider triangle ABC, you will see that the angle at C is  $(x - y)$ . Triangle ABC has no right angles, so we cannot use the Pythagorean Theorem. However, we can use the law of cosines:  $c^2 = a^2 + b^2 - 2ab \cos C$ .

Because we are working in the unit circle, the length of  $a$  and  $b$  is 1:

$$c^2 = 1^2 + 1^2 - 2 \cdot 1 \cdot 1 \cdot \cos(x - y)$$

$$c^2 = 2 - 2 \cos(x - y)$$

This gives one expression for the length of  $c$ . However, because we have the coordinates of point A and point B we can also use the distance formula (Pythagorean Theorem) to calculate  $c^2$ :

$$c^2 = (\sin x - \sin y)^2 + (\cos y - \cos x)^2$$

$$c^2 = \sin^2 x - 2 \sin x \cdot \sin y + \sin^2 y + \cos^2 y - 2 \cos y \cdot \cos x + \cos^2 x$$

$$c^2 = \sin^2 x + \cos^2 x + \sin^2 y + \cos^2 y - 2 \sin x \sin y - 2 \cos x \cos y$$

As we saw before,  $\sin^2 x + \cos^2 x$  is equal to 1, and so is  $\sin^2 y + \cos^2 y$ :

$$c^2 = 1 + 1 - 2\sin x \sin y - 2\cos x \cos y$$

$$c^2 = 2 - 2\sin x \sin y - 2\cos x \cos y$$

However, we just figured out that  $c^2$  is also equal to  $2 - 2\cos(x - y)$ . Therefore:

$$2 - 2\cos(x - y) = 2 - 2\sin x \sin y - 2\cos x \cos y$$

$$-2\cos(x - y) = -2\sin x \sin y - 2\cos x \cos y$$

Dividing both sides by -2, we get

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

That's nice, but we'd rather have an expression for  $\cos(x + y)$ . To change that, we can insert  $(-y)$  into the formula:

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

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

Fortunately, there is some symmetry here. The cosine of  $y$  is always equal to the cosine of  $-y$ . Just look at the unit circle to see that it is so. The sine of  $y$  is equal in distance but opposite in value to the sine of  $-y$ :  $\sin -y = -\sin y$ . Now we can get the formula we want:

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

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

Once mathematicians had this formula, they wanted a similar one for  $\sin(x + y)$ . Here it would help to look back at the graph that showed how the sine and cosine are related. Hopefully you verified that  $\sin x = \cos\left(\frac{\pi}{2} - x\right)$  and  $\cos x = \sin\left(\frac{\pi}{2} - x\right)$ , or maybe you should do that now.

Using this fact we can say that  $\cos(x + y) = \sin\left(\frac{\pi}{2} - (x + y)\right) = \sin\left(\frac{\pi}{2} - x - y\right)$ , and so we can change part of our cosine addition formula:

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

$$\sin\left(\frac{\pi}{2} - x - y\right) = \sin\left(\frac{\pi}{2} - x\right)\cos y - \sin x \sin y$$

Now, angle  $x$  is some arbitrary unknown angle.  $\frac{\pi}{2} - x$  is also an arbitrary unknown angle. Let's call it  $z$  for a moment, so we don't get confused:

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

Hmm, this close, but one bit that still contains  $x$ :  $\sin x$ . We already know that  $\sin x = \cos\left(\frac{\pi}{2} - x\right)$ , and we say that  $\frac{\pi}{2} - x = z$  so we can replace  $\sin x$  with  $\cos z$ :

$$\sin(z - y) = \sin z \cos y - \cos z \sin y$$

The letter  $z$  is not as commonly used for variables as  $x$ , so let's change  $z$  to  $x$ :

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

Oops, that worked out wrong because we would prefer a formula for  $\sin(x + y)$ . Well, we can just change it as we did before, by substituting  $-y$  for  $y$ :

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

Looking at the unit circle, we find that while  $\cos(-y) = \cos y$ ,  $\sin(-y)$  is always  $-\sin y$ :

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

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

Together,  $\cos(x + y) = \cos x \cos y - \sin x \sin y$  and  $\sin(x + y) = \sin x \cos y + \cos x \sin y$  are known as the angle addition formulas.

These useful formulas are even more helpful when angle  $x =$  angle  $y$ . Just put  $x = y$  to translate the above formulas into the following:

$$\sin(2x) = 2 \sin x \cos x$$

$$\cos(2x) = \cos^2 x - \sin^2 x$$

There are times when it is awkward to deal with the square of a sine or cosine function, especially in calculus. Mathematicians have rearranged that last formula to help them get rid of such squared functions without getting an annoying square root:

$$\cos(2x) = \cos^2 x - \sin^2 x$$

$$\cos(2x) = \cos^2 x - \sin^2 x$$

$$\cos^2 x = \cos(2x) + \sin^2 x$$

$$\sin^2 x = \cos^2 x - \cos(2x)$$

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

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

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

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

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

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

Once you have these formulas, you can replace  $x$  with  $x/2$  to derive the half-angle formulas:

$$\cos \frac{x}{2} = \pm \sqrt{\frac{1 + \cos x}{2}}$$

$$\sin \frac{x}{2} = \pm \sqrt{\frac{1 - \cos x}{2}}$$

Although it is not commonly used, there is also a tangent addition formula:

$$\tan(x + y) = \frac{\sin(x+y)}{\cos(x+y)}$$

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

Divide both the top and the bottom of this fraction by  $\cos x \cos y$ :

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

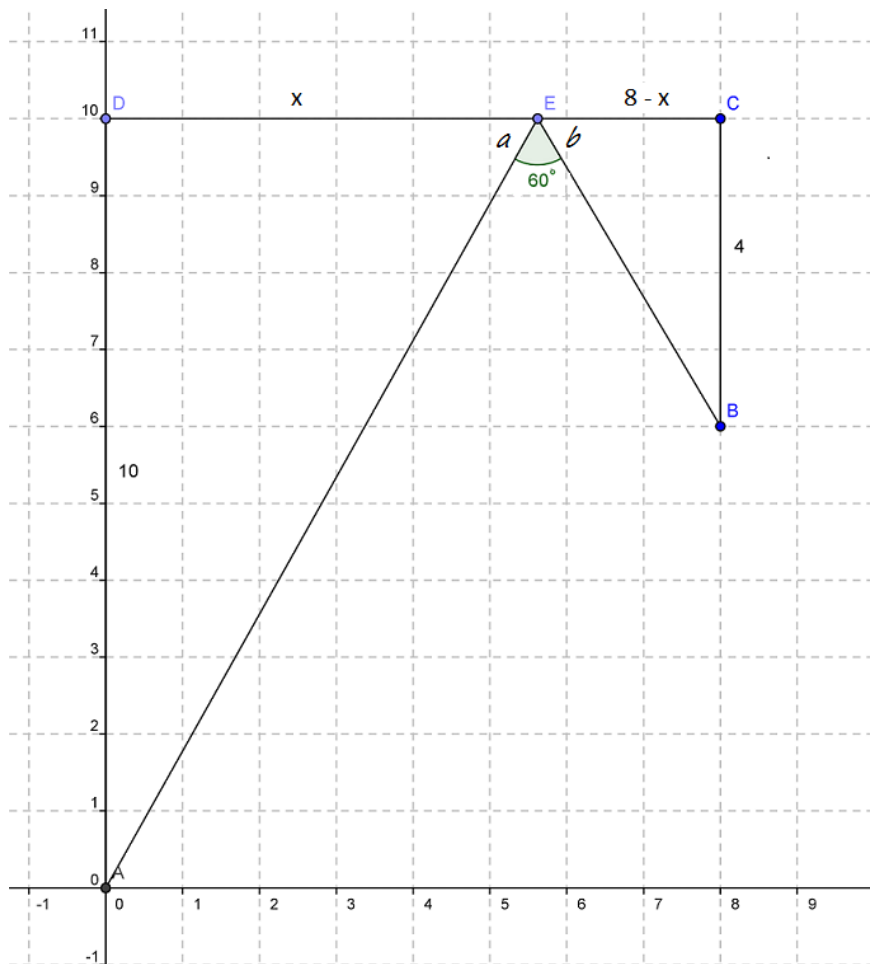
$$\tan(x + y) = \frac{\frac{\sin x}{\cos x} \cdot \frac{\cos y}{\cos y} + \frac{\cos x}{\cos x} \cdot \frac{\sin y}{\cos y}}{\frac{\cos x}{\cos x} \cdot \frac{\cos y}{\cos y} - \frac{\sin x}{\cos x} \cdot \frac{\sin y}{\cos y}}$$

$$\tan(x + y) = \frac{\tan x + \tan y}{1 - \tan x \tan y}$$

Here is the only problem I have encountered that actually creates a use for the angle addition formula for the tangent:

For the following figure,  $E$  is a point that moves along line  $DC$ . Find  $x$  when the angle between  $a$  and  $b$  is 60 degrees as shown.





When the angle is 60 degrees as required, angles  $a$  and  $b$  add up to 120 degrees. Since the tangent of 120 degrees is -1.732, we can say that  $\tan(a + b) = -1.732$ . However,  $\tan(a + b)$  is **not**  $\tan(a) + \tan(b)$ . The tangent addition formula tells us that  $\tan(a + b) = \frac{\tan(a) + \tan(b)}{1 - \tan(a)\tan(b)}$ .

Reading from the figure, we find that  $\tan(a) = 10/x$ , and  $\tan(b) = 4/(8 - x)$ .

$$\begin{aligned} \frac{\tan(a) + \tan(b)}{1 - \tan(a)\tan(b)} &= \frac{\frac{10}{x} + \frac{4}{8-x}}{1 - \left(\frac{10}{x}\right)\left(\frac{4}{8-x}\right)} = \frac{\frac{10(8-x) + 4x}{x(8-x)}}{1 - \left(\frac{10}{x}\right)\left(\frac{4}{8-x}\right)} = \frac{\frac{80-10x+4x}{8x-x^2}}{1 - \frac{40}{8x-x^2}} = \frac{\frac{80-6x}{8x-x^2}}{\frac{1(8x-x^2) - 40}{8x-x^2}} = \\ \frac{80-6x}{8x-x^2-40} &= \frac{80-6x}{8x-x^2} \cdot \frac{8x-x^2}{8x-x^2-40} = \frac{80-6x}{8x-x^2-40}. \end{aligned}$$

Since we know that the tangent we want is -1.732, we solve  $\frac{80-6x}{-x^2+8x-40} = -1.732$ :

$$1.732x^2 - 13.856x + 69.28 = 80 - 6x$$

$$1.732x^2 - 7.856x - 10.72 = 0$$

By using the quadratic formula we can determine that the positive solution for  $x$  is 5.634.

## Trigonometric Equations

Trigonometric equations may be presented to you in different ways. You may be expected to solve an equation and come up with a specific value or set of values. For example, if an equation specifies that  $2 \sin x = 1$ , that implies that we need to find one or more values for  $x$  that will make the equation true. Find the set of angles for which  $\sin x = \frac{1}{2}$ .

To actually solve for an angle, you need to look at whether there is a restriction on the possible values for the angle. Most commonly such a restriction will involve values between 0 and  $2\pi$  (0 to 360 degrees).

### Example

$\sin x + \cos x = 1$ . Solve for  $x$ .

In general, you want to have only a single trigonometric function in your equation. You could convert the cosine to the sine by using the identity  $\sin^2 x + \cos^2 x = 1$ . This means that  $\cos x = \pm\sqrt{1 - \sin^2 x}$ . Substituting that into the equation gives:

$$\sin x \pm \sqrt{1 - \sin^2 x} = 1$$

To solve equations containing a square root, you should isolate the square root on one side and then square everything:

$$\pm\sqrt{1 - \sin^2 x} = 1 - \sin x$$

$$1 - \sin^2 x = (1 - \sin x)^2$$

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

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

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

If you can get to a point where there is a zero on one side, you can often solve the equation by factoring:

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

$$0 = 2 \sin x(\sin x - 1)$$

$$0 = \sin x(\sin x - 1)$$

Either  $\sin x = 0$  or  $\sin x - 1 = 0$ .

Because in this case we squared a quantity that could be either negative or positive, we have to check our answers in the original equation (which is always a good idea anyway). If  $\sin x = 0$  then the first angle in the unit circle where that happens is 0. It happens again at  $\pi/2$  radians or 180 degrees, and every 180 degrees after that. If you put  $x = 0$  in the original equation, you get  $\sin(0) + \cos(0) = 1$ , which is in fact true because the cosine of 0 is 1. On the other hand,  $\sin(\frac{\pi}{2}) + \cos(\frac{\pi}{2}) = -1$ . At  $x = 2\pi$ , things work out again because the cosine is 1. One way of writing this answer is  $x = 0 \text{ degrees} + 2\pi n$ , where  $n$  is a positive or negative whole number, or zero. If  $\sin x - 1 = 0$ , then  $\sin x$  must be equal to 1. This happens at  $\pi/2$  and every 360 degrees after that. When the sine is equal to 1, the cosine is always 0, so there is no problem. We could write that as  $x = 1 \pm 2\pi n$ , where  $n$  is a positive whole number, or zero.

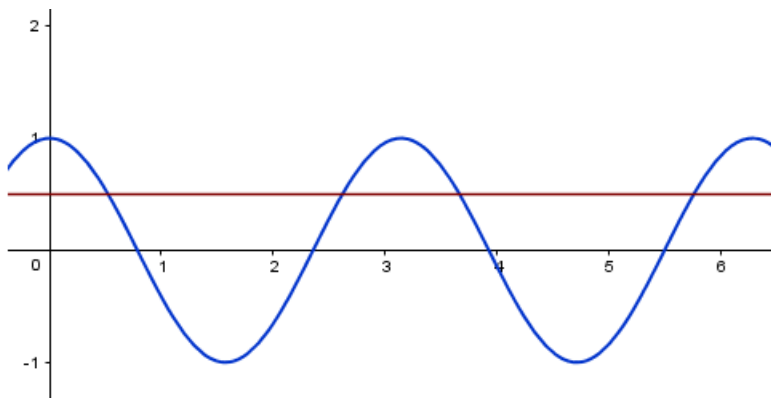
### Example

$\cos(2x) = \frac{1}{2}$ . Solve for  $x$  if  $0 \leq x < 2\pi$ .

Be careful about the period here. The multiplication by 2 means that the period is now  $\pi$  instead of  $2\pi$ , and  $\cos(2x)$  will reach a value of  $\frac{1}{2}$  four times in the interval  $[0, 2\pi]$ . To find all of the values of  $x$ , double the length of the interval where you look for the angles:

$$2x = \frac{\pi}{3} \text{ or } \frac{5\pi}{3} \text{ or } \frac{7\pi}{3} \text{ or } \frac{11\pi}{3}$$

$$x = \frac{\pi}{6} \text{ or } \frac{5\pi}{6} \text{ or } \frac{7\pi}{6} \text{ or } \frac{11\pi}{6}$$



### Example

$\sin^2 x = \sin x$ . Solve for  $x$  if  $0 \leq x < 2\pi$ .

Here you may be tempted to divide both sides of the equation by  $\sin x$ , but remember that this is an unknown. **Make sure that an unknown quantity can't be zero before you divide by it!**

Could  $\sin x$  be zero in this case? Yes, it can, and now we already have one solution. So,

$$\sin x = 0, \text{ or}$$

$$\frac{\sin^2 x}{\sin x} = \frac{\sin x}{\sin x}$$

$$\sin x = 1$$

When  $\sin x = 0$ ,  $x$  is 0. When  $\sin x = 1$ ,  $x = \pi/2$ .

Another way of solving this problem, and avoiding a potential division by zero, is to move the parts with unknown to one side:

$$\sin^2 x = \sin x$$

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

$$\sin x(\sin x - 1) = 0$$

$$\sin x = 0 \text{ or } \sin x - 1 = 0.$$

Other times an equation is presented as a statement that needs to be proven by using trigonometric identities to change one of the sides until it is identical to the other. We'll look at strategies for doing that first, and then go on to find actual values of angles.

It can be quite overwhelming to attempt to prove complex trigonometric identity equations, but it helps to use a systematic approach. First, find out if you are only supposed to change the left side of the equation, or if you are allowed to pick either side. Usually you are not permitted to change both. Here are some basic strategies to help you:

1. Write everything in terms of only  $\sin x$  and  $\cos x$ , especially if that is all you see on the other side.
2. Watch for the three obvious identities:  
 $\sin^2 x + \cos^2 x = 1$        $\tan^2 x + 1 = \sec^2 x$        $1 + \cot^2 x = \csc^2 x$   
 Try to rearrange things to take advantage of these identities.
3. If the equation contains fractions, try putting everything over a common denominator so you have a single term under the division line. Remember that to divide by a fraction, you multiply by the reciprocal. For example:  

$$\frac{\sin x - \cos x}{1 - \frac{\sin x}{\cos x}} = \frac{\sin x - \cos x}{\frac{\cos x}{\cos x} - \frac{\sin x}{\cos x}} = \frac{\sin x - \cos x}{\frac{\cos x - \sin x}{\cos x}} = -(\cos x - \sin x) \cdot \frac{\cos x}{\cos x - \sin x} = -\cos x.$$
4. Reduce the angles, using double angle and/or angle addition formulas. If there is an uneven angle, like  $\sin(3x)$ , split it into  $\sin(2x + x)$ . Note that this is **not**  $\sin(2x) + \sin(x)$ !
5. Factor it if you can.
6.  $\sin^2 x = 1 - \cos^2 x = (1 + \cos x)(1 - \cos x)$
7. If  $\sin^4 x$  or  $\cos^4 x$  are present, factor as a difference of two squares if possible.  $\sin^4 x - 1 = (\sin^2 x - 1)(\sin^2 x + 1) = (\sin x - 1)(\sin x + 1)(\sin^2 x + 1)$ . Otherwise, treat it as  $(\sin^2 x)(\sin^2 x)$  or  $(\cos^2 x)(\cos^2 x)$ .
8. If you see an expression like  $(1 - \cos x)$  or  $(1 + \sin x)$  in a fraction, try to multiply both the numerator and denominator by the complementary terms  $(1 + \cos x)$  or  $(1 - \sin x)$  to take advantage of the difference of two squares.  $1 - \cos^2 x$  changes to  $\sin^2 x$ .
9. If you have terms with  $\sin x$  and  $\cos x$  on one side, and only  $\csc x$  and  $\sec x$  on the side that you must keep the same, you can try changing things like this:  $\sin x = \frac{1}{\frac{1}{\sin x}} = \frac{1}{\csc x}$   
 and  $\cos x = \frac{1}{\frac{1}{\cos x}} = \frac{1}{\sec x}$ .
10. Even if you are not allowed to change the right (or left) side of the equation, it may help to do so anyway and then work backwards to get the answer the proper way.

## Modeling Change with Sine (or Cosine) Functions

You should follow along with the explanations in this section by using some kind of graphing software, like Math GV Function Plotting Software (free download at <http://www.mathgv.com>).

Select File -> New 2D Cartesian Graph, and then Graph -> New 2D Function. You can use pi to represent  $\pi$ , and put the value of the sine in parentheses, like this:  $\sin(2\pi \cdot x)$  or  $\sin(2\pi x)$ . Start with the basic sine function  $y = \sin(x)$ .

Not only is the sine curve an interesting curve by itself, there are also many real events that can be described by this type of function because they go up and down periodically in the same way as the sine function. We can model these events with a sine curve by stretching or squishing the curve horizontally and/or vertically, and shifting it to the right or left, or up and down. This works just like you would do it with other functions. It just looks scarier because of the sine notation. One reason that educators like these problems is that they show horizontal stretches and compressions much better than those that don't use trig functions.

First, let's move our sine function over a little. The sine of 0 is 0, and while many events start out with a value of zero at time zero, some do not. Suppose that we are modeling a process that starts out at its maximum value at time 0. Our sine function has a maximum at  $\pi/2$ , so let's shift the whole function over to the left to get the maximum value positioned at time zero. To move a function to the left, we add something to  $x$ . When you do that, you create a situation where a smaller  $x$  can do the same job that a bigger  $x$  did previously. I can shift the whole curve  $\pi/2$  units to the left by adding  $\pi/2$  to  $x$ , like this:  $f(x) = \sin\left(x + \frac{\pi}{2}\right)$ . Now  $x$  can be 0 and still give the function a value of  $\sin\left(\frac{\pi}{2}\right)$ . In fact, the entire curve has shifted to the left. This makes it identical to the curve of the cosine function. Your course may direct you to use the cosine function for modeling if the maximum is at time 0 or if you need to have the maximum at a specified value. That is more efficient than starting with the sine function.

Next, we will consider that our imaginary process has a period corresponding to 1 second. Seconds you say? I thought we were working in radians, with stuff like  $2\pi$ ? Although this situation would be potentially confusing, we are fortunately just graphing things on a piece of paper or on a computer screen. We only need to have the function line up with simple "units" along the  $x$ -axis, and then we can call those units seconds, or hours, or inches or whatever is appropriate to the situation. Right now we are calling the units radians, and one period of the function extends over  $2\pi$  of them. Let's change that a little. I am going to multiply  $x$  by  $2\pi$ . This way, when the input is actually 1, the multiplication by  $2\pi$  blows the value up to  $2\pi$ . Once that has happened, the sine function really doesn't see the difference. The function still does the same job, but it now can do that job with a scaled-down value of  $x$ . As a result, our graph has become squished up in the horizontal direction by a factor of  $2\pi$ . The period is now 1 instead of  $2\pi$ . Just graph  $y = \sin(2\pi x)$  to see that. When  $x$  is 0.5,  $2\pi x$  is  $\pi$  so we are at the half-way point in the period. The entire period is 1 unit, and instead of radians we'll label that seconds on our graph.

But wait, what about our shift to the left? Well, because the function is already where we wanted it at  $x = 0$  it won't move away from that as it gets squished:  $y = \sin\left(2\pi x + \frac{\pi}{2}\right)$  still has a maximum value of  $\sin \frac{\pi}{2}$  when  $x$  is zero. The only thing that needs to be multiplied by  $2\pi$  is  $x$ . But suppose you had started with the graph  $y = \sin(2\pi x)$ , and you wanted the maximum value at  $x = 0$ . You would look at the period, which is now 1, and reason that a change of  $\frac{1}{4}$  unit in  $x$  would be needed to shift the graph to the left. Add  $\frac{1}{4}$  *directly to*  $x$  to create the shift. Graph  $y = \sin(2\pi(x + \frac{1}{4}))$ . You get a sine wave with a period of 1 that is shifted  $\frac{1}{4}$  units to the left.  $2\pi(x + \frac{1}{4})$  is the same as  $2\pi x + \pi/2$ . The function can be written as  $f(x) = \sin(2\pi x + \pi/2)$ , or as  $f(x) = \sin(2\pi(x + \frac{1}{4}))$ . This last form has the advantage of allowing you to read the actual magnitude of the shift directly. The bottom line here is that **whatever you want to change, do it directly to  $x$ , and you'll get what you want.**

Once we have reached this point, it is a really a simple matter to stretch the period to have it occupy as many seconds as we want. For example, if we want a period of 6 seconds we just divide the part with the  $x$  by 6. If you graph  $y = \sin\left(\frac{2\pi x}{6} + \frac{\pi}{2}\right)$ , you'll see that you get the desired result. The actual magnitude of the shift can be seen by factoring out  $\frac{2\pi}{6}$  and rewriting the equation as  $y = \sin\left(\frac{2\pi}{6}\left(x + \frac{3}{2}\right)\right)$ . Here we are using seconds, but the units are really arbitrary.  $y = \sin\left(\frac{2\pi}{365}x\right)$  would nicely model a process that starts at zero and has a period of 365 days.  $y = \sin\left(\frac{2\pi}{365}x + \frac{\pi}{2}\right)$  will give you a maximum at zero and a period of 365 days. If you need to shift either of these equations, just add or subtract the required values directly to or from  $x$ .

Problems in your course may ask you to determine the frequency. In physics, the frequency is thought of as how many waves go by a given point per second, or some other unit of time. Suppose that a wave completes a full cycle in 0.1 second. We say that the period is 0.1 second. The frequency is 10 waves per second. If the period is 2 seconds, then only 0.5 waves go by every second. The frequency is simply *1/the period*.

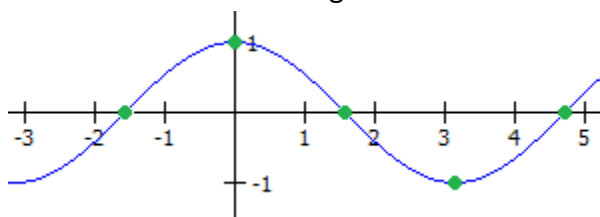
Well, that takes care of the horizontal adjustments. The vertical changes we make are actually simpler to understand because they can be made as direct changes to the output. If we want the output of the function  $f(x) = \sin(x)$  to be 3 times as big, we just write  $f(x) = 3 \sin(x)$ . This causes the value of the function to vary between -3 and 3 rather than between -1 and 1. The function becomes stretched out in the vertical direction. Shifting the entire function up or down is a simple matter of adding or subtracting the desired value. If we wanted  $f(x) = 3 \sin(x)$  to return only positive values, we just add 3 to the output, like this:  $f(x) = 3 \sin(x) + 3$ . Now the value of the function will vary between 0 and 6 rather than between -3 and 3.

Instead of using the sine function, you could use the cosine function. This is often convenient when a process has a maximum value at time zero, since that is where the cosine has its maximum.

## Graphing a Modified Trigonometric Function

Example: Graph  $y = 3 \cos\left(\frac{1}{\pi}x + \frac{1}{\pi}\right) + 1$

1. Make a sketch of the original cosine function.



I chose to include a full period between  $-\pi/2$  and  $3\pi/2$ . Notice that the length of the period is  $2\pi$ , or about 6.28. The green dots mark minimum and maximum values, and the zero points (points on the midline of the function).

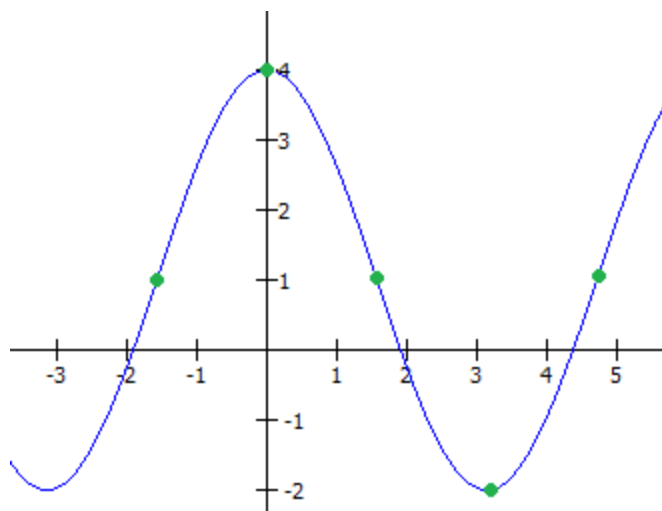
Location of green dots from left to right:

$$\left(-\frac{\pi}{2}, 0\right) \quad (0, 1) \quad \left(\frac{\pi}{2}, 0\right) \quad (\pi, -1) \quad \left(\frac{3\pi}{2}, 0\right)$$

2. Look at the problem like this:  $y = 3 \cdot (\text{cosine of something}) + 1$ .

We know that the value of the cosine always goes from -1 to 1. This means that 3 times the cosine has a range of -3 to 3. The last part, + 1, says that 1 unit must be added to the output of the cosine function. Now the output values of the function will range from -2 to 4. Graph this using MathGV or another program as  $y = 3 \cos(x) + 1$  so you can see what it looks like:



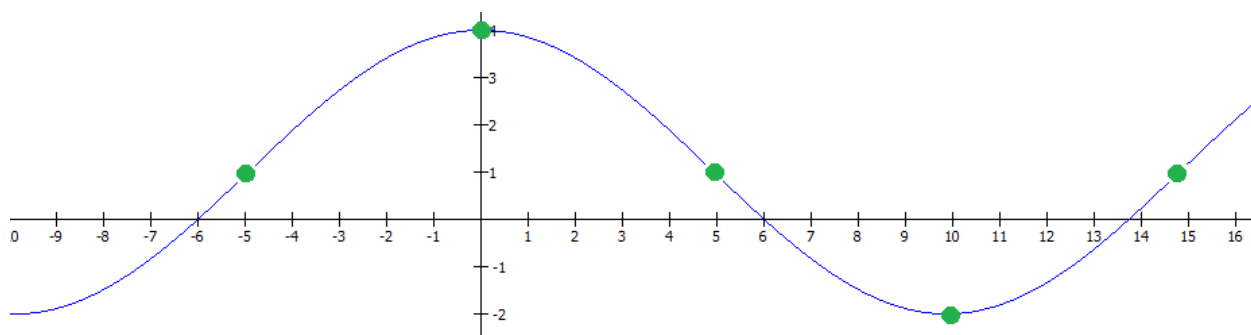


Location of green dots from left to right:

$$\left(-\frac{\pi}{2}, 1\right) \quad (0, 4) \quad \left(\frac{\pi}{2}, 1\right) \quad (\pi, -2) \quad \left(\frac{3\pi}{2}, 1\right)$$

The midline of the function is now at  $y = 1$ .

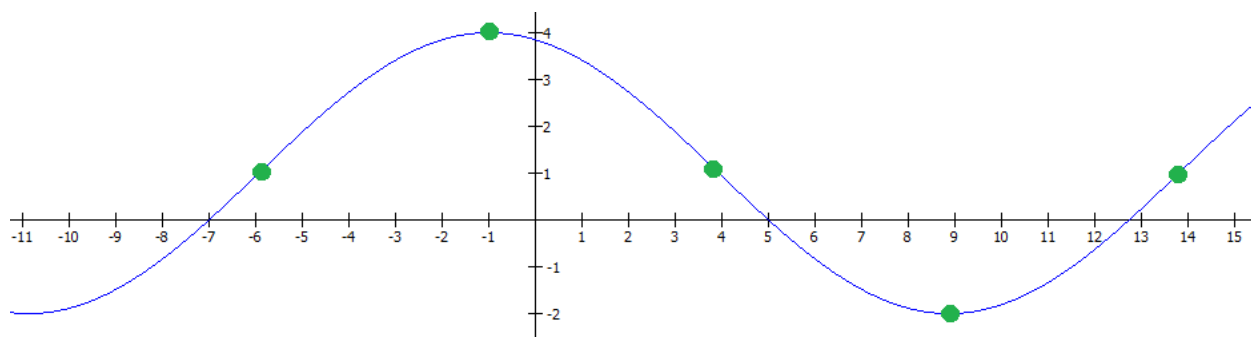
3. Find the new period.  $y = 3 \cos\left(\frac{1}{\pi}x + \frac{1}{\pi}\right) + 1$ . Ignore the other parts for a moment, and look only at  $\cos\left(\frac{1}{\pi}x\right)$ . When  $x = 0$ , the cosine function “sees” a value of zero and returns its maximum value. For the cosine function to “see” a full input of  $2\pi$ ,  $x$  must be  $2\pi^2$ . The multiplication by  $\frac{1}{\pi}$ , which is the same as a division by  $\pi$ , caused the period to stretch by a factor of  $\pi$ . The new period is  $2\pi^2$ , or about 19.7 radians.  $y = 3 \cos\left(\frac{1}{\pi}x\right) + 1$  looks like this (notice that we have ignored the horizontal shift for the moment):



Location of green dots (from left to right):

$$\left(-\frac{\pi^2}{2}, 1\right) \quad (0, 4) \quad \left(\frac{\pi^2}{2}, 1\right) \quad (\pi^2, -2) \quad \left(\frac{3\pi^2}{2}, 1\right)$$

4. Now we rewrite the equation so that we can quickly see the shift. Factor out  $\frac{1}{\pi}$  to see that  $y = 3 \cos\left(\frac{1}{\pi}x + \frac{1}{\pi}\right) + 1$  is the same as  $y = 3 \cos\left(\frac{1}{\pi}(x + 1)\right) + 1$ . The graph of  $y = 3 \cos\left(\frac{1}{\pi}x\right) + 1$  must be shifted 1 unit to the left:



Location of green dots (from left to right):

$$\left(-\frac{\pi^2}{2} - 1, 1\right) \quad (-1, 4) \quad \left(\frac{\pi^2}{2} - 1, 1\right) \quad (\pi^2 - 1, -2) \quad \left(\frac{3\pi^2}{2} - 1, 1\right)$$

5. Convert to numbers so you can see that the dots are in the right place:

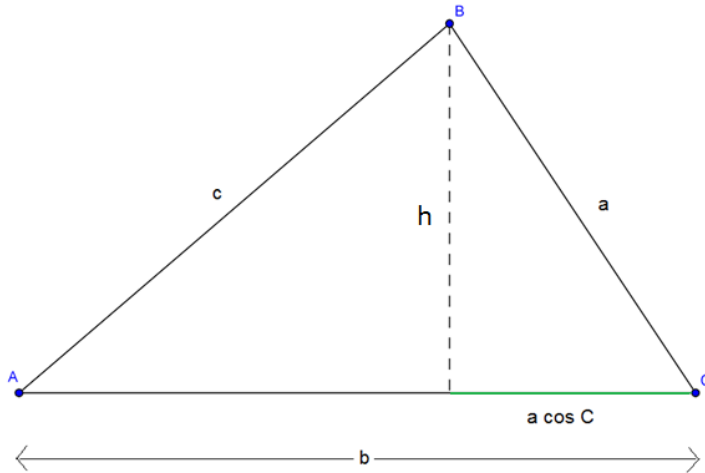
$$(-5.93, 1) \quad (-1, 4) \quad (3.93, 1) \quad (8.87, -2) \quad (13.80, 1)$$

## Heron's Formula

You might remember from geometry that once you choose the length of all three sides of a triangle, you have completely determined that triangle. The angles are now fixed, and so is the area. As a result, you should be able to calculate the area of a triangle if the length of the sides are given. People commonly use Heron's Formula for this:

$$A = \sqrt{s(s-a)(s-b)(s-c)}, \text{ where } s \text{ is the semiperimeter of the triangle. } s = \frac{a+b+c}{2}$$

Heron's formula can be derived using the Pythagorean Theorem. Take triangle ABC and draw an altitude  $h$ .



From the law of cosines, we know that  $c^2 = a^2 + b^2 - 2ab \cos C$ . So,  $c^2 - a^2 - b^2 = -2ab \cos C$ , and  $\cos C = \frac{c^2 - a^2 - b^2}{-2ab}$ . The green line in the image above has length  $a \cos C$ , which is equal to  $\frac{c^2 - a^2 - b^2}{-2b}$ . We can use this expression in the Pythagorean Theorem.  $h^2 + (a \cos C)^2 = a^2$ .

Because we are interested in  $h$ , we'll write that as  $h^2 = a^2 - (a \cos C)^2$ . Substituting  $\frac{c^2 - a^2 - b^2}{-2b}$  for  $a \cos C$  gives  $h^2 = a^2 - \left(\frac{c^2 - a^2 - b^2}{-2b}\right)^2$ , which is the same as  $h^2 = a^2 - \frac{(c^2 - a^2 - b^2)^2}{4b^2}$ . Since we need to take a square root here, it will be best to create a single fraction:  $h^2 = \frac{4a^2b^2}{4b^2} - \frac{(c^2 - a^2 - b^2)^2}{4b^2} = \frac{4a^2b^2 - (c^2 - a^2 - b^2)^2}{4b^2}$ .

That last expression is rather convenient because the denominator is a nice square. Recall that we can take the square root of the numerator and the denominator separately:

$$h = \sqrt{\frac{4a^2b^2 - (c^2 - a^2 - b^2)^2}{4b^2}} = \frac{\sqrt{4a^2b^2 - (c^2 - a^2 - b^2)^2}}{2b}$$

The area of the square will be one-half the base times the height  $h$ . The base in this example is

$$b, \text{ so } A = \frac{1}{2}bh = \frac{1}{2}b \cdot \frac{\sqrt{4a^2b^2 - (c^2 - a^2 - b^2)^2}}{2b} = \frac{\sqrt{4a^2b^2 - (c^2 - a^2 - b^2)^2}}{4}$$

Although you could use this formula the way it is, Heron's Formula looks much nicer. To fix things up a bit, we can rewrite the top part of the fraction as the difference of two squares:

$$A = \frac{\sqrt{(2ab)^2 - (c^2 - a^2 - b^2)^2}}{4}$$

The difference of two squares can be factored  $[(a + b)^2 = (a + b)(a - b)]$ :

$$A = \frac{\sqrt{(2ab+c^2-a^2-b^2)(2ab-c^2+a^2+b^2)}}{4}$$

If you remember that  $(a + b)^2 = a^2 + 2ab + b^2$  and  $(a - b)^2 = a^2 - 2ab + b^2$ , you can use that to write each of the two factors of the numerator as the difference of two squares:

$$A = \frac{\sqrt{-(-2ab-c^2+a^2+b^2)(2ab-c^2+a^2+b^2)}}{4}$$

$$A = \frac{\sqrt{-(a^2-2ab+b^2-c^2)(a^2+2ab+b^2-c^2)}}{4}$$

$$A = \frac{\sqrt{-((a-b)^2-c^2)((a+b)^2-c^2)}}{4}$$

$$A = \frac{\sqrt{-(a-b+c)(a-b-c)(a+b+c)(a+b-c)}}{4}$$

$$A = \frac{\sqrt{(a-b+c)(-a+b+c)(a+b+c)(a+b-c)}}{4}$$

$$A = \frac{\sqrt{(a-b+c)(-a+b+c)(a+b+c)(a+b-c)}}{4}$$

$$A = \frac{\sqrt{(p-2b)(p-2a)p(p-2c)}}{4}, \text{ where } p \text{ is the perimeter } a + b + c$$

$$A = \frac{\sqrt{2s(2s-2a)(2s-2b)(2s-2c)}}{4}$$

$$A = \frac{\sqrt{2s \cdot 2(s-a) \cdot 2(s-b) \cdot 2(s-c)}}{4} = \frac{\sqrt{16s(s-a)(s-b)(s-c)}}{\sqrt{16}} = \sqrt{\frac{16s(s-a)(s-b)(s-c)}{16}}$$

$$A = \sqrt{s(s-a)(s-b)(s-c)}$$

## Equations of the Form $A \sin x + B \cos x$

Rearranging things into different forms is very important in math. There are many instances

where rearranging equations allows us to simplify something or get a form that is suitable for a particular method.

Let's go back to the trigonometric identity  $\sin(x + y) = \sin x \cos y + \cos x \sin y$ . Once we select a specific value for  $y$ ,  $\cos y$  and  $\sin y$  become fixed numbers. I'll pick an easy value, say  $60^\circ$  or  $\frac{\pi}{3}$ , to illustrate this point. I don't have the sine and cosine of  $\frac{\pi}{3}$  memorized, but I can quickly draw a unit circle and figure out that  $\sin \frac{\pi}{3} = \sqrt{3}/2$  and  $\cos \frac{\pi}{3} = 1/2$

$$\sin\left(x + \frac{\pi}{3}\right) = \cos \frac{\pi}{3} \sin x + \sin \frac{\pi}{3} \cos x$$

$$\sin\left(x + \frac{\pi}{3}\right) = \frac{1}{2} \sin x + \frac{\sqrt{3}}{2} \cos x$$

In fact, for every value of  $y$  that I choose I can create such an equation. If I choose the value  $n$  for  $y$ , it would look like this:

$$\sin(x + n) = A \sin x + B \cos x$$

Here the number  $A$  represents  $\cos y$ , and the number  $B$  is  $\sin y$ . Flip the equation around to get

$$A \sin x + B \cos x = \sin(x + n)$$

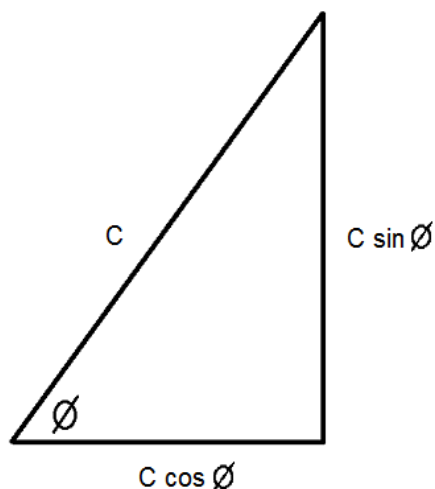
However, the value that we choose for  $y$  represents an angle, not just a number. The Greek letter phi, pronounced as either "fy" or "fee", is commonly used for this purpose. It is written as  $\phi$ :

$$A \sin x + B \cos x = \sin(x + \phi)$$

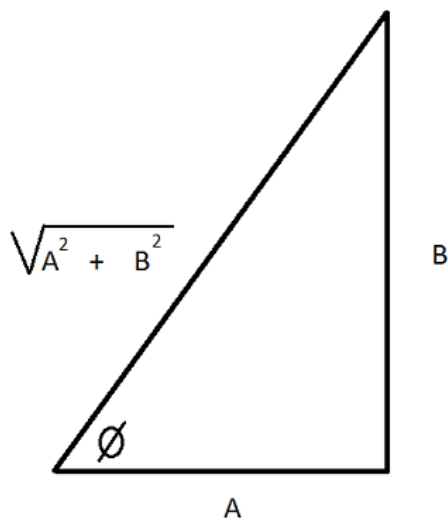
This shows that we can take the sum of a sine function, and a cosine function (which is really just a modified sine function) and express it as a single sine function. That is not very important to you and me, but folks in various engineering and physics fields actually have a use for this. Try it out for yourself by entering some functions into MathGV or your graphing calculator. For example, you can use  $y = 3\sin(x) + 4\cos(x)$ . When you graph this you get a nice modified sine function.

Also,  $\sin(x + \phi) = A \sin x + B \cos x$ , which shows that you could decompose a single sine wave into two new modified sine waves. If you were actually working with this equation you would not find it quite satisfactory because of the restrictions on the numbers  $A$  and  $B$ . Since these numbers represent the sine and cosine of angle  $\phi$  they are restricted to values between -1 and 1. Well at least we have negative values available, but they are inconveniently small. You can of course multiply the entire equation by something, like this:

$5 \sin(x + \phi) = 5A \sin x + 5B \cos x$ , which is fine in this direction but not so helpful when you're working the other way. Just look at  $25 \sin x + 14 \cos x$ : what was  $A$  and what was  $B$ ? Here our only reasonable option is to let 25 be  $A$  and 14 be  $B$ , and use some unknown multiplier  $C$  on the other side:  $A \sin x + B \cos x = C \sin(x + \phi)$ . Now  $A$  represents  $C \cos \phi$  and  $B$  is  $C \sin \phi$ . That may seem complicated, but we can draw a simple picture showing angle  $\phi$  and its associated values:



Look at this picture carefully and verify that it is correct. Now, since  $C \cos \phi = A$  and  $C \sin \phi = B$ :



This is nice, because we can see right away that  $C = \sqrt{A^2 + B^2}$ . So, instead of  $A \sin x + B \cos x = C \sin(x + \phi)$  we can write  $A \sin x + B \cos x = \sqrt{A^2 + B^2} \sin(x + \phi)$ . That's perfect. Starting with the function  $y = 25 \sin x + 14 \cos x$ , we can change it into a single sine function like this:

$$25 \sin x + 14 \cos x = \sqrt{25^2 + 14^2} \sin(x + \emptyset)$$

The specific angle  $\emptyset$  is easy to find, because we can see from the picture above that  $\tan \emptyset = \frac{B}{A}$ .

If  $\tan \emptyset = \frac{14}{25} = .56$ , then we can use the inverse tangent to find  $\emptyset$ . After I remember to set my calculator to radians, I get a value for  $\emptyset$  of .5105 radians. It is very important to use radians here rather than degrees because we are working with functions that we want to graph. (Recall that using degrees instead of radians would produce a distorted graph that is not really meaningful.)

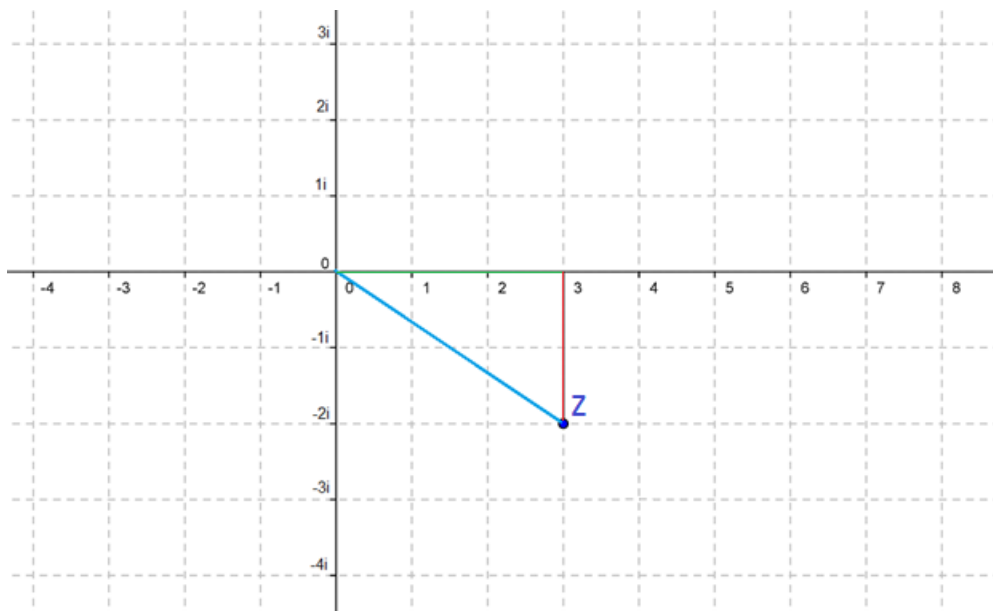
$$25 \sin x + 14 \cos x = 28.65 \sin(x + .5105)$$

To check your work, graph both  $y = 25 \sin x + 14 \cos x$  and  $y = 28.65 \sin(x + .5105)$ . Both functions will coincide exactly. You will have to zoom out a fair bit since we picked such large numbers for A and B. Try your own example with smaller numbers.

Once you start using negative numbers for A, which represents  $C \cos \emptyset$ , you will end up with an angle  $\emptyset$  outside the range of the inverse tangent function. This is always something to watch for when using the inverse trigonometric functions. In the case of the tangent your result will usually be off by  $\pi$  radians in a clockwise or counterclockwise direction. For example, if A is negative and B is positive then you should look for the angle  $\emptyset$  in the second quadrant (between  $\frac{\pi}{2}$  and  $\pi$  radians), where the cosine is negative and the sine is positive.

## Complex Numbers

In the section on imaginary numbers, we saw that the “size”, or modulus, of a complex number is its distance from the origin on the complex number plane:



This shows the modulus of the complex number  $z = 3 - 2i$  as a blue line. The length of the blue line, according to the Pythagorean Theorem, is  $\sqrt{3^2 + 2^2}$ , which is  $\sqrt{13}$ .  $|z| = \sqrt{13}$ .

Important mathematical discoveries were made by expressing complex numbers in trigonometric form, so of course you will be expected to be able to do that. That is, you'll be expected to convert long rows of numbers, not make mathematical discoveries. It never seems to occur to anyone that students might want to do something original.

The image above already contains a triangle, so you can just apply trigonometry to it. If the blue line makes an angle  $\theta$  with the origin, then  $\tan \theta = \frac{-2}{3}$ , and  $\theta = -33.69^\circ$ . The length of the green line would be  $\sqrt{13} \cos \theta$ , while the length of the red line is  $\sqrt{13} \sin \theta$ . We can therefore express the number  $z$  as  $\sqrt{13}(\cos -33.69^\circ + i \sin -33.69^\circ)$ , or as  $\sqrt{13}(\cos 326.3^\circ + i \sin 326.3^\circ)$ . The angle may be expressed in degrees or in radians.

Complex numbers can be multiplied:

### Example

Multiply  $3 - 2i$  [trig form  $\sqrt{13}(\cos 326.3^\circ + i \sin 326.3^\circ)$ ] by  $4 + 3i$  [trig form  $5(\cos 36.87^\circ + i \sin 36.87^\circ)$ ], and express the result in trigonometric form.

$$(3 - 2i)(4 + 3i) = 12 + 9i - 8i - 6i^2 = 12 + i + 6 = 18 + i$$



The modulus of this number is  $\sqrt{18^2 + 1^2} = \sqrt{325} = 5\sqrt{13}$ . To find the associated angle, take the inverse tangent of  $1/18$ , which is approximately 3.18 degrees. This is the actual angle, since we are expecting that it will be in the first quadrant.

In general, we could express two numbers in trigonometric form and then multiply:

$$r_1(\cos a + i \sin a) r_2(\cos b + i \sin b) = r_1 r_2 (\cos a \cos b + i \cos a \sin b + i \sin a \cos b + i^2 \sin a \sin b)$$

That can be simplified quite a bit:

$$r_1 r_2 (\cos a \cos b + i^2 \sin a \sin b + i \cos a \sin b + i \sin a \cos b)$$

$$r_1 r_2 (\cos a \cos b - \sin a \sin b + i (\cos a \sin b + \sin a \cos b))$$

Now use your trigonometric identities,  $\cos(x + y) = \cos x \cos y - \sin x \sin y$  and  $\sin(x + y) = \sin x \cos y + \cos x \sin y$ :

$$r_1 r_2 (\cos(a + b) + i \sin(a + b)).$$

We can use this to multiply  $3 - 2i [\sqrt{13}(\cos 326.3^\circ + i \sin 326.3^\circ)]$  and  $4 + 3i [5(\cos 36.87^\circ + i \sin 36.87^\circ)]$ , to get  $5\sqrt{13}(\cos 363.18^\circ + i \sin 363.18^\circ) = 5\sqrt{13}(\cos 3.18^\circ + i \sin 3.18^\circ)$ .

## De Moivre's Theorem: Proof by Induction

De Moivre's Theorem states that  $(r(\cos x + i \sin x))^n = r^n (\cos(nx) + i \sin(nx))$ . That makes sense since we just saw that you can multiply two complex numbers to get

$$r_1 r_2 (\cos(a + b) + i \sin(a + b)).$$

If both numbers are the same, that would give you  $r^2(\cos(2a) + i \sin(2a))$ . So, the theorem is correct for  $n = 2$ . But will things still work out for  $n = 3$ ? Hmm, I'd better try that to make sure. To save some time I can start with  $\cos(2x) + i \sin(2x)$  and multiply that by  $(\cos x + i \sin x)$ :

$$\begin{aligned} (\cos(2x) + i \sin(2x)) (\cos x + i \sin x) = \\ \cos(2x) \cos(x) + i \cos(x) \sin(2x) + i \sin(x) \cos(2x) - \sin(x) \sin(2x) \end{aligned}$$

That result can be rearranged to take advantage of the angle addition formulas  $\cos(x + y) = \cos x \cos y - \sin x \sin y$  and  $\sin(x + y) = \sin x \cos y + \cos x \sin y$ :

$$\cos(2x)\cos(x) - \sin(2x)\sin(x) + i (\sin(2x)\cos(x) + \cos(2x)\sin(x)) = \cos(2x + x) + i \sin(2x + x)$$

Well, it looks like  $(\cos x + i \sin x)^3$  is in fact equal to  $\cos(3x) + i \sin(3x)$ . At this point I would guess that the theorem works for  $n = 4$ , but I really don't feel like going on. And even if I did, all I would accomplish is to prove the theorem for a few specific cases out of an infinite total number. While I would settle for "It's definitely probably true," mathematicians feel the need to prove everything. A proof by induction can show that something is true not only for the first few cases, but on into infinity as  $n$  gets larger and larger. Induction may feel like some kind of trick, so we should see how it works with something that is relatively simple.

### Example

Prove that the sum of the first  $n$  numbers is given by the formula  $S_n = \frac{n}{2}(1 + n)$ .

In the section on Series, we saw that you can add the first 100 numbers up very quickly by creating 50 pairs:  $1 + 100$ ,  $2 + 99$ ,  $3 + 98$ , and so on. All of these pairs add up to 101, so the sum of the first 100 numbers is 50 times 101. In general, the sum of the first  $n$  numbers is  $\frac{n}{2}(1 + n)$ . You can try that out for several values of  $n$ , and find that it always works. Induction helps us to prove that  $S_n = \frac{n}{2}(1 + n)$  by showing that if it is true for  $n$  is equal to some number  $k$ , it must also be true for  $k + 1$ . Proof by induction usually starts by showing that a formula is true for  $n = 1$ :

$$S_1 = \frac{1}{2}(1 + 1) = 1$$

Next, assume that it is true for  $n = k$ :  $S_k = \frac{k}{2}(1 + k)$

Now we will show that if the formula is true for  $n = k$ , then it is also true for  $n = k + 1$ . If you put the number  $k + 1$  into the formula, you get

$$S_{k+1} = \frac{k+1}{2} (1 + (k + 1))$$

If the formula is correct, you should get the same result by simply adding the number  $(k + 1)$  to the sum of the first  $k$  numbers:

$$S_{k+1} = \frac{k}{2}(1 + k) + (k + 1)$$

So,  $\frac{k+1}{2} (1 + (k + 1))$  should be equal to  $\frac{k}{2}(1 + k) + (k + 1)$ . Let's rearrange that a bit to see if it really is the same:

$$\frac{k+1}{2} (1 + (k + 1)) = \frac{k}{2}(1 + k) + (k + 1)$$

$$\frac{k+1}{2} (1 + (k + 1)) = \frac{k}{2} + \frac{k^2}{2} + k + 1$$

$$\frac{k+1}{2} (1 + (k + 1)) = \frac{k^2}{2} + \frac{k}{2} + \frac{2k}{2} + \frac{2}{2}$$

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

$$\frac{k+1}{2} (1 + (k + 1)) = \frac{(k + 1)(k + 2)}{2}$$

$$\frac{k+1}{2} (1 + (k + 1)) = \frac{k + 1}{2} \cdot (k + 2)$$

$$\frac{k+1}{2} (1 + (k + 1)) = \frac{k + 1}{2} \cdot (k + 1 + 1)$$

This looks better if you only change one of the sides, but I actually got there by changing both sides until they looked equal. Then I went back and did it this way.

Although this may feel like you've been cheated somehow, it does work. Just try it out with an incorrect formula like  $S_n = \frac{n}{2}(n)$ . If that is true,  $S_{k+1}$  should be equal to  $\frac{k}{2}(k) + k + 1$ . By directly filling in the formula for  $k + 1$ , you would get  $S_{k+1} = \frac{k+1}{2} (k + 1)$ . Check these expressions to see that no amount of manipulation will make them equal to each other.

Now, to get back to De Moivre's Theorem. As we said, it is obviously true for  $n = 1$ . If it is actually true for  $n = k$ , then for  $n = k + 1$ :

$$(r(\cos x + i \sin x))^{k+1} = r^k(\cos(kx) + i \sin(kx)) \cdot r(\cos x + i \sin x)$$

And, by filling in the formula directly:

$$(r(\cos x + i \sin x))^{k+1} = r^{k+1}(\cos((k+1)x) + i \sin((k+1)x))$$

So, we need to show that

$$r^k(\cos(kx) + i \sin(kx)) \cdot r(\cos x + i \sin x) = r^{k+1}(\cos((k+1)x) + i \sin((k+1)x))$$

Just as we did for  $n = 3$ , we can multiply and rearrange the part on the left:

$$r^{k+1}(\cos(kx) + i \sin(kx))(\cos x + i \sin x)$$

$r^{k+1}$  on the left matches  $r^{k+1}$  on the right, so we won't worry about that part.

$$\cos(kx) \cos x + i \sin x \cos(kx) + i \sin(kx) \cos x - \sin(kx) \sin x$$

$$\cos(kx) \cos x - \sin(kx) \sin x + i (\sin(kx) \cos x + \sin x \cos(kx))$$

Again use the angle addition formulas to get:

$$\cos(kx + x) + i \sin(kx + x), \text{ which is equal to the part on the right, } (\cos((k+1)x) + i \sin((k+1)x)).$$

## Parametric Equations

Parametric equations can express both  $x$  and  $y$  as functions of a parameter. Often that parameter is  $t$ , for time.

The first time I had to graph parametric equations, I was confused. Where was  $t$  supposed to go on the graph?? The answer to that is that it doesn't have to go anywhere. Let's take a look at a set of parametric equations:

$$x = t$$

$$y = t^2$$

Or, if you are bothered by the simplicity of these equations, you can rewrite them in function notation:  $x(t) = t$  and  $y(t) = t^2$ . At  $t = 0$  both  $x$  and  $y$  are 0, so the first point on the graph is  $(0,0)$ . When  $t = 1$ ,  $x$  is 1 and  $y$  is 1, so we get the point  $(1,1)$ . Next we set  $t = 2$  to obtain the point  $(2,4)$ . The next point is  $(3,9)$ , etc. You can either just graph this, or you can put some dots on your curve and write the corresponding value of  $t$  beside each point. Some graphs also show arrows to indicate the direction in which the curve is traced out as time increases. By using substitution, you can convert the parametric equations into an equation in rectangular coordinates. If  $y = t^2$  and  $t = x$ , then  $y = x^2$ . You may want to take note of the fact that this conversion would seem a lot more complex if you use function notation.

### Example

Graph     $x(\theta) = \cos \theta$   
            $y(\theta) = \sin \theta$

This set of parametric functions traces out a circle, in a counterclockwise direction.  $\theta$  is the parameter. By converting this to a set of regular equations and using a clever substitution, you can show that it is the description of a circle with a radius of 1:  $x^2 + y^2 = \cos^2 \theta + \sin^2 \theta = 1$ . Still, you should probably actually graph this once to convince yourself that it is really true. I would recommend graphing  $x(\theta) = 10 \cos \theta$  and  $y(\theta) = 10 \sin \theta$  instead so you have a bit more room to work.

When you express a set of parametric equations as a relation between  $x$  and  $y$ , be careful about the sine and cosine. If you say that  $x = \sin t$ , that means that  $x$  can only have values between  $-1$  and  $1$ . So, we can convert

$$x = \sin t$$

$$y = \sin^2 t$$

to  $y = x^2$ , but both  $y$  and  $x$  can only have values between  $-1$  and  $1$ .

You will also be asked to convert regular functions to parametric equations. If you want to express  $y$  as a function of  $x$ , it is usually best to just set  $x$  equal to  $t$ , or whatever parameter you are asked to use.  $y = x^3$  turns into  $x = t$ ,  $y = t^3$ . Sometimes you want  $x$  to start at a certain value, say  $4$ . In that case, let  $x = t + 4$ , so that it can be  $4$  at the starting time  $t = 0$ .

### Example

A particle starts at the point  $(4, 2)$ , and moves in a straight line to the point  $(8, 7)$  at a constant speed. Write a set of parametric equations that describe this motion.

The slope of the line segment traced out by the particle is rise/run, or  $(7 - 2)/(8 - 4)$ , which equals  $5/4$ . Here you could use the expression  $y - y_1 = m(x - x_1)$ . Use the starting point,  $(4, 2)$  for  $(x_1, y_1)$ :  $y - 2 = \frac{5}{4}(x - 4)$ . So,  $y = \frac{5}{4}(x - 4) + 2$ . That is just the equation of a line, and it doesn't contain the information that we are starting at  $x = 4$ . We have to specify that  $4 \leq x \leq 8$ . Set  $x - 4 = t$ , so we can start  $t$  at zero, and write the parametric equations:

$$x = t + 4$$

$$y = \frac{5}{4}t + 2$$

Specify that  $0 \leq t \leq 4$ , so that  $x$  stays between  $4$  and  $8$ . If that doesn't seem clear to you, just start with  $4 \leq x \leq 8$  and substitute for  $x$ :  $4 \leq t + 4 \leq 8$ . Then subtract  $4$  from all three parts.

## Falling Objects

When Galileo used ramps to slow down the effects of gravity, he noticed that a ball that rolls down a ramp moves faster and faster. Measuring carefully, he found that if the ramp was four times as long, the ball would only take twice as long to reach the bottom. If the ramp would be 9 times as long, the time to reach the bottom would increase by a factor of 3. So, if an object falls for 3 seconds as opposed to 1 second, we would expect it to fall 9 times as far.

Because of this squared relationship between time and distance, motion influenced by gravity has the shape of a parabola. You can see this even for a falling object when you graph its position against time. Let's see how that works.

We know that  $\text{speed} = \frac{\text{distance}}{\text{time}}$ , so  $\text{distance} = \text{speed} \cdot \text{time}$ . Velocity is speed with direction, so in physics you will be using  $v$  to represent speed (and direction). Distance may be represented by  $x$  for horizontal motion, by  $y$  for vertical motion, or generically by the letter  $s$ . Distance traveled = velocity multiplied by the time.

**$s = vt$** , plus some initial distance  $s_i$  which may be 0 if the motion starts at a given zero point:

$$\mathbf{s = s_i + vt}$$

We also know that acceleration is velocity/time. Acceleration measures how much faster the speed gets per unit of time, so it is usually measured in meters per second/second, or  $\text{m/sec}^2$ .

If  $a = \frac{v}{t}$ , then  $v = at$ .

Because the acceleration that we will consider in our physics problems will be constant, such as the acceleration produced by gravity, we can simply take the average speed when doing our position calculations. The speed at the start of an interval is  $v_i$ , and at the end of the interval it will be  $v_i + at$ . The final speed will often be denoted by  $v_f$ .

$$\mathbf{v_f = v_i + at}$$

As a result, the average speed over the entire interval is  $\frac{v_i + v_f}{2}$ , which is  $\frac{v_i + (v_i + at)}{2}$ , or  $v_i + \frac{1}{2}at$ .

Plug that in for  $v$  in the position equation,  $s = s_i + vt$ :

$$s = s_i + (v)t$$

$$s = s_i + (v_i + \frac{1}{2}at)t$$

$$\mathbf{s = s_i + v_it + \frac{1}{2}at^2}$$

Already you can see that this is a quadratic equation. If you graph the position  $s$  against the time  $t$ , you'll get a parabola as you would for any quadratic equation. This equation will work for any motion that involves a constant acceleration produced by some force (Force = mass · acceleration).

The acceleration due to gravity at the Earth's surface is about  $9.8 \text{ m/s}^2$ , or approximately 32 feet/sec<sup>2</sup>. We use the letter  $g$  to indicate acceleration due to gravity, and  $g$  is the positive quantity  $9.8 \text{ m/s}^2$ . Since this acceleration is directed downward people commonly use  $-g$ . Here the motion is vertical, so we will use  $y$  for the distance:

$$y = y_i + v_i t - \frac{1}{2} g t^2$$

If you drop a ball from height  $h$ , how long will it take to hit the ground? Well,  $h$  is the initial position, and the initial speed is zero. When the ball hits the ground, its position  $s$  will be 0:

$$0 = h + 0t - \frac{1}{2} g t^2$$

$$\frac{1}{2} g t^2 = h$$

$$t^2 = \frac{2h}{g} \text{ so } t = \sqrt{\frac{2h}{g}}$$

It should not be necessary to memorize this, but you can use it to check your calculations when you are looking at specific examples.

What is the impact velocity of the ball?

The velocity after time  $t$  is given by  $v = v_i + at$ , and  $v_i = 0$ , while the acceleration  $a$  is equal to  $-g$ .

$$t \text{ is } \sqrt{\frac{2h}{g}}:$$

$$v = -g \frac{\sqrt{2h}}{\sqrt{g}} = -\sqrt{g} \sqrt{2h} = -\sqrt{2gh}$$

If an object is launched vertically with a speed  $v_i$ , how long will it take before it hits the ground?

$y = v_i t - \frac{1}{2} g t^2 + y_i$ . We will assume that we start on the ground, so  $y_i = 0$ .

$y = v_i t - \frac{1}{2} g t^2$ . When the object hits the ground  $y = 0$

$$0 = v_i t - \frac{1}{2} g t^2$$

$$0 = (v_i - \frac{1}{2} g t) t$$

$$t = 0 \text{ or } v_i - \frac{1}{2} g t = 0. \quad v_i = \frac{1}{2} g t, \text{ so } t = \frac{2v_i}{g}$$


---



Interestingly, if you are standing at the edge of a cliff and throw a ball straight up with a velocity of, say 10 m/s, its impact velocity at the bottom of the cliff will be exactly the same as if you had thrown it straight down with a velocity of 10 m/s. This does make sense, because throwing the ball up at 10 m/sec will cause it to be falling back down with a speed of 10 m/s when it is level with the edge of the cliff. Although the impact velocity is the same, the time to impact is not.

The basic equations of motion can be rearranged to get a useful relationship between displacement, acceleration and velocity that does not involve time.

$$v_f = v_i + at \rightarrow at = v_f - v_0 \rightarrow t = \frac{v_f - v_i}{a}$$

You could substitute that into  $S = S_i + v_i t + \frac{1}{2} at^2$ , but that is a bit time-consuming because of the square. Remember that  $v_{ave} = \frac{v_f + v_i}{2}$ . We can use the average speed because the acceleration will be constant. We also know that  $S = S_i + v_{ave}t$ . Substitute for  $v_{ave}$  and  $t$  to get:

$$S = S_i + \frac{v_f + v_i}{2} \cdot \frac{v_f - v_i}{a}$$

$$S = S_i + \frac{v_f^2 - v_i^2}{2a}$$

This formula allows you to calculate the final position of an object if you know the initial position, the initial and final velocity, and the acceleration. If you think about that, it does make sense that if you know those particular things you should be able to predict where the object will end up. Notice that it makes no difference if  $v_i$  is positive or negative. If the object is going to fall, the acceleration will be negative and everything will work out correctly.

## Projectiles

Force and velocity vectors that are at an angle from horizontal can be decomposed into horizontal and vertical components. If an object is launched at angle  $\theta$  with velocity  $v$ , the vertical component is  $|v|\sin \theta$ , and horizontal component of the velocity is  $|v|\cos \theta$ . Here  $|v|$  represents the magnitude of the velocity vector.

Each component of the velocity can be considered separately. The vertical component is a parabola when graphed against time, and the steady horizontal velocity then creates a parabolic trajectory for the object.

The upward component of the velocity determines how long the object is in the air, and the horizontal component determines how far it travels horizontally during this time.

First consider the vertical component of the motion.  $v_y$ , the *initial* vertical speed, is  $|v|\sin\theta$ . Because of the influence of gravity, the vertical speed will change constantly.

$y$  is 0 at the beginning and the end of the parabola.  $0 = v_y t - \frac{1}{2}gt^2$ .

$0 = (v_y - \frac{1}{2}gt)t$  so  $t = 0$  or  $v_y - \frac{1}{2}gt = 0$ . Solve that last equation for  $t$  to get  $t = \frac{2v_y}{g}$ . The object is in the air for a total time of  $\frac{2v_y}{g}$ .

Maximum height occurs at  $t = \frac{v_y}{g}$  (parabolas are symmetrical). To find the actual height at this point insert this value for  $t$  into the position equation:  $s = -\frac{1}{2}g\left(\frac{v_y}{g}\right)^2 + \frac{v_y^2}{g} = \frac{v_y^2}{2g}$ .

Now look at the horizontal component. The horizontal speed will stay constant (for practical purposes we will ignore air friction).

To find how far the object travels in a horizontal direction, multiply the horizontal velocity by the total time the object is in the air:

$x = x_i + v_h t$  where  $v_h$  is the horizontal component of the velocity.

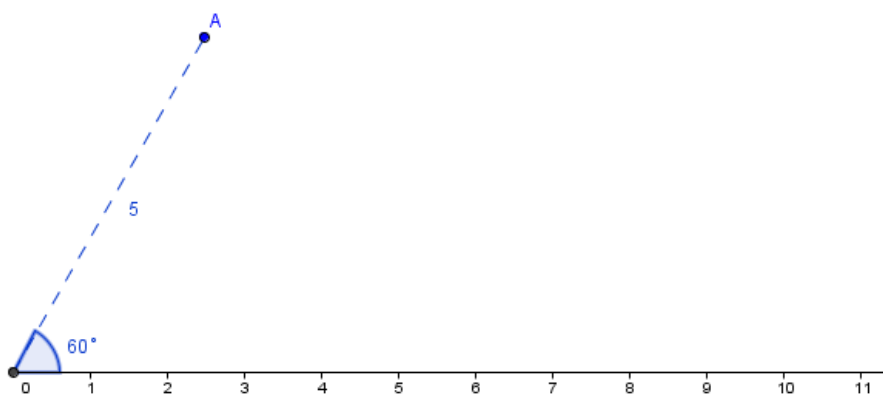
If you need the actual magnitude and direction of the impact velocity, consider the vertical and horizontal components of the velocity as perpendicular vectors. You can use the Pythagorean Theorem to find the magnitude, and the inverse tangent to find the angle of the impact velocity.

## Polar Coordinates

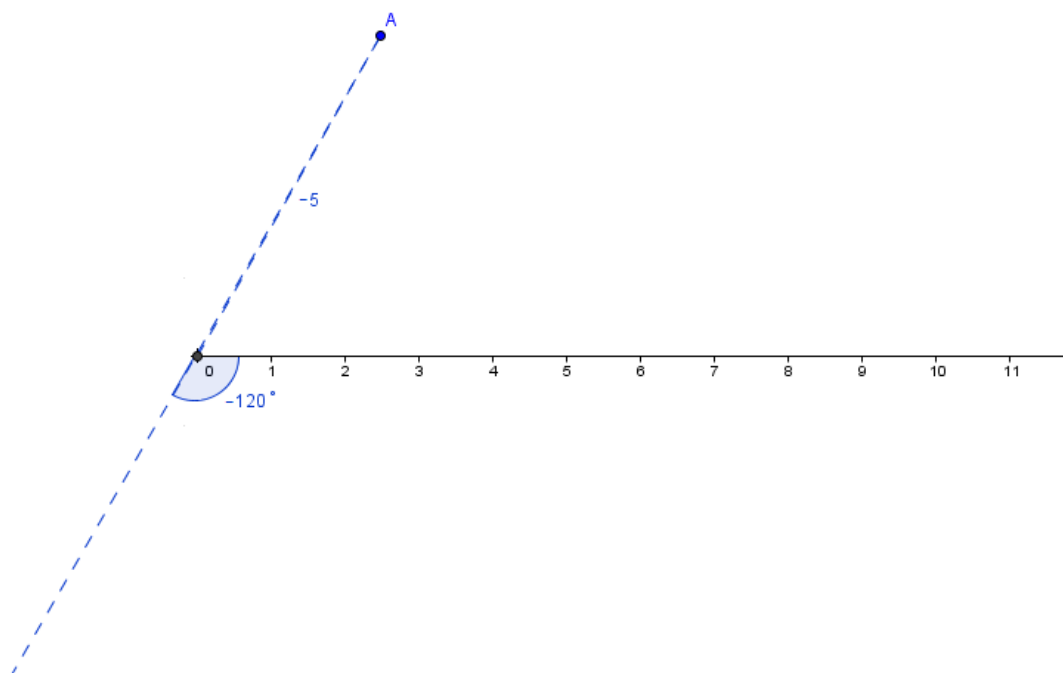
Polar coordinates were already used in the time of the ancient Greeks, which means that this system is much older than the Cartesian coordinates you are familiar with. Our regular x-y system sees a flat surface as an infinite square, and Cartesian coordinates are also referred to as rectangular coordinates. However, you could also see a flat surface as an infinitely large circle. This might be what you would do if you were trying to specify the positions of stars in the sky. To use polar coordinates, all you need is a central point, kind of like the origin, and a polar axis from which to measure an angle.

The free graphing software [MathGV](#) lets you experiment with polar graphs so you can learn to understand them better. Go to File, and select New Polar Graph. A window appears with many different axes. If you want your graphs to look like the ones in your textbook, go to Graph → Graph Settings. Highlight all of the settings on the axis list except for 0 degrees and delete them.

To actually specify the location of a point, you need to provide the distance  $r$  of the point from the origin, and the angle between the polar axis and the line along which the point lies. In the image below, you can see the point A located at a distance of 5 units from the origin as measured along a line that makes a 60 degree angle with the polar axis.



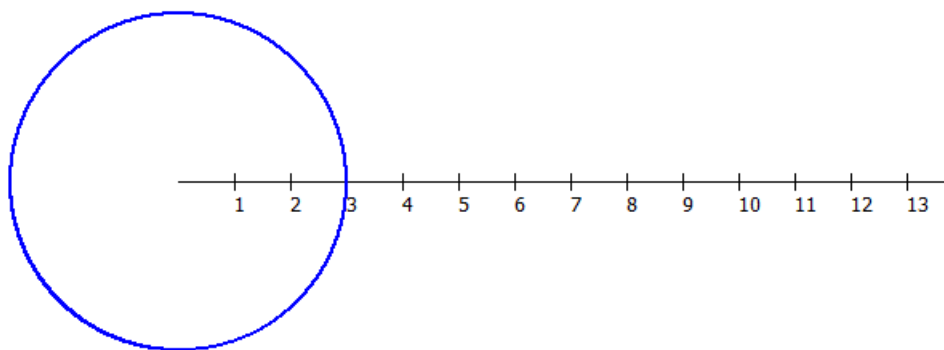
The angle is always called  $\theta$ . We could also describe the point A as being located at a distance of -5 units from the origin, along a line that makes a -120 degree angle with the polar axis:



In this way, you can describe any point on a plane. Just as we place  $x$  before  $y$  in rectangular coordinates, we put  $r$  before  $\theta$  in polar:  $(r, \theta)$ . Oddly enough however, polar graphs are usually in an  $r = \dots$  format so that  $\theta$  is the input of the function.

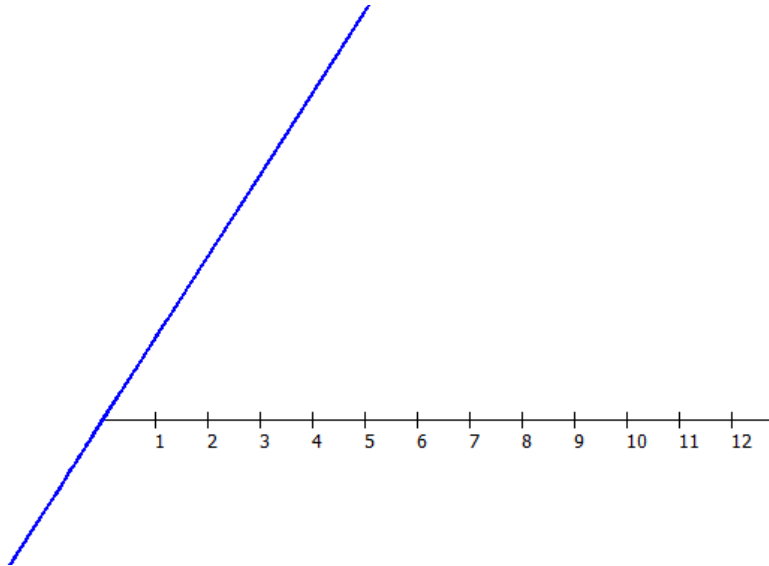
Let's look at some basic polar graphs:

**1.  $r = 3$ .** The radius is always 3, but you can use any value for  $\theta$ :



Any equation of the form  $r = c$  creates a circle with its center at the origin.

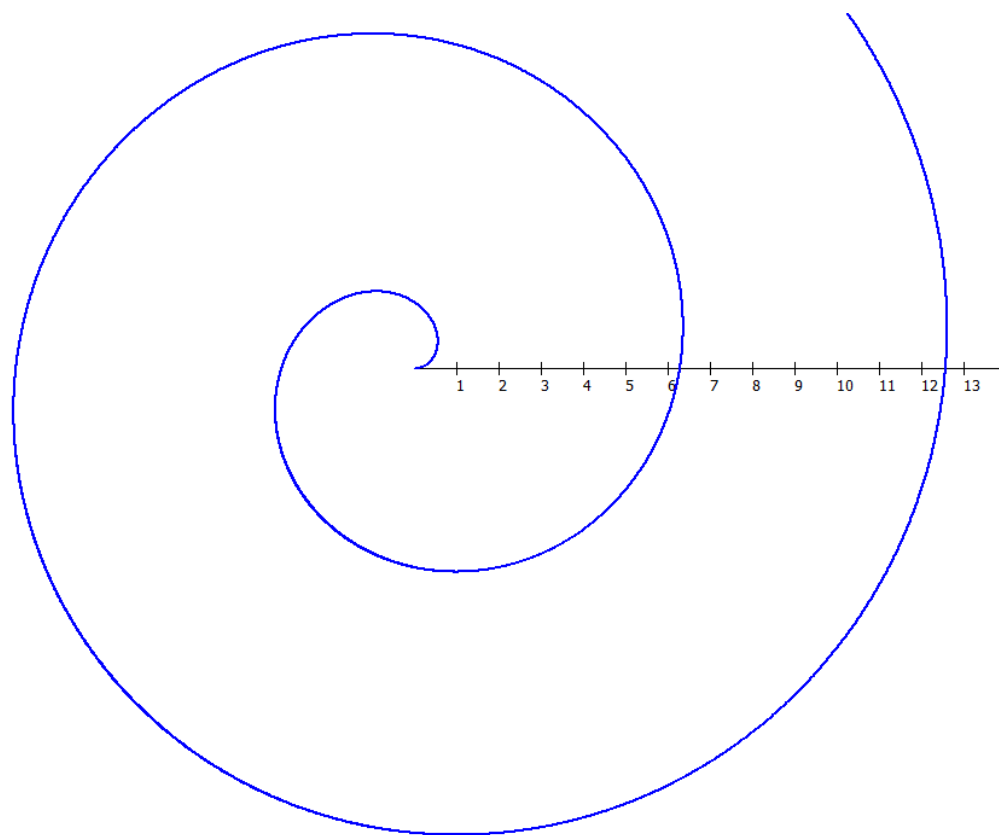
2.  $\theta = 1$ . The angle is constant at 1 radian, while the radius varies:



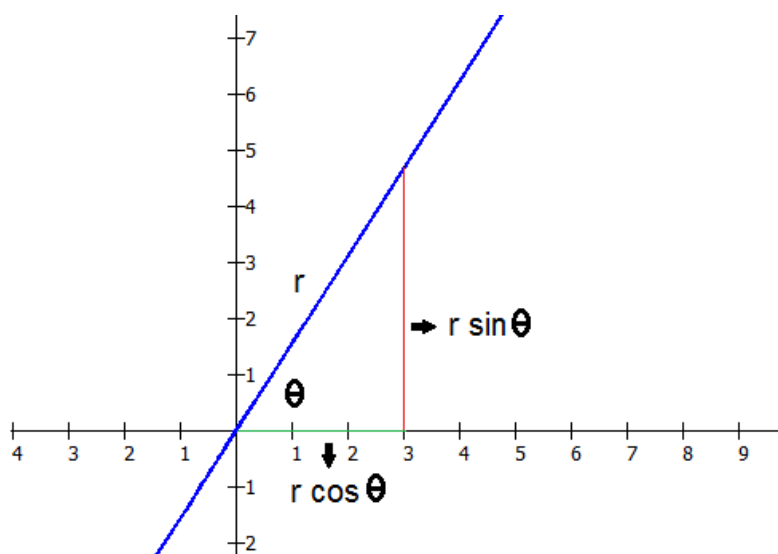
When you specify only that  $\theta = 1$ ,  $r$  can be any value. That is similar to  $x = 1$  in rectangular coordinates, which creates a line because  $y$  can have any value.

The line extends below the polar axis due to negative values of  $r$ , not because the angle changes.  $\theta = c$  specifies a line that always passes through the polar origin. Unfortunately though you can't just move that line up or down by adding something as you would in a regular coordinate system.

3.  $r = \theta$ . The radius is always equal to the size of the angle in radians, which results in a spiral:



Converting from polar to rectangular coordinates is not that difficult. If you lay a regular coordinate system on top of the polar one, you can see that  $y = r \sin \theta$  and  $x = r \cos \theta$ :



So, the point  $(5, \frac{\pi}{3})$  can be translated by setting  $x = 5 \cos \frac{\pi}{3}$  and  $y = 5 \sin \frac{\pi}{3}$ . That gives you  $(2.5, 2.5\sqrt{3})$  in Cartesian coordinates.

Going the other way, from rectangular into polar, is just slightly more trouble.

### Example

Change the point  $(-3, -4)$  from rectangular into polar coordinates.

First, we have to calculate the value of  $r$ . You can use the distance formula, which is really just the Pythagorean Theorem, to determine that  $r = \sqrt{3^2 + 4^2} = 5$ . Next, we need the angle. Here is where you need to be careful. In the picture above, you see that  $\tan \theta$  is just  $y/x$ . Just make sure that you end up in the right quadrant with the correct angle in reference to the polar axis. Here  $y/x$  is  $-4/-3$ , or  $\frac{4}{3}$ . The inverse tangent of  $\frac{4}{3}$  is  $53.13^\circ$ , but because the point you want is in the third quadrant you need to measure out a negative distance along the line that makes a  $53.13$  degree angle with the polar axis. The coordinates you want are  $(-5, 53.13)$ , which can of course be expressed in many different ways, such as  $(5, -126.87)$  or  $(5, 233.13^\circ)$ . Fortunately, there are conversion calculators available online so you can check your work.

We already saw that  $x = r \cos \theta$  and  $y = r \sin \theta$ . By looking at the image above you can also see that  $r^2 = x^2 + y^2$ . This provides all the information you need to change polar equations into rectangular equations.

### Example

Change the polar equation  $r = 6$  to an equation in rectangular coordinates.

If  $r = 6$ , then  $6^2 = x^2 + y^2$ , so  $x^2 + y^2 = 36$ . This equation describes a circle with a radius of 6.

### Example

Change the polar equation  $\theta = \frac{\pi}{4}$  into an equation in rectangular coordinates.

For this equation, just insert the values for  $x$  and  $y$ :  $x = r \cos \frac{\pi}{4}$ , and  $y = r \sin \frac{\pi}{4}$ . That translates to  $x = r \frac{\sqrt{2}}{2}$  and  $y = r \frac{\sqrt{2}}{2}$ . A regular equation would look like  $y = \text{something with } x$  instead of  $y = r \frac{\sqrt{2}}{2}$ . The obvious thing to do is to replace  $r$  with something containing  $x$ . We can get that something by rearranging  $x = r \frac{\sqrt{2}}{2}$  to  $r = \frac{x}{\frac{\sqrt{2}}{2}}$ . This means that  $y = \frac{\sqrt{2}}{2} \cdot \frac{x}{\frac{\sqrt{2}}{2}}$  or  $y = x$ . Alternatively, you can say that  $\frac{y}{x} = \frac{r \frac{\sqrt{2}}{2}}{r \frac{\sqrt{2}}{2}} = 1$ . If  $\frac{y}{x} = 1$  then  $y = x$ . The rectangular equation is  $y = x$ , which is a line with a slope of 1. The inverse tangent of 1 is  $\frac{\pi}{4}$  radians. The equation is correct because the slope of the line corresponds to  $\theta = \frac{\pi}{4}$ .

### Example

Change the polar equation  $r = 5 \sec \theta$  into an equation in rectangular coordinates.

You may find this easier if you rewrite the equation as  $r = \frac{5}{\cos \theta}$ . It may take a bit of thought, but eventually you would then think of multiplying both sides by  $\cos \theta$  to take advantage of the equation  $x = r \cos \theta$ . If  $r \cos \theta = 5$ , then  $x = 5$ . Use MathGV to graph the polar equation.

### Example

Change the polar equation  $r = 4 \cos \theta$  into an equation in rectangular coordinates.

Here it is advantageous to multiply both sides by  $r$ :  $r^2 = 4r \cos \theta$ . That translates to  $x^2 + y^2 = 4x$ . Although this is a valid equation, your teacher would probably prefer that you change it to a proper conic section format by completing the square:

$$x^2 + y^2 - 4x = 0$$

$$x^2 - 4x + y^2 = 0$$

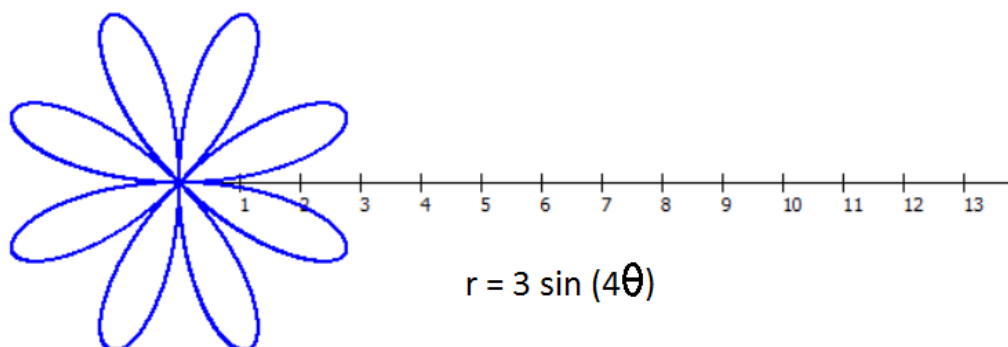
$$x^2 - 4x + 4 + y^2 = 4$$

$$(x - 2)^2 + y^2 = 4$$

This is a circle with center  $(2, 0)$  and a radius of 2. Use MathGV to graph the polar equation.

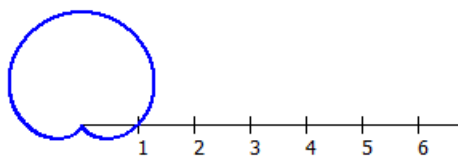


Equations of the form  $r = a \sin \theta$  and  $r = a \cos \theta$  produce circles, but when we graph  $r = a \sin (n\theta)$  and  $r = a \cos (n\theta)$ , where  $n$  is an integer, we get flower-like shapes called roses:



Hmmm, that looks a bit more like a daisy to me. Obviously no botanists were consulted when these shapes were named. You can experiment in MathGV to see how to produce different kinds of roses.

When we graph  $r = \sin \theta$ , we get a circle that sits just above the polar origin. This happens because  $\sin \theta$  ranges from 1 to -1 in a nice orderly way. In a sense we are reversing the process that created the sine wave from the circle. Now look at  $r = 1 + \sin \theta$ . Here  $r$  is never negative. The smallest value is 0, which pulls the graph to the polar origin at the bottom:



In general,  $r = a + b \sin \theta$  and  $r = a + b \cos \theta$  generate lima-bean shaped curves called **limaçons** (pronounced as “lemahsons”). This name is actually derived from the Latin word for snail or slug, which is limax.

If  $a$  is larger than  $b$ , the zero point is not reached so the curve shows a dimple that doesn’t reach the origin, or it just flattens a bit if the difference is large. If  $b$  is larger than  $a$  there is a range of negative values which results in a small second loop that appears inside the first. The limaçons with a sharp pinch ( $a = b$  as shown above) or an inside loop ( $b > a$ ) are called **cardioids** because they resemble the shape of a heart. Caustics are patterns formed when light is reflected by a curved surface. When the surface is a perfect circle we may be able to see an actual mathematically correct cardioid. For interesting images, search for “caustic optics”.

## Polar Graph Shapes

$\theta = a$       A line that makes an angle of  $a$  radians with the polar axis

$r = a$       A circle with radius  $a$  centered at the polar origin

$r = \sin \theta$       A circle with radius  $1/2$  and center at  $(1/2, \pi/2)$

$r = a \sin \theta$       A circle with radius  $a/2$ , above the origin if  $a > 0$ , and below if  $a < 0$

$r = \cos \theta$       A circle with radius  $1/2$  and center at  $(1/2, 0)$

$r = a \cos \theta$       A circle with radius  $a/2$ , right of the origin if  $a > 0$  and left if  $a < 0$

$r = a + b \sin \theta$  and  $r = a + b \cos \theta$  generate lima-bean shaped curves called limaçons. Limaçons with a sharp pinch resemble the shape of a heart and are called cardioids.

.

$r = a + b \sin \theta$ , where  $b > a$       A cardioid above the origin

$r = a - b \sin \theta$ , where  $b > a$       A cardioid below the origin

$\theta = \text{some constant}$  specifies a line that always passes through the polar origin. Unfortunately though you can't just move that line up or down by adding something as you would in a regular coordinate system.

Horizontal and vertical lines can be easily translated from regular equations.  $x = 5$  turns into  $r \cos \theta = 5$ , so that  $r = \frac{5}{\cos \theta}$ .  $y = 5$  turns into  $r = \frac{5}{\sin \theta}$ .

To create any line in polar coordinates we can translate from regular Cartesian coordinates. The equation of a line,  $y = mx + b$ , translates as  $r \sin \theta = m r \cos \theta + b$ . Rearranging the equation to find  $r$ , we get  $r \sin \theta - m r \cos \theta = b$ , so  $r(\sin \theta - m \cos \theta) = b$  and  $r = \frac{b}{\sin(\theta) - m \cos(\theta)}$ , where  $m$  is the slope.

Using this equation, you can draw any line. It is not necessary for that line to go through the polar origin. The polar axis intercept of a line occurs when  $\theta = 0$ . When that happens,  $r = -b/m$ , since  $\sin \theta$  will be 0 and  $\cos \theta$  will be 1.  $-b/m$  is the polar axis intercept.

## Polar Equations of Conics

$e$  is the eccentricity.  $e = 1$  for a parabola.  $e < 1$  for an ellipse and  $e > 1$  for a hyperbola.

$d$  is the distance from the origin to the directrix.

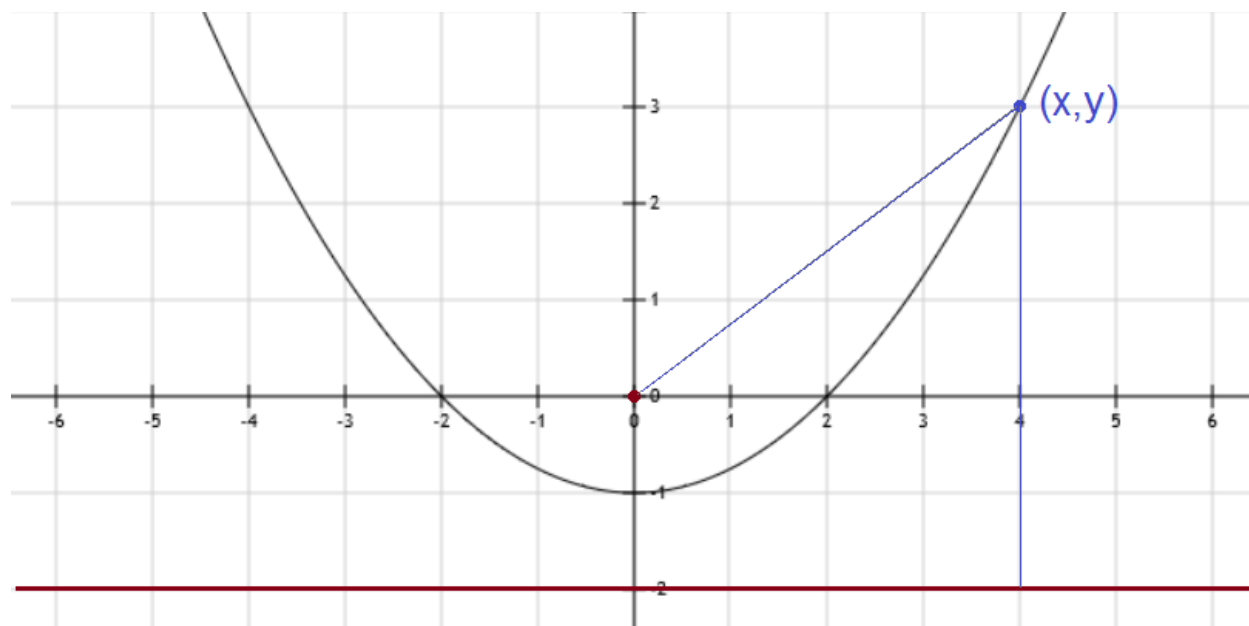
$$r = \frac{de}{1 + e \sin \theta} \quad \text{directrix at } y = d$$

$$r = \frac{de}{1 - e \sin \theta} \quad \text{directrix at } y = -d$$

$$r = \frac{de}{1 + e \cos \theta} \quad \text{directrix at } x = d$$

$$r = \frac{de}{1 - e \cos \theta} \quad \text{directrix at } x = -d$$

A parabola has a directrix such that for any point  $(x, y)$  on the parabola the distance to the focus is equal to the distance to the directrix. By placing the focus of the parabola at the origin, we can translate its equation to polar coordinates. The distance from  $(x, y)$  to the origin is now  $\sqrt{x^2 + y^2}$ . The directrix has moved over along with the focus. We will say that it is at a distance  $d$  from the origin. The distance between  $(x, y)$  and the directrix is  $d + y$ :



$$\sqrt{x^2 + y^2} = d + y$$

$$r = d + r \sin \theta$$

$$r - r \sin \theta = d$$

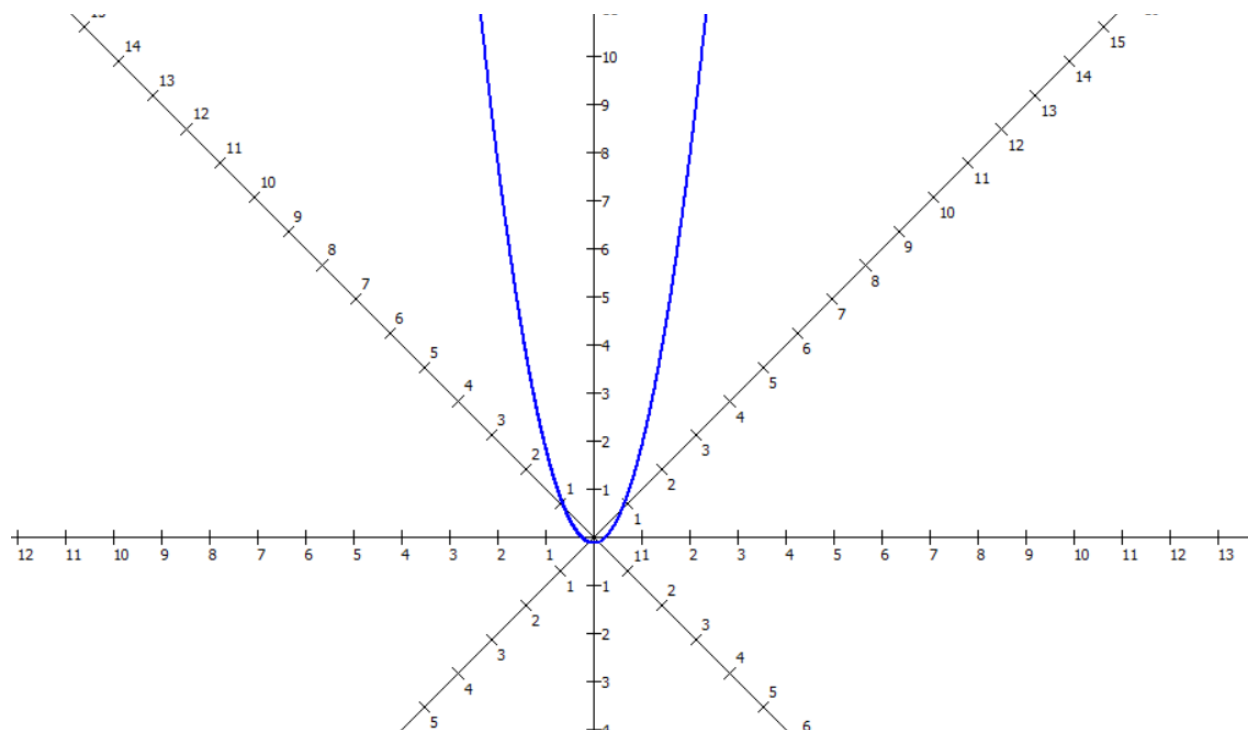
$$r(1 - \sin \theta) = d$$

$$r = \frac{d}{1 - \sin \theta}$$

Here  $d$  is equal to  $2p$ , the distance between the focus and the directrix. Recall that the equation  $y = ax^2$  is equivalent to  $y = \frac{1}{4p}x^2$ . So, if we want to represent the parabola  $y = 2x^2$  in

polar coordinates, we would set  $\frac{1}{4p} = 2$ , so that  $8p = 1$  and  $p = 1/8$ . And  $d$  is  $2p$ , so

$r = \frac{1/4}{1 - \sin \theta}$ . The polar graph, drawn in [MathGV](#), is shown below. (You can also use an online graphing calculator with polar capabilities, like [FooPlot](#).) Notice that the focus of the parabola is at the origin rather than the vertex, so the graph extends slightly below the polar axis.



If the parabola opens down instead of up, the directrix is located above the parabola, and the distance between  $(x, y)$  and the directrix is  $d - y$ . You may have to draw yourself a picture to see how that works.

$$\sqrt{x^2 + y^2} = d - y$$

$$r = d - r \sin \theta$$

$$r + r \sin \theta = d$$

$$r(1 + \sin \theta) = d$$

$$r = \frac{d}{1 + \sin \theta}$$

For side-opening parabolas, the distance from the generic point  $(x, y)$  to the directrix is  $d + x$  or  $d - x$ . Since  $x = r \cos \theta$ , the polar equation is  $r = \frac{d}{1 \pm \cos \theta}$ .

Parabolas are created by placing each point at an equal distance from the focus and the directrix. That makes the ratio  $\frac{\text{distance to focus}}{\text{distance to directrix}}$  equal to 1. This ratio is the **eccentricity** of a conic section, and we can make it less than 1, or greater than 1. When it is less than 1, the graph eventually has to curve back in on itself, creating an ellipse. When the eccentricity is greater than 1, the curve is pulled closer to the directrix which creates both branches of a hyperbola. Note that the ellipse and hyperbola have two potential directrices, and either one can be used to create the same graph.

Ellipses or hyperbolas can be created using the polar equations

$$r = \frac{de}{1 \pm e \sin \theta} \text{ or } r = \frac{de}{1 \pm e \cos \theta}$$

where  $e$  is the eccentricity and  $d$  is the distance from the polar origin to the closest directrix.

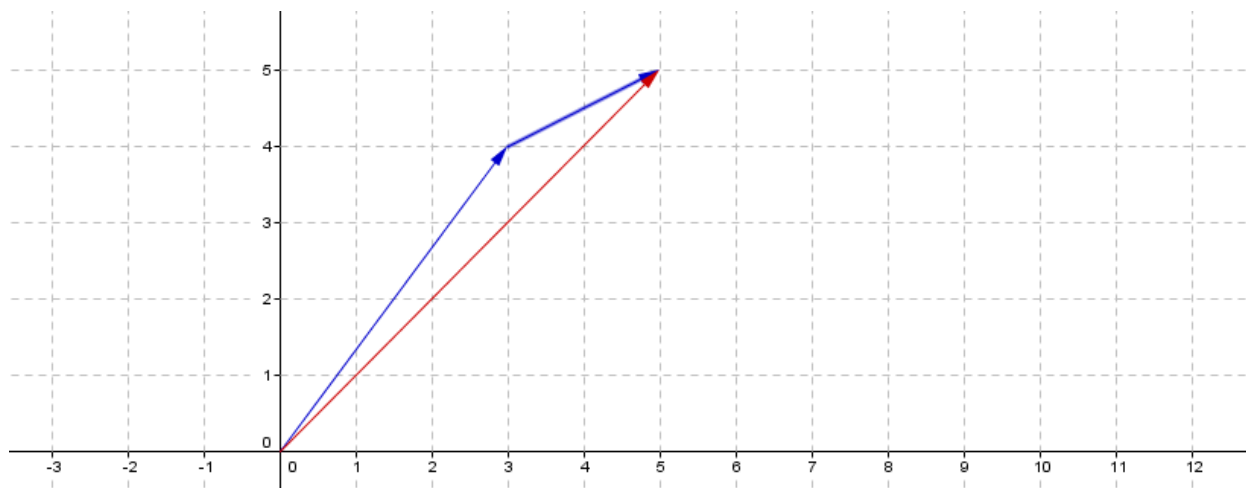
## Vectors

Vectors are arrows with a given length and direction. That is all they are, so there is no need to feel intimidated. Vectors are very useful in physics, where they are used to represent forces, velocity, and other things that have both a magnitude (size) and a direction. The length of the vector can represent the magnitude, and you just make the arrow point in the direction you need. In mathematics, vectors are normally used along with a coordinate system, and it is most convenient to position them so they start at the origin. That makes them easier to describe, because you only have to specify the position of the tip. For example, a vector that starts at the origin and ends at (3, 4) can be easily specified as  $\langle 3, 4 \rangle$ . The pointy brackets indicate that we are referring to a vector rather than just a point. The length of this vector can now be calculated easily from the Pythagorean Theorem (the Distance Formula in this case). The length, or magnitude, of a vector  $\vec{v}$  is indicated as  $|\vec{v}|$ , or sometimes as  $\|\vec{v}\|$ .  $|\vec{v}| = \sqrt{3^2 + 4^2} = \sqrt{25} = 5$ . Vectors are still arrows when placed in a coordinate system, but you could also think of them as a set of directions. The vector  $\langle 3, 4 \rangle$  would indicate your net movement if you started at the origin and moved 3 units to the right, and 4 units up. 3 is referred to as the x-component of the vector, and 4 is the y-component.

Even though vectors are normally positioned to start at the origin, they are still just arrows with a particular length and direction. Don't hesitate to pick them up and place them somewhere else. If they don't start at the origin, you can still describe them like this: vector  $\overrightarrow{PQ}$ , where P is (2,1) and Q is (5, 5). Vector  $\overrightarrow{PQ}$  has the same length and direction as  $\langle 3, 4 \rangle$ , because to get from point P to point Q you would move 3 units to the right and 4 units up.

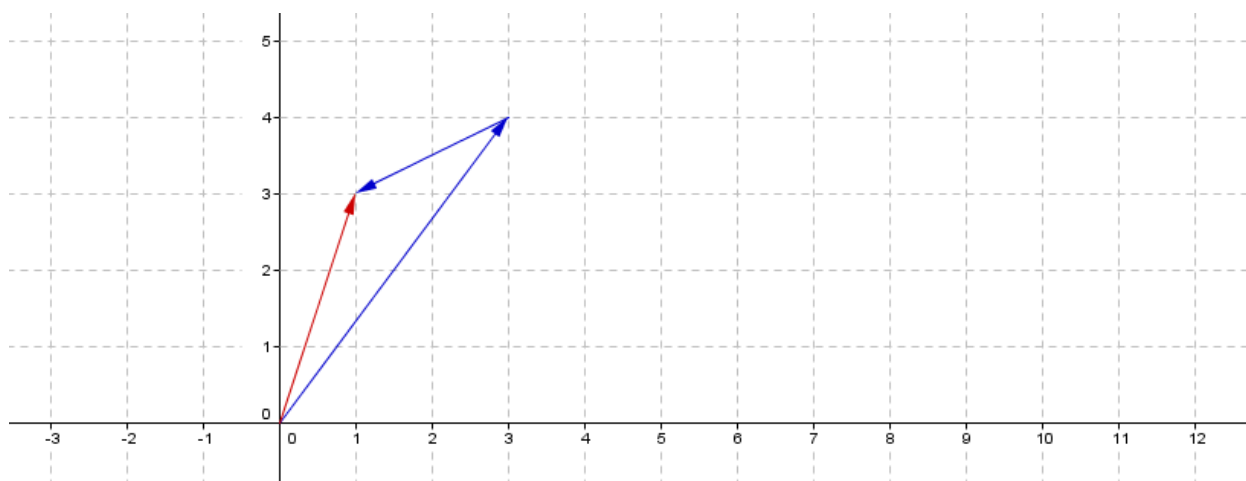
## Adding or Subtracting Vectors

To add vectors, just think of them as directions. Suppose you were adding the vectors  $\langle 3, 4 \rangle$  and  $\langle 2, 1 \rangle$ . Your directions would say: First move right 3 units, then up 4 units. Next, move right 2 units, then up 1 unit. If you were following along with this in a coordinate system, you would see that you have now moved to the right a total of 5 units, and up a total of 5 units.  $\langle 3, 4 \rangle + \langle 2, 1 \rangle$  is  $\langle 5, 5 \rangle$ . All you have to do is add the two x-components, and the two y-components to create the **resultant vector**. This addition is really the same thing as picking up one of the vectors and placing it to start where the other one ends. The resultant vector is drawn as a new vector from the start of the first vector to the end of the second:



Amazingly, simple vector addition works out really well in physics. You can use vectors to create a diagram of the forces acting on an object. In physics, just like in real life, objects react as if forces are applied directly to their center of mass. You can draw the forces as vectors with their origin at that center of mass. If you then superimpose a coordinate system with its origin at the center, you can add up the x and y components of all of the vectors. The resultant force that you calculate this way will be exactly equal to the force that is measured by the speed and direction of the movement of the object.

To subtract a vector from another vector, just subtract its x and y components. To create a picture of vector subtraction, consider that subtraction is really an addition of a negative number.  $10 - 7 = 10 + -7$ . To subtract a vector, just make it negative and then use addition. For example,  $\langle 3, 4 \rangle - \langle 2, 1 \rangle = \langle 3, 4 \rangle + \langle -2, -1 \rangle$ . The vector  $\langle -2, -1 \rangle$  points in the exact opposite direction from  $\langle 2, 1 \rangle$ . Addition of  $\langle 3, 4 \rangle + \langle -2, -1 \rangle$  produces the vector  $\langle 1, 3 \rangle$ :





## Multiplying a Vector by a Scalar

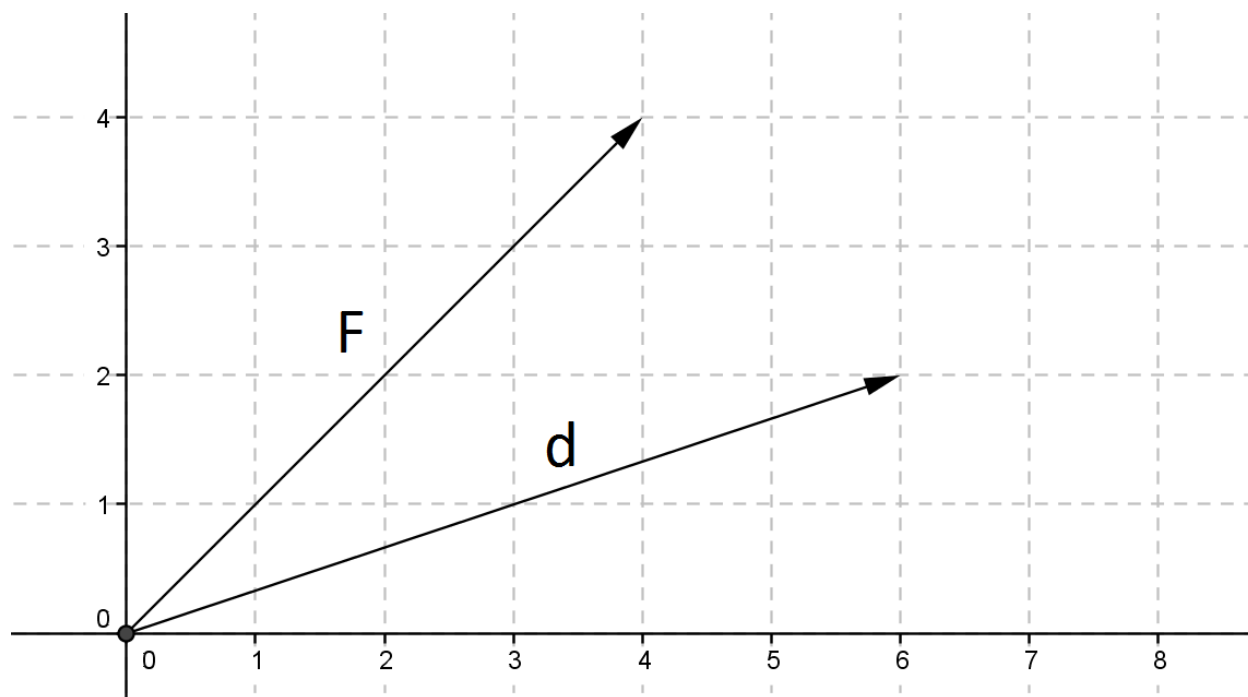
If you draw a vector that is 5 cm long to represent a force of 50 Newtons, you would probably want to draw a vector with a length of 10 cm to represent a force of 100 Newtons. If the new vector is to have the same direction as the original, you can easily create a vector that is twice the length by simply multiplying the x- and y-components by 2. Thanks to the magic of similar triangles, which you may vaguely remember from geometry,  $\langle 6, 8 \rangle$  is twice as long as  $\langle 3, 4 \rangle$ . When you are working with both vectors and numbers, the numbers are called a scalars. A scalar has a magnitude, but no direction associated with it. You can multiply any vector by any scalar. To shrink your vector, use a scalar smaller than 1. If you multiply by a negative scalar, the new vector will point in the opposite direction from the original (a 180 degree difference).

## The Dot Product: Vectors and Work

The dot product of two vectors  $\vec{A} \langle A_x, A_y \rangle$  and  $\vec{B} \langle B_x, B_y \rangle$  is  $A_x B_x + A_y B_y$ . Note that the dot product is not a vector; it is just a number (a scalar). This means that the dot product has no direction; it just has a magnitude.

The definition of the dot product may seem arbitrary, but it has a real physical meaning. In physics, **work** is defined as **Force** multiplied by **displacement**. One Joule (one Newton-meter) of work is done when a Force of one Newton moves an object a distance of 1 meter. Often the object in question ends up moving in a direction that is not parallel with the force. For example, try to push an eraser across the surface of your desk. Chances are that you pushed at an angle, and the eraser was hard to move because some of your applied force pressed the eraser down into the desk. To get the maximum “work” output, you should keep your finger parallel to the desk so that all of the force moves the eraser forward.

The actual work done when the force and the displacement do not line up perfectly is  $|\vec{F}| \cos \theta$  multiplied by the displacement,  $|\vec{d}|$ . Here  $\theta$  is the angle between the force vector and the displacement vector, with both vectors drawn so that they start at the center of the mass that is being moved.  $|\vec{F}| |\vec{d}| \cos \theta$  is the dot product of  $\vec{F}$  and  $\vec{d}$ . If  $\alpha$  is the angle of  $\vec{F}$  with a horizontal axis, and  $\beta$  is the angle of  $\vec{d}$  with that same axis, then  $\cos \theta = \cos (\alpha - \beta)$ . Trigonometry tells us that  $\cos (\alpha - \beta) = \cos \alpha \cos \beta + \sin \alpha \sin \beta$ .



$\vec{F}$  has x-component  $F_x$  and y-component  $F_y$ .  $\vec{F} = (F_x, F_y)$ .  $\vec{d}$  has x-component  $d_x$  and y-component  $d_y$ .  $\vec{d} = (d_x, d_y)$ .

$$\text{Therefore, } \cos \theta = \frac{F_x}{|\vec{F}|} \cdot \frac{d_x}{|\vec{d}|} + \frac{F_y}{|\vec{F}|} \cdot \frac{d_y}{|\vec{d}|} = \frac{(F_x \cdot d_x + F_y \cdot d_y)}{|\vec{F}| |\vec{d}|}$$

$|\vec{F}| |\vec{d}| \cos \theta$ , the dot product of  $\vec{F}$  and  $\vec{d}$ , is  $F_x d_x + F_y d_y$ .

## Limits and Continuity

Knowing how to factor is important in this section. Remember that  $a^2 - b^2 = (a + b)(a - b)$ .

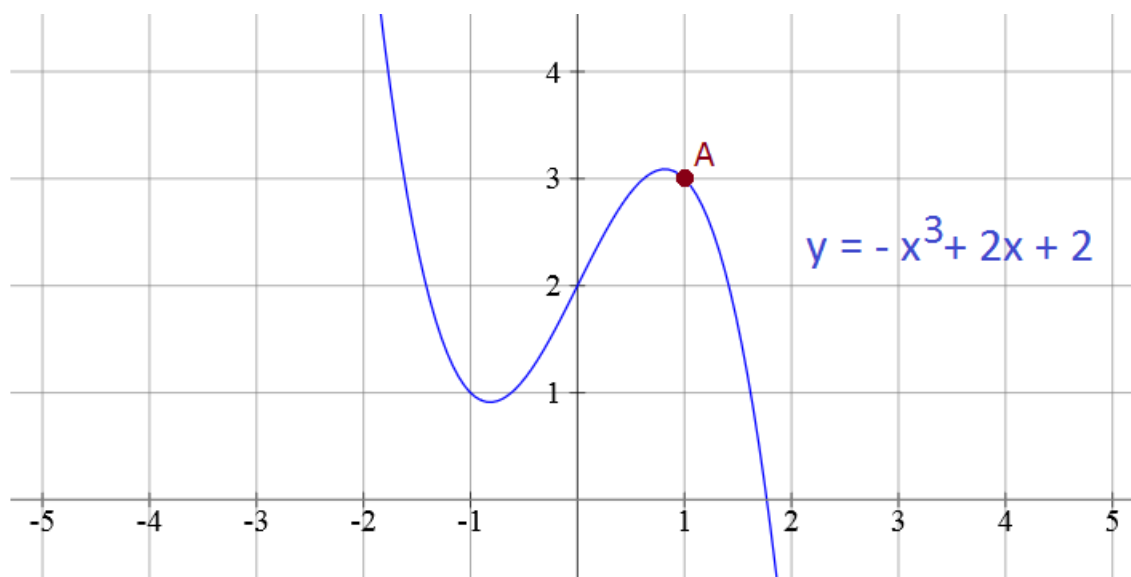
Continuous functions have no holes or breaks in their graphs.

An infinitely small hole (removable discontinuity) appears when a rational function has a factor in the numerator that cancels out a factor in the denominator. The limit exists at the hole.

In order for a limit to exist, the left-hand limit must equal the right-hand limit.

Limit =  $\infty$  means that the limit doesn't exist.

The easiest way to learn about limits is to use them on something that you are already familiar with: finding the value of a function. In the picture below, you can see the graph of a function.



At point A, the value of  $x$  is 1. When you put that into the equation of the function,  $y = -x^3 + 2x + 2$ , you get a  $y$ -value of 3. We say that the value of the function is 3 when  $x$  is equal to 1. If  $x$  is

not 1, but just a little bit bigger, you can say that the value of the function is close to 3. The closer  $x$  gets to 1, the closer the function value gets to 3. In this case we would say that  $x$  is approaching 1 from the right, since we are considering values of  $x$  that are to the right of 1 on the  $x$ -axis. We could also look at  $x$  values smaller than 1, and then get closer and closer to 1 from the left. Either way, the function value approaches 3 as  $x$  approaches 1 more and more closely. We say that the **limit** of the function value is 3 as  $x$  approaches 1. By the way, I will use  $y = -x^3 + 2x + 2$  interchangeably with  $f(x) = -x^3 + 2x + 2$ . The “ $y = \dots$ ” notation looks simpler and it is what you actually use to construct a graph of the function. The “ $f(x) = \dots$ ” way of writing a function allows us to specify what value of  $x$  was used to obtain a particular value of  $y$  shown on the graph:  $f(1) = 3$  is a quick way to write that  $y$  is equal to 3 when  $x$  is equal to 1.

Limits are very straightforward when the function has a continuous graph. However, some functions have a break or a hole in the graph. For example,  $y = \frac{x^2 - 9}{x - 3}$  does not have a value when  $x = 3$ . The best we can do is get really, really close to 3 and see what happens. You may want to take your calculator and try that out. You should find that  $y$  gets closer and closer to 6 as  $x$  gets closer and closer to 3. This happens whether you start at 2.9 and go up, or at 3.1 and go down. Even though the function does not have a value at  $x = 3$ , we say that the limit of the function value, as  $x$  approaches 3, is 6.

As you do your calculations, you might notice that when you get really close to 3, the bottom part of the fraction gets to be very small. Normally, when you divide something by a very small decimal number, like .0001, you end up with a large result. In this case however, the top part of the fraction also gets smaller. Just calculate the top and the bottom parts separately to see that. The reason is that  $x^2 - 9$  is the same as  $(x - 3)$  times  $(x + 3)$ . The bottom part of the fraction can get extremely small, but it is always cancelled out by the  $(x - 3)$  in the top part. In the end, the actual value of the fraction is always what is left,  $x + 3$ . Even though  $x$  is not permitted to be 3, the value of the function,  $x + 3$ , approaches 6 as  $x$  approaches 3.

You should become comfortable with the idea that a limit is something that you may not be able to reach but that you can come as close to as you want. To make the definition of the word limit even more solid, textbook examples will feature piecewise defined functions. The image below shows the piecewise function  $y = \frac{x^2 - 9}{x - 3}$  if  $x \neq 3$  and  $y = 4$  if  $x = 3$ :

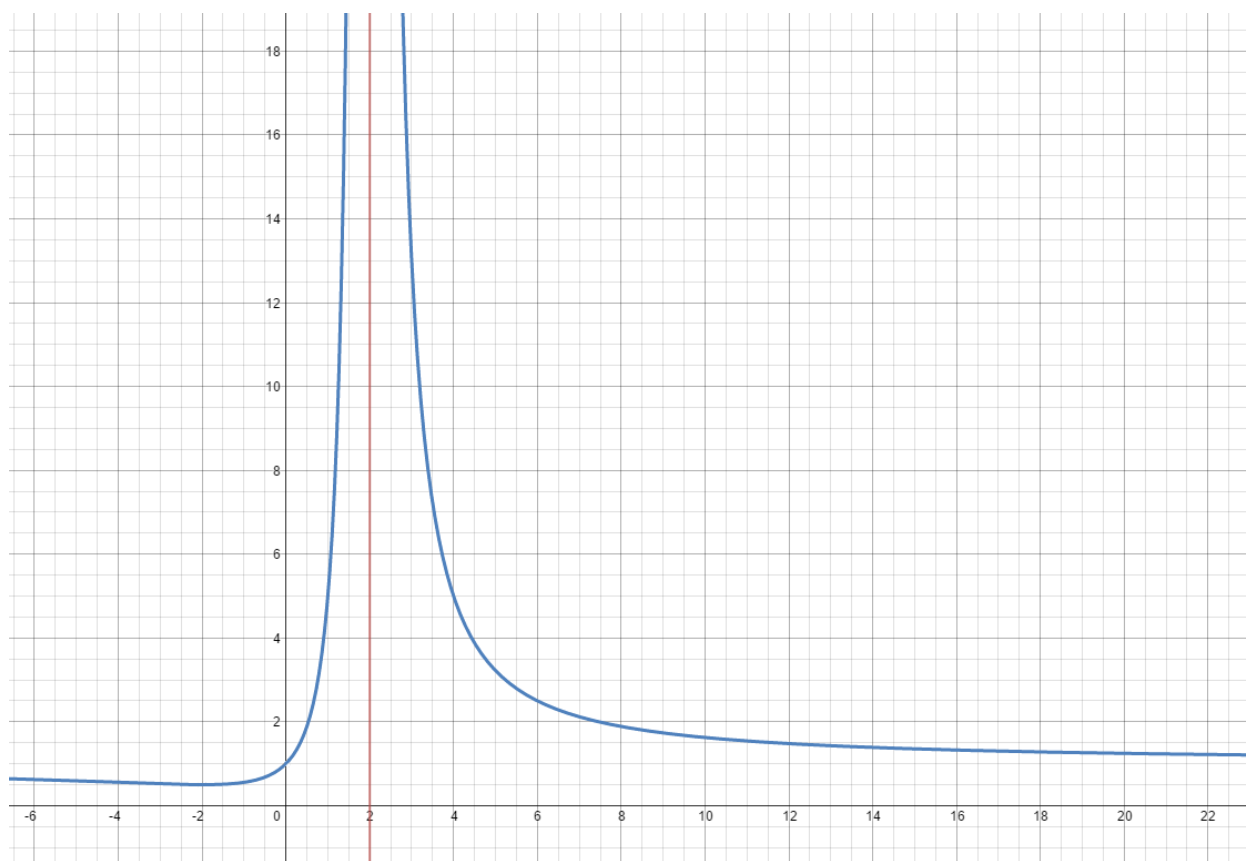
Caution: the open circle shown on the graph at the point  $(3, 6)$  is misleading in the sense that it is a lot smaller than it looks. The “hole” that it represents is infinitely small so you shouldn’t be able to see it at all. In fact,  $x$  could be 3.00000000000000000000000000000001 without causing any problems. Realizing this may make it easier for you to understand that there is still a limit at such a point.

As you go through your chapter you will learn about left-hand limits and right-hand limits. These limits will be indicated by  $+$  and  $-$  signs. To get a right-hand limit, put your finger on the function graph to the right of the point at which you are trying to find the limit, and then move closer to it. For the left-hand limit you start to the left and move right. If you end up with two different values when you do this, *the* Limit does not exist. This can happen if there is a break or gap in the function. Another case where this happens is if there is no limit because the function values keep getting larger and larger or smaller and smaller near the point in question.

We say that the limit is positive or negative infinity, but that really means it does not exist. It is like saying, “The sky is the limit.” This expression means that there is no limit.

**Lim =  $\infty$  means there is no limit.**

This kind of thing is usually caused by a rational function that doesn't have a factor in the top part that cancels out the division on the bottom. For example, if  $f(x) = \frac{x^2 + 4}{(x-2)^2}$ , a value of  $x$  very close to 2 causes a division by a very tiny number, and there is nothing in the numerator that cancels it out. That makes the function value shoot to infinity as  $x$  gets close to 2, creating a vertical asymptote at  $x = 2$ . An asymptote is a line that the graph of a function approaches more and more closely, but never reaches. The function  $f(x) = \frac{x^2 + 4}{(x-2)^2}$  is shown below, along with the asymptote at  $x = 2$ . The limit, as  $x$  approaches 2, is infinity. The limit does not exist.



Looking for  $\lim_{x \rightarrow a}$  of a rational algebraic (no trig, logs or  $b^x$ ) function? There are 3 possibilities:

1. The function exists at the point in question so you can plug in the value directly
2. The top and bottom have a common factor (there is a removable discontinuity)
3. There is no common factor so the function has a vertical asymptote ( $\lim = \pm\infty$ ).

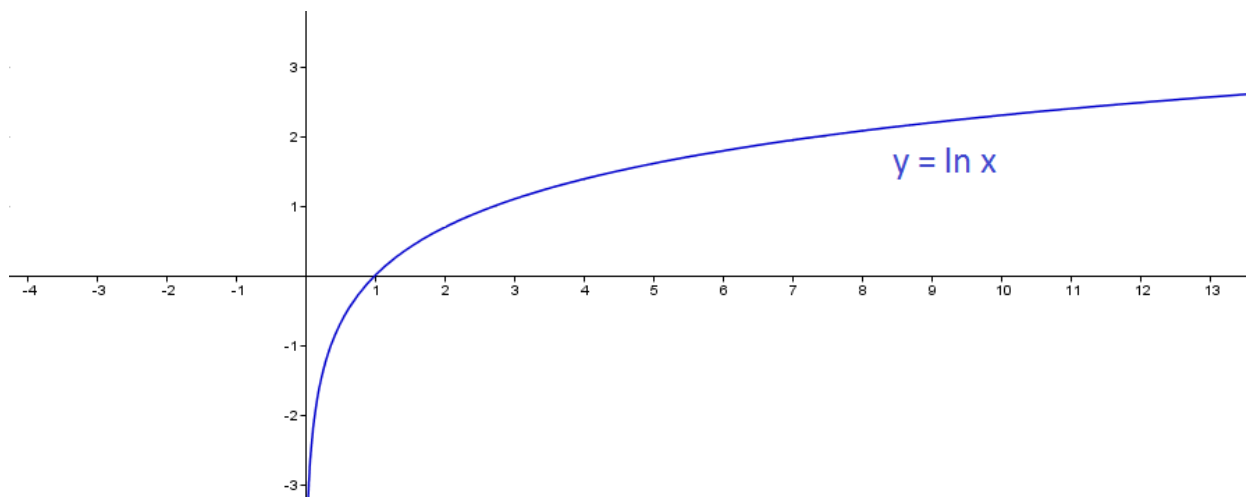
Don't hesitate to use your graphing calculator or graphing software to see what is going on.

Limits can often be computed by using algebraic manipulations. Usually these manipulations involve: factoring, the difference of two squares, the difference or sum of two cubes, and the difference of two squares in reverse. That last part is useful when you see a fractional expression that has  $\sqrt{\dots} - \sqrt{\dots}$  on the top or the bottom. Multiply by  $\frac{\sqrt{\dots} + \sqrt{\dots}}{\sqrt{\dots} + \sqrt{\dots}}$  to eliminate the radicals on one end, and don't worry about the other end. Something will probably cancel out, allowing you to find the limit.

**As you rearrange limit expressions using various tricks, do not rush to multiply out the terms on the top or the bottom. Since you are looking for things to cancel out, the factored form is usually better.**

If there is no way to change the expression in your problem to a more favorable one, you may be expected to find the limit simply by considering what actually happens when  $x$  approaches the indicated value.

The next example involves the natural logarithm function,  $\ln(x)$ . You may recall that this function expresses any number  $x$  as the number  $e$  raised to some power. If that doesn't sound familiar, look at the "Exponential Functions and Logarithms" chapter. Here is a picture of  $y = \ln x$ :



### **Example**

Find  $\lim_{x \rightarrow 0^+} \frac{1 - \ln x}{x}$

$x \rightarrow 0^+$  means the right-hand limit. Because  $\ln x$  doesn't exist to the left of zero, we can't possibly find the left-hand limit here. So, consider what happens to  $\ln x$  as  $x$  gets very small. You can see that the graph takes a nosedive when it gets near zero. The reason for that is that the only way to create a very small number using  $e^x$  is to put a large negative exponent on  $e$ . The natural log function gives you that exponent, so it returns an extremely large negative number when  $x$  is near 0. Try it out on your calculator. This means that  $1 - \ln x$  will be an extremely large positive number when  $x$  is near 0. And what happens when you take a really large number and divide it by a really tiny number? It just gets even bigger.  $\lim_{x \rightarrow 0^+} \frac{1 - \ln x}{x} = \infty$ , which means that it doesn't exist. You can check your work by entering  $y = \frac{1 - \ln x}{x}$  in your favorite graphing app.

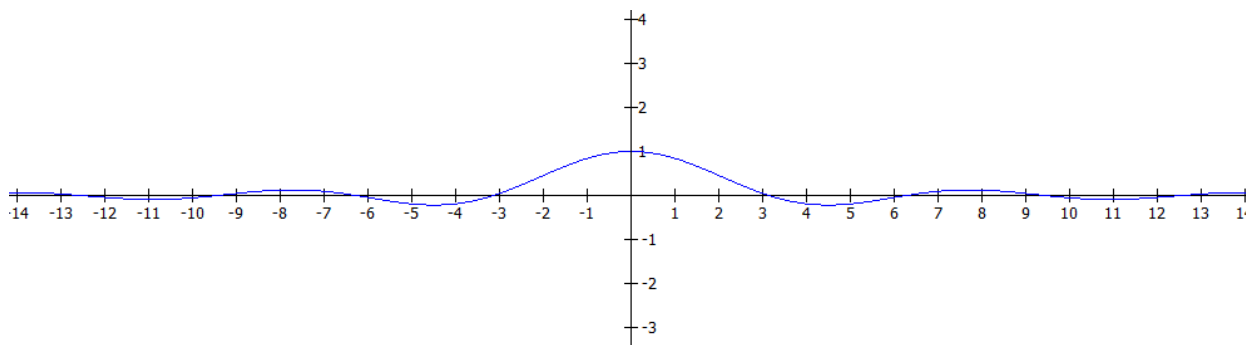


## The Squeeze Theorem and a Special Limit

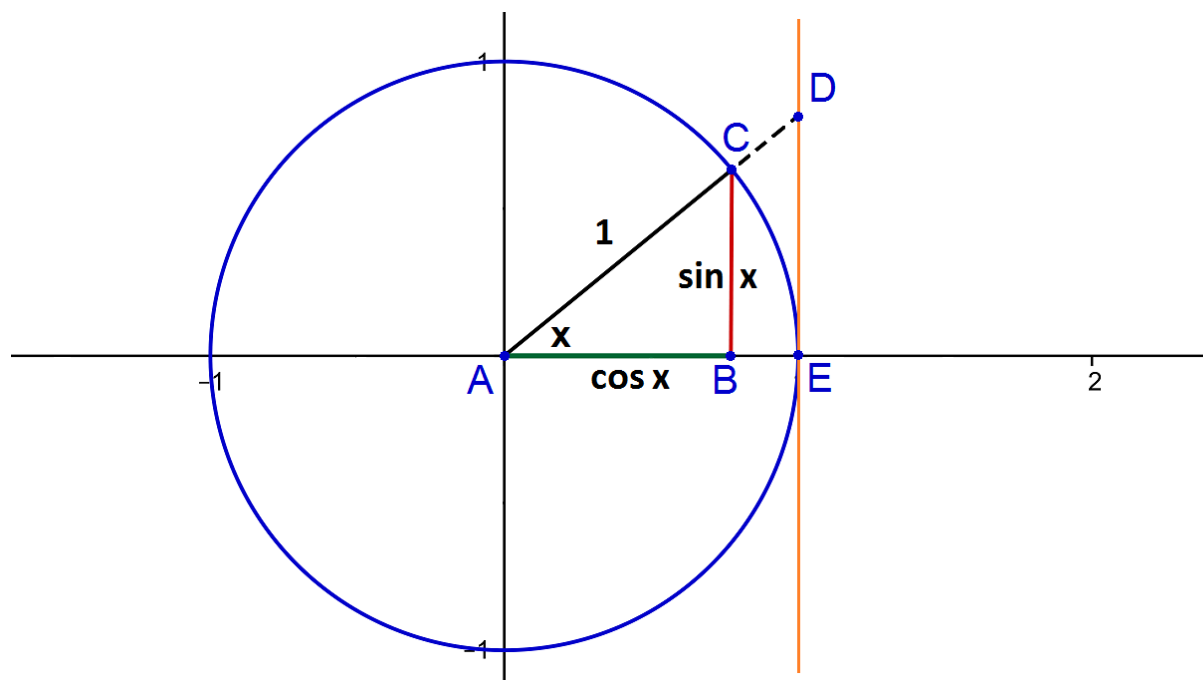
$$\lim_{x \rightarrow 0} \frac{\sin(x)}{x} = \lim_{x \rightarrow 0} \frac{x}{\sin(x)} = 1$$

If you can show that the value of  $f(x)$  is always between that of two other functions, you can use the other functions to determine the limit of  $f(x)$  at a problem point.

If you look at the graph of  $y = \frac{\sin x}{x}$  you can guess that the limit should be 1. This function has an infinitely small hole at  $x = 0$ :



Although we can see that the limit of  $\frac{\sin x}{x}$  is 1, mathematicians like to prove important facts like this. In this case the proof illustrates an important principle called the Squeeze Theorem, so we'll take a close look at it. Below is a picture of the unit circle with an angle  $x$  drawn in the first quadrant.



The sine of the angle is marked in red, and the cosine is marked in green. The orange line is a tangent line to the circle. Because triangle ABC is similar to triangle AED,  $\frac{CB}{AB} = \frac{\sin x}{\cos x} = \tan x = \frac{DE}{AE}$ . This is a unit circle so the length of AE is 1, and the length of DE represents the value of the tangent of angle x.

This incidentally is the real tangent of an angle; it is the length of the part of the tangent line that is intercepted by the angle, which happens to be the same as the sine divided by the cosine. The area of triangle ABC is  $\frac{1}{2}$  times the base AB times the height BC, which is  $\frac{1}{2} \cos x \sin x$ . That area is obviously smaller than the piece of the circle that has endpoints A, E and C. In turn, the piece of the circle is smaller than triangle AED. To find the area of the slice of the circle (the sector of the circle), we have to figure out what portion of the circle it represents. Recall that there are  $2\pi$  radians in a circle, so a piece with angle x is  $x/2\pi$  times the area of the circle. For a unit circle like this, the area of the circle is just  $\pi$ .  $\frac{x}{2\pi} \cdot \pi = \frac{x}{2}$ . Still larger than the sector of the circle is the triangle AED. Its area is  $\frac{1}{2} \cdot 1 \cdot \tan x$ , which is  $\frac{1}{2} \tan x$ .

Now we can set up an inequality:

$$\frac{\cos x \sin x}{2} \leq \frac{x}{2} \leq \frac{\tan x}{2}$$

Multiply everything by 2 to get

$$\cos x \sin x \leq x \leq \frac{\sin x}{\cos x}$$

Divide everything by  $\sin x$  while x is not yet zero:

$$\cos x \leq \frac{x}{\sin x} \leq \frac{1}{\cos x}$$

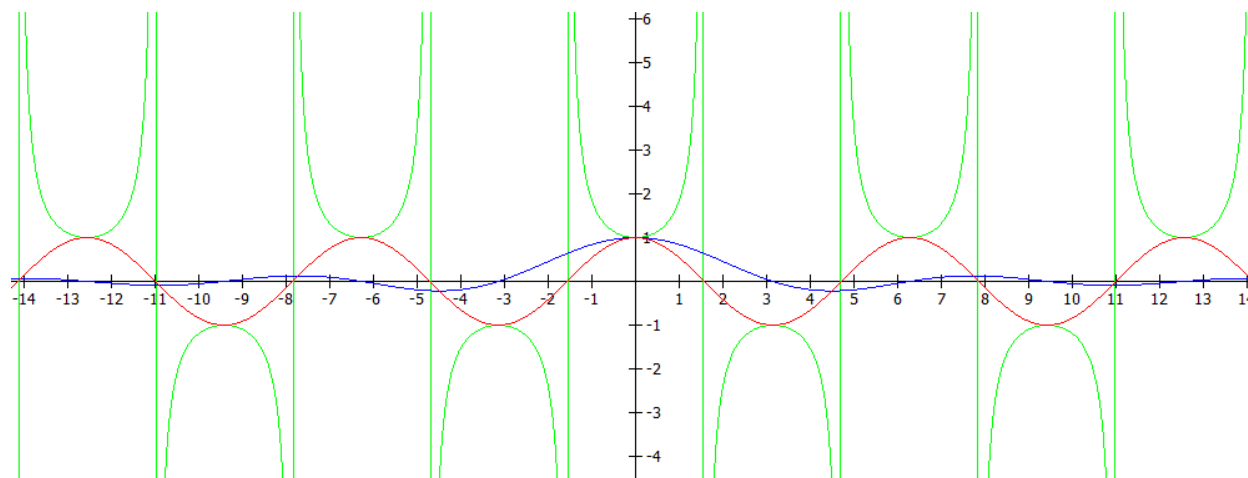
Note that if  $\frac{a}{b} \leq \frac{c}{d}$ , where  $a, b, c$  and  $d$  are all positive numbers, then  $a \leq \frac{bc}{d}$ . Multiply by  $d$  to get  $ad \leq bc$ . Dividing by  $a$  gives  $d \leq \frac{bc}{a}$ , and dividing by  $c$  we get  $\frac{d}{c} \leq \frac{b}{a}$ . So, if you take the reciprocals of the reciprocals of the fractions, reverse the less than or equals sign:

$$\frac{1}{\cos x} \geq \frac{\sin x}{x} \geq \cos x$$

This statement shows that for an angle  $x$  in the first quadrant,  $\frac{\sin x}{x}$  is always between  $\frac{1}{\cos x}$  and  $\cos x$ . As  $x$  goes to zero,  $\frac{1}{\cos x}$  goes to 1, and so does  $\cos x$ . The limit we want,  $\lim_{x \rightarrow 0} \frac{\sin x}{x}$ , is squeezed between the other two limits which are both 1. We conclude that  $\lim_{x \rightarrow 0} \frac{\sin x}{x} = 1$ .

(From the explanation above, you can also see that  $\lim_{x \rightarrow 0} \frac{x}{\sin x} = 1$ ). Here we approached this limit from the right by drawing our triangles in the first quadrant. We can draw these same triangles in the fourth quadrant to approach the limit from the left with the same result.

The picture below shows  $y = \frac{\sin x}{x}$  being squeezed between the graph of  $\sec x$  (green) and  $\cos x$  (red):



## Trigonometric Limits

$$\lim_{x \rightarrow 0} \frac{\sin(mx)}{mx} = 1$$

$$\lim_{x \rightarrow 0} \frac{\tan(mx)}{mx} = 1$$

$$\lim_{x \rightarrow 0} \frac{1 - \cos(mx)}{mx} = \lim_{x \rightarrow 0} \frac{\cos(mx) - 1}{mx} = 0$$

If you can determine that  $\lim_{x \rightarrow 0} \frac{f(x)}{g(x)} = \frac{m}{n}$ , then  $\lim_{x \rightarrow 0} \frac{g(x)}{f(x)} = \frac{n}{m}$ .

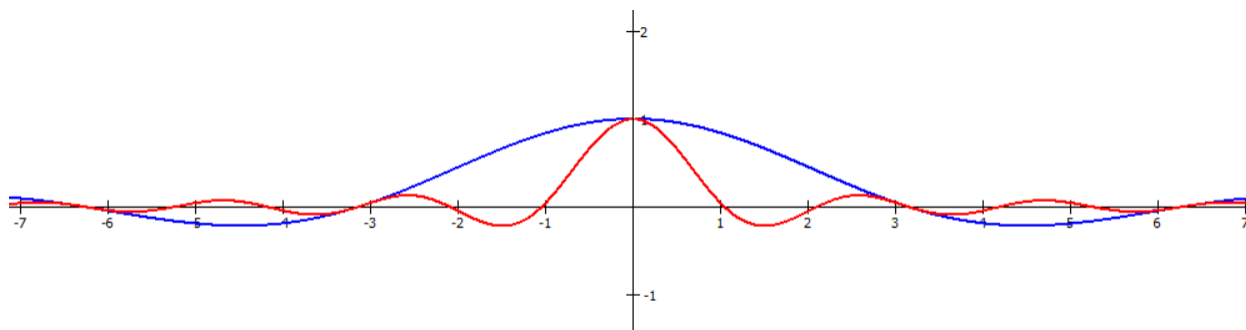
In the last section we saw that  $\lim_{x \rightarrow 0} \frac{\sin x}{x} = 1$ , and  $\lim_{x \rightarrow 0} \frac{x}{\sin x} = 1$ . In general, the limit of the reciprocal is the reciprocal of the original limit. So, if you can determine that  $\lim_{x \rightarrow 0} \frac{f(x)}{g(x)} = \frac{m}{n}$ , then

$$\lim_{x \rightarrow 0} \frac{g(x)}{f(x)} = \frac{n}{m}.$$

We can also make simple substitution for  $x$ :

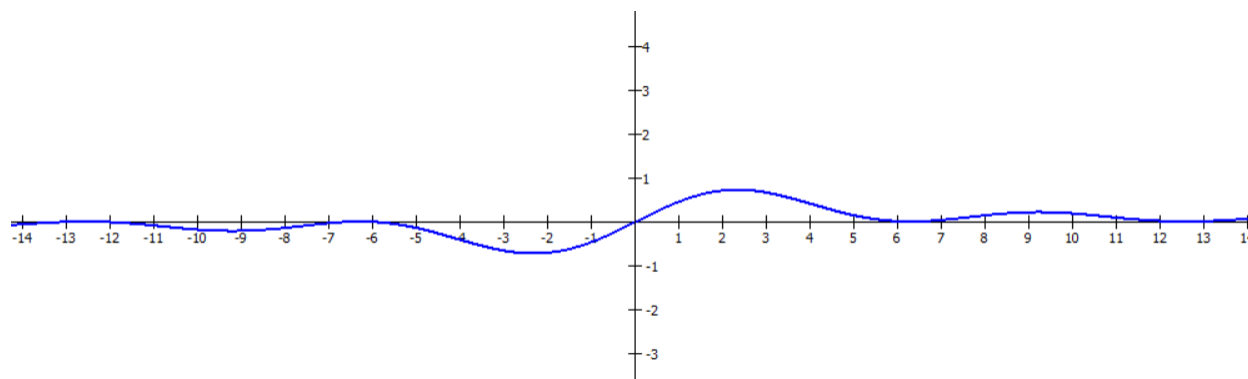
$$\lim_{x \rightarrow 0} \frac{\sin x}{x} = 1 \quad \lim_{x \rightarrow 0} \frac{\sin h}{h} = 1 \quad \lim_{x \rightarrow 0} \frac{\sin(3x)}{3x} = 1$$

That last one works because multiplying  $x$  by 3 only compresses the graph horizontally, as shown by the red graph below.



In general, we can say that  $\lim_{x \rightarrow 0} \frac{\sin(mx)}{mx} = 1$

Unfortunately, there is no limit for  $f(x) = \frac{\cos x}{x}$ . A reason for that is that this limit does not have the form  $\frac{0}{0}$  like  $\frac{\sin x}{x}$ . Instead, the cosine approaches 1 as  $x$  gets smaller and smaller. That causes  $\frac{\cos x}{x}$  to shoot to negative and positive infinity as  $x$  gets close to 0. The limit does not exist. We can get around that problem by considering the limit of  $\frac{1 - \cos x}{x}$  as  $x$  goes to zero:



You can see from the graph that this limit is zero.

Once you know a few basic limits, you can find more.

### Example

Find  $\lim_{x \rightarrow 0} \frac{\tan x}{x}$

$\lim_{x \rightarrow 0} \frac{\tan x}{x} = \lim_{x \rightarrow 0} \frac{\frac{\sin x}{\cos x}}{x}$ . To divide  $\frac{\sin x}{\cos x}$  by  $x$ , we should multiply it by  $\frac{1}{x}$ . Then we get

$$\lim_{x \rightarrow 0} \frac{\tan x}{x} = \lim_{x \rightarrow 0} \frac{\sin x}{x \cos x}$$

Thanks to the limit laws, we know that we can split this into

$$\lim_{x \rightarrow 0} \left( \frac{\sin x}{x} \cdot \frac{1}{\cos x} \right) = \lim_{x \rightarrow 0} \frac{\sin x}{x} \cdot \lim_{x \rightarrow 0} \frac{1}{\cos x}$$

$\lim_{x \rightarrow 0} \frac{1}{\cos x}$  can just be determined by inserting 0 for  $x$ :  $\lim_{x \rightarrow 0} \frac{1}{\cos 0} = 1$ .

$$\lim_{x \rightarrow 0} \frac{\sin x}{x} \cdot \lim_{x \rightarrow 0} \frac{1}{\cos x} = 1 \cdot 1$$

$$\lim_{x \rightarrow 0} \frac{\tan x}{x} = 1$$

Just like we did with the sine limit, you can make a simple substitution for  $x$  in  $\lim_{x \rightarrow 0} \frac{\tan x}{x}$ , so

$$\text{that } \lim_{x \rightarrow 0} \frac{\tan(mx)}{mx} = \lim_{x \rightarrow 0} \frac{mx}{\tan(mx)} = 1.$$

$$\text{And that also works for the limit with the cosine: } \lim_{x \rightarrow 0} \frac{1 - \cos(mx)}{mx} = \lim_{x \rightarrow 0} \frac{\cos(mx) - 1}{mx} = 0$$

Now look at  $\lim_{x \rightarrow 0} \frac{\sin(3x)}{4x}$ . Notice that the parts with the  $x$  don't match, so we can't just substitute and say that the limit is 1. We can however multiply the top and bottom by  $3x$ :

$$\lim_{x \rightarrow 0} \frac{3x \sin(3x)}{3x \cdot 4x}$$

Now separate that into two parts:

$$\lim_{x \rightarrow 0} \frac{3x \sin(3x)}{3x \cdot 4x} = \lim_{x \rightarrow 0} \frac{\sin(3x)}{3x} \cdot \frac{3x}{4x}$$

Even though  $x$  will get closer and closer to 0, it is never officially 0 because we are using a limit. As a result, we can just cancel  $x$ . Alternatively, we could have multiplied the top and bottom of the fraction by 3 initially, rather than by  $3x$ :

$$\lim_{x \rightarrow 0} \frac{\sin(3x)}{4x} \cdot \frac{3}{3} = \lim_{x \rightarrow 0} \frac{\sin(3x)}{3x} \cdot \frac{3}{4} = 1 \cdot \frac{3}{4} = \frac{3}{4}$$

$$\text{In general, we can say that } \lim_{x \rightarrow 0} \frac{\sin(mx)}{nx} = \frac{m}{n}$$

$$\text{In the same way, with just a bit more work, you can show that } \lim_{x \rightarrow 0} \frac{\tan(mx)}{nx} = \frac{m}{n}$$

Here are a few more:

$$\lim_{x \rightarrow 0} \frac{\sin(mx)}{\sin(nx)} = \lim_{x \rightarrow 0} \frac{\tan(mx)}{\tan(nx)} = \lim_{x \rightarrow 0} \frac{\tan(mx)}{\sin(nx)} = \lim_{x \rightarrow 0} \frac{\sin(mx)}{\tan(nx)} = \frac{m}{n}$$

$$\lim_{x \rightarrow 0} \frac{\csc (mx)}{\csc (nx)} = \lim_{x \rightarrow 0} \frac{\cot (mx)}{\cot (nx)} = \lim_{x \rightarrow 0} \frac{\cot (mx)}{\csc (nx)} = \lim_{x \rightarrow 0} \frac{\csc (mx)}{\cot (nx)} = \frac{n}{m}$$