Precalculus Notes

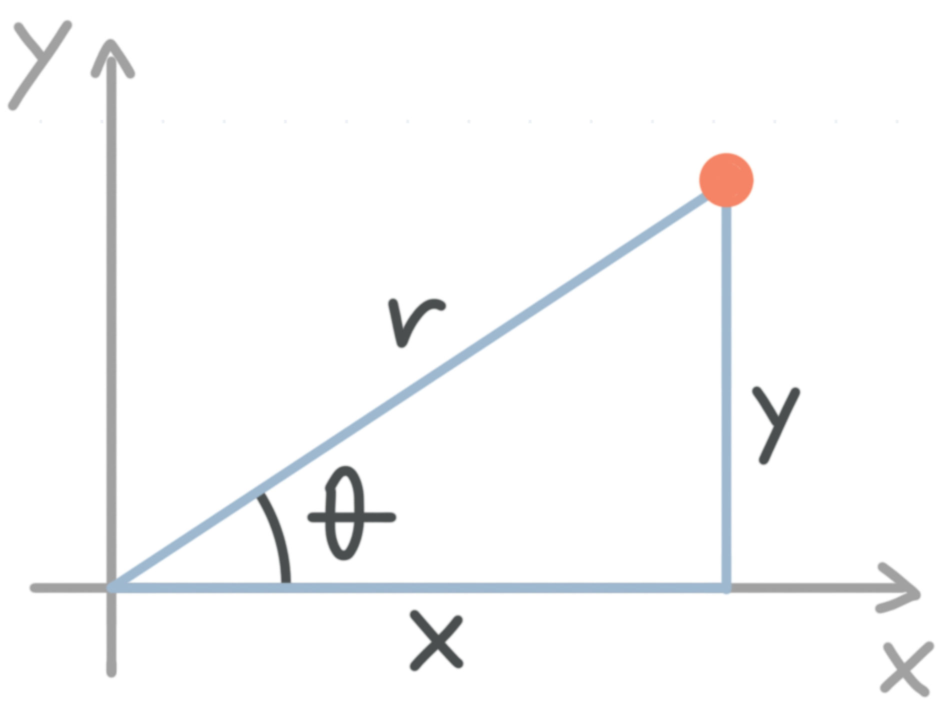
krista king MATH

Polar coordinates

We're already familiar with rectangular coordinates (x, y). We've used them in Algebra, Geometry, and Trigonometry to locate points in the Cartesian (rectangular) plane. We know that x represents the horizontal distance of the point from the origin, while y represents the vertical distance of the point from the origin.

Now we want to change those (x, y) rectangular points into (r, θ) polar points. Whereas (x, y) points give us the horizontal distance x from the origin and the vertical distance y from the origin, (r, θ) points give us the distance r from the origin and the angle θ between the point and the positive direction of the horizontal axis.

We can sketch the x, y, r, and θ, all together in the same plane, as they relate to the same coordinate point.



Notice how both the rectangular system and the polar system can get us to the same point. If we use the rectangular system, we move horizontally parallel to the x-axis, then vertically parallel to the y-axis, until we arrive at our point. But if we use the polar system, we stand at the origin and rotate an angle equivalent to θ, until we're facing the coordinate point, and then we walk a distance of r straight from the origin until we arrive at the point.

The triangle we drew in the diagram is a right triangle, which means we can use right-triangle trigonometry to describe relationships between x, y, r, and θ. Remembering SOH-CAH-TOA (Sine=Opposite/Hypotenuse, Cosine=Adjacent/Hypotenuse, and Tangent=Opposite/Adjacent), we can say

$$\sin\theta = \frac{\text{opposite}}{\text{hypotenuse}} = \frac{y}{r}$$

$$\cos\theta = \frac{\text{adjacent}}{\text{hypotenuse}} = \frac{x}{r}$$

$$\tan\theta = \frac{\text{opposite}}{\text{adjacent}} = \frac{y}{x}$$

If we solve the both the sine and cosine equations for y and x respectively, by multiplying both sides by r, we get

$$y = r\sin\theta$$

$$x = r\cos\theta$$

These are the two most important equations we'll use in order to convert back and forth between rectangular and polar coordinates. Given a point or equation in rectangular coordinates (x, y), we can just substitute $r\sin\theta$ for y and $r\cos\theta$ for x, and we'll have converted into polar coordinates.

Let's do an example so that we can see how to use these equations to convert a polar coordinate point into a rectangular coordinate point.

Example

Convert the polar point into rectangular coordinates (x, y).

$$(r, \theta) = \left(2, \frac{\pi}{3}\right)$$

To rewrite the polar point in terms of rectangular coordinates, we'll plug $r = 2$ and $\theta = \pi/3$ into the conversion equations.

$$x = r\cos\theta$$

$$x = 2\cos\left(\frac{\pi}{3}\right) = 2\left(\frac{1}{2}\right) = 1$$

and

$$y = r\sin\theta$$

$$y = 2\sin\left(\frac{\pi}{3}\right) = 2\left(\frac{\sqrt{3}}{2}\right) = \sqrt{3}$$

So the polar point $(r, \theta) = (2, \pi/3)$ is equivalent to the rectangular point $(x, y) = (1, \sqrt{3})$. The two points are actually exactly the same point in space, we're just expressing that specific point in two different ways.

In this last example, we converted a point in the first quadrant. Let's do another example where we convert a point in the second quadrant.

Example

Convert the polar point into rectangular coordinates (x, y).

$$(r, \theta) = \left(6, \frac{3\pi}{4}\right)$$

To rewrite the polar point in terms of rectangular coordinates, we'll plug $r = 6$ and $\theta = 3\pi/4$ into the conversion equations.

$$x = r\cos\theta$$

$$x = 6\cos\left(\frac{3\pi}{4}\right) = 6\left(-\frac{\sqrt{2}}{2}\right) = -3\sqrt{2}$$

and

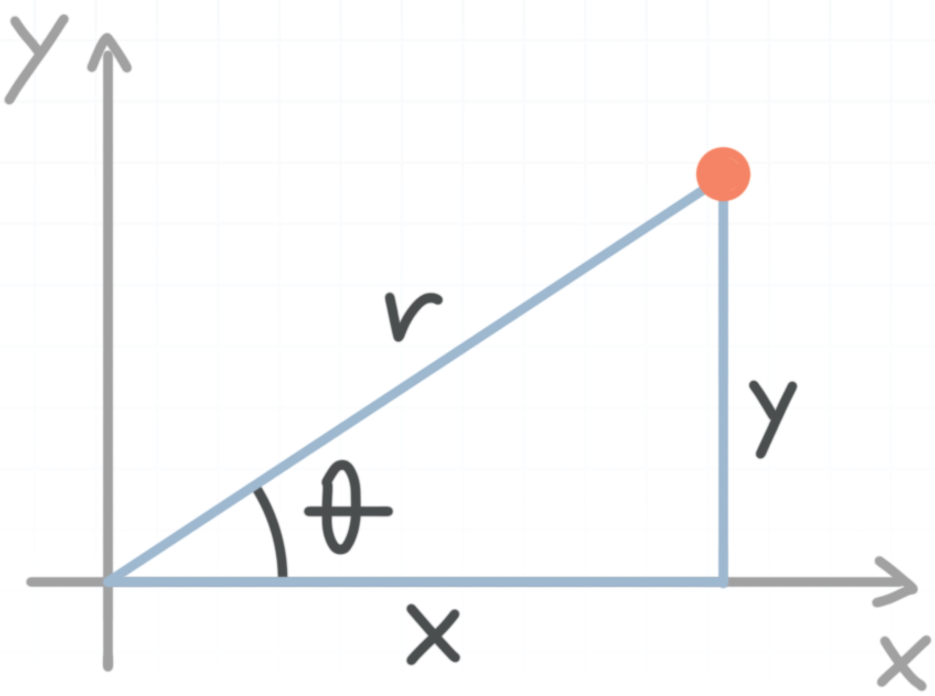
$$y = r\sin\theta$$

$$y = 6\sin\left(\frac{3\pi}{4}\right) = 6\left(\frac{\sqrt{2}}{2}\right) = 3\sqrt{2}$$

So the polar point $(r, \theta) = (6, 3\pi/4)$ is equivalent to the rectangular point $(x, y) = (-3\sqrt{2}, 3\sqrt{2})$.

Two more conversion formulas

We've used $x = r\cos\theta$ and $y = r\sin\theta$ to convert between polar to rectangular coordinates, but there are two other conversion formulas worth mentioning, the first of which also comes from the right triangle that we sketched out earlier.



Remember the Pythagorean Theorem, which tells us that, for right triangles, the sum of the squared leg lengths is equivalent to the squared hypotenuse length. In other words, $a^2 + b^2 = c^2$ for a right triangle with legs a and b and hypotenuse c.

Since the legs of the right triangle above are x and y, and the hypotenuse is r, we can say

$$x^2 + y^2 = r^2$$

So to convert between polar and rectangular coordinates, it's also helpful to know

$$r^2 = x^2 + y^2$$

$$r = \sqrt{x^2 + y^2}$$

The last conversion formula we want to build comes again from

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

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

We know that $\tan\theta = \sin\theta/\cos\theta$, so

$$\tan\theta = \frac{\sin\theta}{\cos\theta} = \frac{\frac{y}{r}}{\frac{x}{r}} = \frac{y}{x}$$

If we apply the inverse tangent $\tan^{-1}$ to both sides of this equation, we get

$$\tan^{-1}(\tan\theta) = \tan^{-1}\left(\frac{y}{x}\right)$$

$$\theta = \tan^{-1}\left(\frac{y}{x}\right)$$

Let's do an example where we use these two new conversion formulas in order to convert a rectangular coordinate point into a polar coordinate point.

Example

Convert $(x, y) = (6,11)$ to polar coordinates.

The value of r for this point will be

$$r = \sqrt{x^2 + y^2} = \sqrt{6^2 + 11^2} = \sqrt{36 + 121} = \sqrt{157} \approx 12.5$$

and the value for θ will be

$$\theta = \tan^{-1}\left(\frac{y}{x}\right) = \tan^{-1}\left(\frac{11}{6}\right) \approx \tan^{-1}(1.83) \approx 1.07 \text{ radians}$$

So we could rewrite the rectangular point $(x, y) = (6,11)$ in polar coordinates as approximately

$$(r, \theta) \approx (12.5,1.07)$$

Multiple ways to express polar points

In Trigonometry we learned about **coterminal angles**, which are angles that differ by one full 2π rotation. For instance, the angles $\pi/4$ and $9\pi/4$ are coterminal angles because they differ by 2π.

$$\frac{9\pi}{4} - \frac{\pi}{4} = \frac{9\pi - \pi}{4} = \frac{8\pi}{4} = 2\pi$$

Given an angle θ, the expression

$\alpha = \theta + n(2\pi)$, where n is any integer

represents the complete set of angles (all the angles α) that are coterminal with θ.

Finding equivalent points by changing θ

So given a polar point like $(2,\pi)$, we know that we can find equivalent points just by changing the θ value by an integer-multiple of 2π. For example, all of these angles are coterminal with $\theta = \pi$:

For $n = -3$ $\quad \alpha = \pi - 3(2\pi) = \pi - 6\pi = -5\pi$

For $n = -2$ $\quad \alpha = \pi - 2(2\pi) = \pi - 4\pi = -3\pi$

For $n = -1$ $\quad \alpha = \pi - 1(2\pi) = \pi - 2\pi = -\pi$

For $n = 1$ $\quad \alpha = \pi + 1(2\pi) = \pi + 2\pi = 3\pi$

For $n = 2$ $\quad \alpha = \pi + 2(2\pi) = \pi + 4\pi = 5\pi$

For $n = 3$ $\qquad \alpha = \pi + 3(2\pi) = \pi + 6\pi = 7\pi$

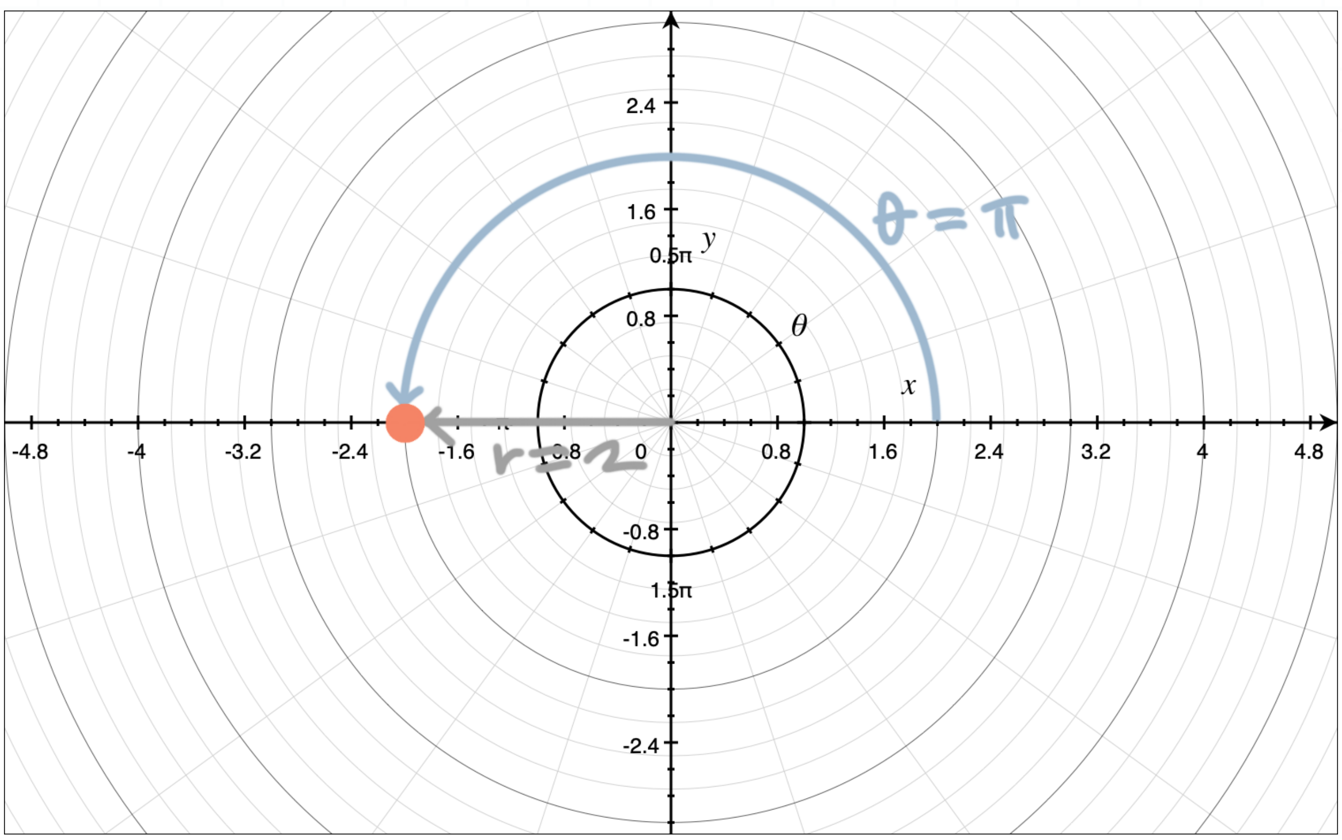
Therefore, all of these polar points are equivalent to the polar point $(2,\pi)$:

$(2, -5\pi)$ $\qquad (2,3\pi)$

$(2, -3\pi)$ $\qquad (2,5\pi)$

$(2, -\pi)$ $\qquad (2,7\pi)$

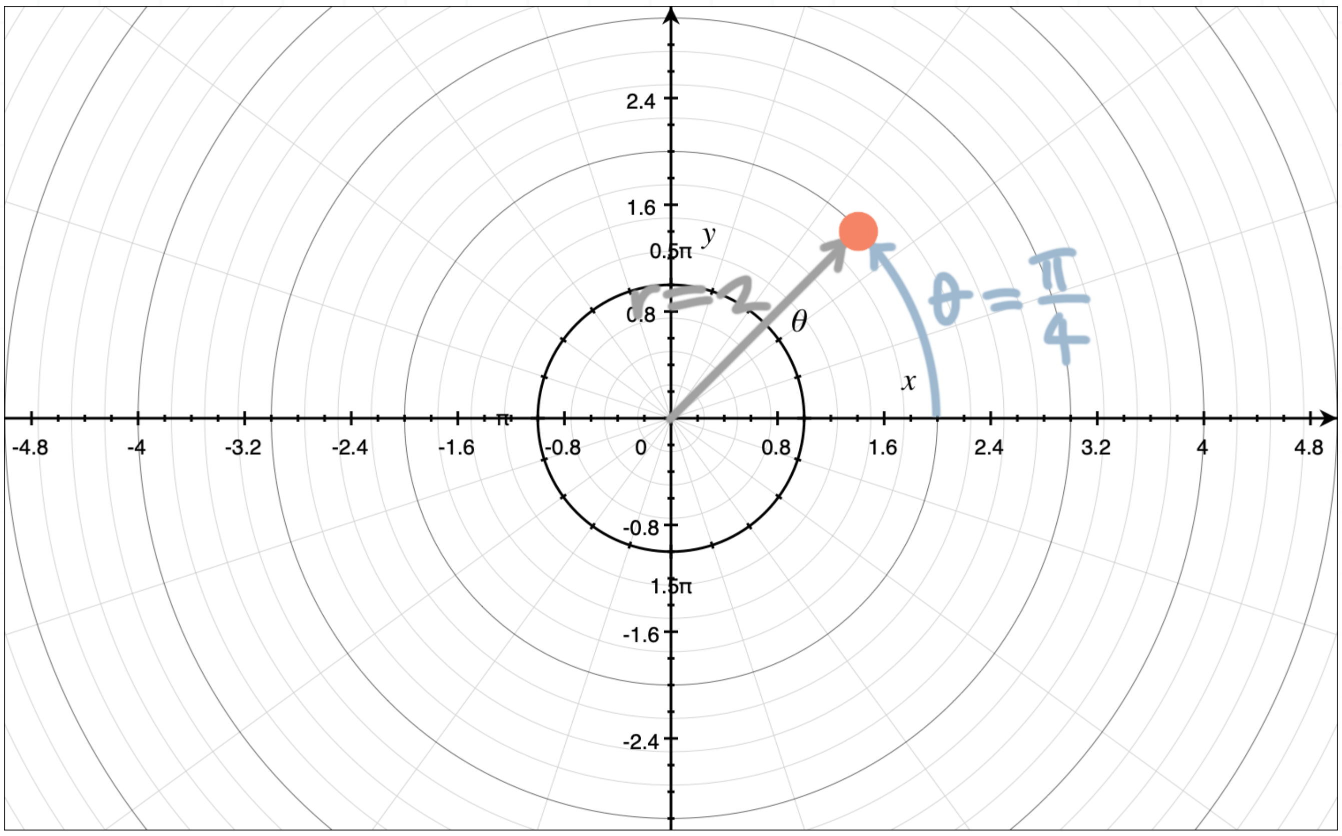
Notice how all of these points have the same r-value as the original point $(2,\pi)$, but we've just changed the θ-value by some multiple of 2π. All of these points are plotted in the same spot,



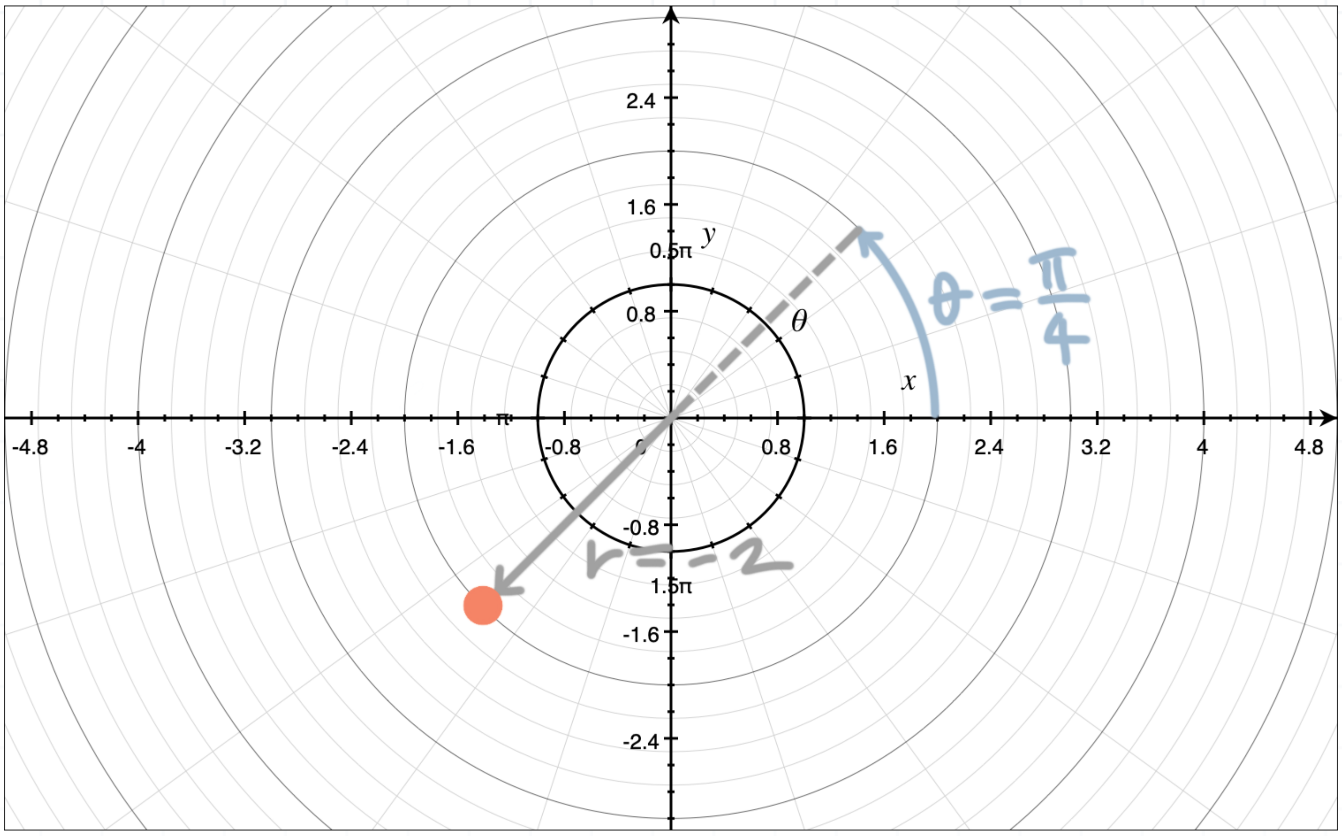
so we've now shown that we can express the same polar point in infinitely many ways, just by continuing to add or take away an extra 2π rotation from the angle θ.

Finding equivalent points by changing r

When we see a polar point with a positive r-value, it tells us to walk forward, from the **pole**, $(0,0)$, toward the angle θ. For instance, given the point $(2,\pi/4)$, we stand on the pole and rotate toward the angle $\pi/4$ in the first quadrant, and then walk forward into the first quadrant, to arrive at $(2,\pi/4)$.



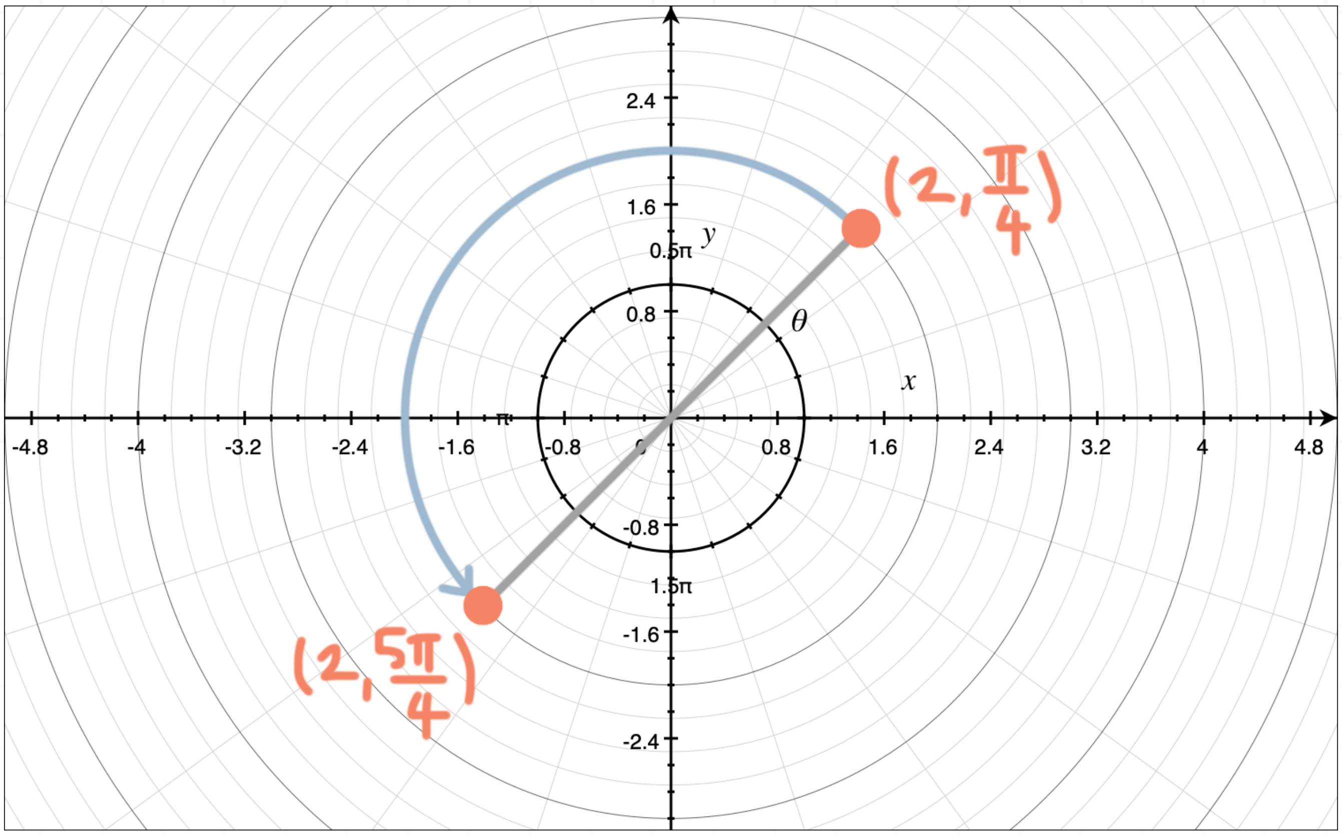
But when we see a polar point with a negative r-value, it tells us to walk backward away from the angle θ. For instance, given the point $(-2,\pi/4)$, we rotate toward the angle $\pi/4$ in the first quadrant, and then walk straight backward into the third quadrant, to arrive at $(-2,\pi/4)$.



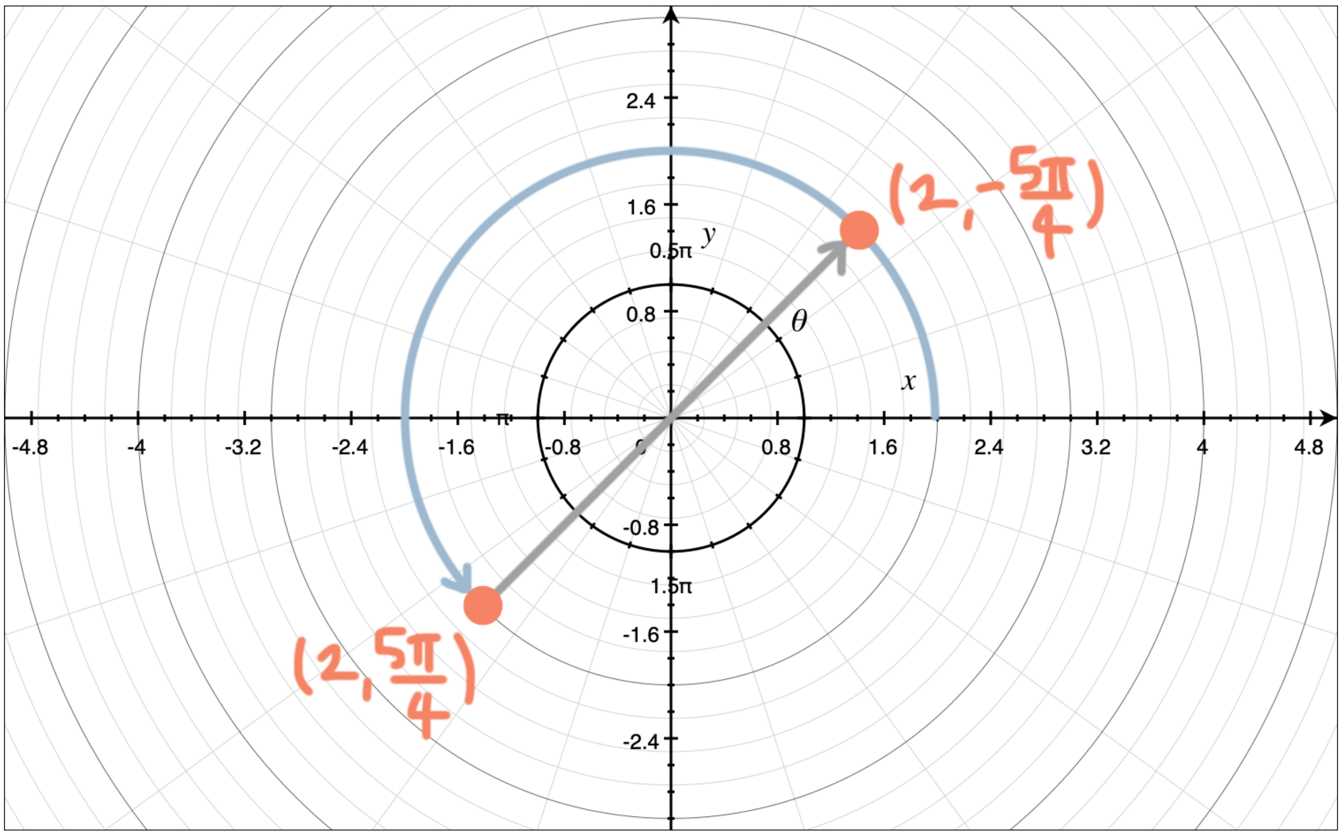
We want to notice from this example that the point $(-2,\pi/4)$ is a half-circle rotation away (a π rotation away) from $(2,\pi/4)$. Which means we could also express $(-2,\pi/4)$ as

$$\left(2,\frac{\pi}{4}+\pi\right)=\left(2,\frac{\pi}{4}+\frac{4\pi}{4}\right)=\left(2,\frac{5\pi}{4}\right)$$

So the point $(-2,\pi/4)$ is equivalent to the point $(2,5\pi/4)$. But remember, we're looking for points that are equivalent to $(2,\pi/4)$. What we've learned is that we can rotate a half circle from $(2,\pi/4)$ to get to $(2,5\pi/4)$,



and then change the value of r from 2 to -2 in order to get back to the original point.



So $(-2,5\pi/4)$ is the same point as $(2,\pi/4)$. And because we know that we can change θ by any multiple of 2π to find an equivalent point, all of these points are equivalent to $(-2,5\pi/4)$, and therefore also equivalent to $(2,\pi/4)$:

For $n = -3$ $\qquad \left(-2, \frac{5\pi}{4} - 3(2\pi)\right) = \left(-2, -\frac{19\pi}{4}\right)$

For $n = -2$ $\qquad \left(-2, \frac{5\pi}{4} - 2(2\pi)\right) = \left(-2, -\frac{11\pi}{4}\right)$

For $n = -1$ $\qquad \left(-2, \frac{5\pi}{4} - 1(2\pi)\right) = \left(-2, -\frac{3\pi}{4}\right)$

For $n = 1$ $\qquad \left(-2, \frac{5\pi}{4} + 1(2\pi)\right) = \left(-2, \frac{13\pi}{4}\right)$

For $n = 2$ $\qquad \left(-2, \frac{5\pi}{4} + 2(2\pi)\right) = \left(-2, \frac{21\pi}{4}\right)$

For $n = 3$ $\qquad \left(-2, \frac{5\pi}{4} + 3(2\pi)\right) = \left(-2, \frac{29\pi}{4}\right)$

To summarize, we can say that there are an infinite number of ways to express the same point in space in polar coordinates. We can

1. Keep the value of r the same but add or subtract any multiple of 2π from θ.

2. Change the value of r to $-r$ while we add or subtract any odd integer multiple of π from θ.

Both of these options produce an infinite number of equivalent points for the original polar point. Let's do an example where we find a different way to express the same polar coordinate point.

Example

Find a point in polar coordinates that's equivalent to $(14,31\pi/7)$.

We can find equivalent points by adding or subtracting multiples of 2π from the angle θ, so all of these are examples of equivalent polar points:

For $n = -3$ $\quad\left(14,\frac{31\pi}{7}-3(2\pi)\right)=\left(14,-\frac{11\pi}{7}\right)$

For $n = -2$ $\quad\left(14,\frac{31\pi}{7}-2(2\pi)\right)=\left(14,\frac{3\pi}{7}\right)$

For $n = -1$ $\quad\left(14,\frac{31\pi}{7}-1(2\pi)\right)=\left(14,\frac{17\pi}{7}\right)$

For $n = 1$ $\quad\left(14,\frac{31\pi}{7}+1(2\pi)\right)=\left(14,\frac{45\pi}{7}\right)$

For $n = 2$ $\quad\left(14,\frac{31\pi}{7}+2(2\pi)\right)=\left(14,\frac{59\pi}{7}\right)$

For $n = 3$ $\quad\left(14,\frac{31\pi}{7}+3(2\pi)\right)=\left(14,\frac{73\pi}{7}\right)$

We could also find an equivalent point by simultaneously changing $r = 14$ to $r = -14$, and adding π to $31\pi/7$ to get

$$\left(-14,\frac{31\pi}{7}+\pi\right)=\left(-14,\frac{38\pi}{7}\right)$$

Then we can find more equivalent polar points with the $r = -14$ value by adding and subtracting multiples of 2π from the angle θ, so this is another set of points which are also equivalent to $(14,31\pi/7)$:

For $n = -3$ $$\left(-14,\frac{38\pi}{7} - 3(2\pi)\right) = \left(-14, -\frac{4\pi}{7}\right)$$

For $n = -2$ $$\left(-14,\frac{38\pi}{7} - 2(2\pi)\right) = \left(-14,\frac{10\pi}{7}\right)$$

For $n = -1$ $$\left(-14,\frac{38\pi}{7} - 1(2\pi)\right) = \left(-14,\frac{24\pi}{7}\right)$$

For $n = 1$ $$\left(-14,\frac{38\pi}{7} + 1(2\pi)\right) = \left(-14,\frac{52\pi}{7}\right)$$

For $n = 2$ $$\left(-14,\frac{38\pi}{7} + 2(2\pi)\right) = \left(-14,\frac{66\pi}{7}\right)$$

For $n = 3$ $$\left(-14,\frac{38\pi}{7} + 3(2\pi)\right) = \left(-14,\frac{80\pi}{7}\right)$$

These are all examples of points that are equivalent to $(14,31\pi/7)$, but we could list infinitely more.

Let's do one more, but this time we'll start with a negative value of r.

Example

Find some polar points that are equivalent to $(-20, -18\pi/11)$.

We can find equivalent points by adding or subtracting multiples of 2π from the angle θ, so these are examples of some equivalent polar points:

For $n = -3$ $\quad \left(-20, -\frac{18\pi}{11} - 3(2\pi)\right) = \left(-20, -\frac{84\pi}{11}\right)$

For $n = -2$ $\quad \left(-20, -\frac{18\pi}{11} - 2(2\pi)\right) = \left(-20, -\frac{62\pi}{11}\right)$

For $n = -1$ $\quad \left(-20, -\frac{18\pi}{11} - 1(2\pi)\right) = \left(-20, -\frac{40\pi}{11}\right)$

For $n = 1$ $\quad \left(-20, -\frac{18\pi}{11} + 1(2\pi)\right) = \left(-20, \frac{4\pi}{11}\right)$

For $n = 2$ $\quad \left(-20, -\frac{18\pi}{11} + 2(2\pi)\right) = \left(-20, \frac{26\pi}{11}\right)$

For $n = 3$ $\quad \left(-20, -\frac{18\pi}{11} + 3(2\pi)\right) = \left(-20, \frac{48\pi}{11}\right)$

We could also find an equivalent point by simultaneously changing $r = -20$ to $r = 20$, and adding π to $-18\pi/11$ to get

$$\left(20, -\frac{18\pi}{11} + \pi\right) = \left(20, -\frac{7\pi}{11}\right)$$

Then we can find more equivalent polar points with the $r = 20$ value by adding and subtracting multiples of 2π from the angle θ, so this is another set of points which are also equivalent to $(-20, -18\pi/11)$:

For $n=-3$ $$\left(20,-\frac{7\pi}{11}-3(2\pi)\right)=\left(20,-\frac{73\pi}{11}\right)$$

For $n=-2$ $$\left(20,-\frac{7\pi}{11}-2(2\pi)\right)=\left(20,-\frac{51\pi}{11}\right)$$

For $n=-1$ $$\left(20,-\frac{7\pi}{11}-1(2\pi)\right)=\left(20,-\frac{29\pi}{11}\right)$$

For $n=1$ $$\left(20,-\frac{7\pi}{11}+1(2\pi)\right)=\left(20,\frac{15\pi}{11}\right)$$

For $n=2$ $$\left(20,-\frac{7\pi}{11}+2(2\pi)\right)=\left(20,\frac{37\pi}{11}\right)$$

For $n=3$ $$\left(20,-\frac{7\pi}{11}+3(2\pi)\right)=\left(20,\frac{59\pi}{11}\right)$$

These are all examples of points that are equivalent to $(-20,-18\pi/11)$, but we could list infinitely more.

Converting equations

Now that we know how to convert between polar coordinate points (r, θ) and rectangular coordinate points (x, y) using the conversion formulas,

$$x = r\cos\theta \qquad r^2 = x^2 + y^2$$

$$y = r\sin\theta \qquad \tan\theta = \frac{y}{x}$$

we can start using these same conversion formulas to convert between polar and rectangular equations.

Converting polar equations

It's common to see polar equations like $r = 8\cos\theta$, where the equation is defined for r in terms of θ, in the same way that it's common to see rectangular equations like $y = x + 3$, where the equation is defined for y in terms of x.

But, just like we can have rectangular equations with only one variable, like $y = 2$ and $x = 4$, we can also have polar equations with only one variable, like $r = 2$ and $\theta = \pi$. While $y = 2$ represents a perfectly horizontal line and $x = 4$ represents a perfectly vertical line, $r = 2$ represents a perfect circle around the pole and $\theta = \pi$ represents a line from the pole out toward the angle $\theta = \pi$.

We can convert all of these polar equations into rectangular equations. Let's do an example with an equation like $r = 2$.

Example

Convert the polar equation $r = 7$ into rectangular coordinates.

To convert a polar equation in this form, we'll plug $r = 7$ into the conversion equation $r^2 = x^2 + y^2$.

$$7^2 = x^2 + y^2$$

$$x^2 + y^2 = 49$$

So both $r = 7$ and $x^2 + y^2 = 49$ represent a circle centered at $(0,0)$ with radius 7. They are identical curves, but $r = 7$ is defined in polar coordinates, while $x^2 + y^2 = 49$ is defined in rectangular coordinates.

Now let's look at an example with an equation like $\theta = \pi$.

Example

Convert the polar equation $\theta = \pi/3$ into rectangular coordinates.

To convert a polar equation in this form, take the tangent of both sides of the equation.

$$\tan\theta = \tan\left(\frac{\pi}{3}\right)$$

Now we can use the conversion equation $\tan\theta = y/x$ to get

$$\frac{y}{x} = \tan\left(\frac{\pi}{3}\right)$$

To simplify the right side, we'll use the quotient identity for tangent to rewrite tangent as sine over cosine.

$$\tan\left(\frac{\pi}{3}\right) = \frac{\sin\left(\frac{\pi}{3}\right)}{\cos\left(\frac{\pi}{3}\right)} = \frac{\frac{\sqrt{3}}{2}}{\frac{1}{2}} = \frac{\sqrt{3}}{2}\left(\frac{2}{1}\right) = \sqrt{3}$$

Substituting this result, we get

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

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

Let's do one last example, this time with the equation $r = 8\cos\theta$, so that we can see what happens when both r and θ are present in the equation.

Example

Convert the polar equation $r = 8\cos\theta$ into rectangular coordinates.

First, we'll rewrite the conversion equation $x = r\cos\theta$ as

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

Now we can replace $\cos\theta$ in $r = 8\cos\theta$ with x/r.

$$r = 8\left(\frac{x}{r}\right)$$

$$r^2 = 8x$$

Then with the conversion equation $r^2 = x^2 + y^2$, we can substitute $x^2 + y^2$ for r^2.

$$x^2 + y^2 = 8x$$

$$x^2 - 8x + y^2 = 0$$

We weren't asked to do this, but we could rewrite this rectangular equation by completing the square with respect to x,

$$(x^2 - 8x + 16) - 16 + y^2 = 0$$

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

to see that this is the equation of a circle with radius 4 centered at $(4,0)$. Of course, that also gives us a preview of the fact that the polar equation $r = 8\cos\theta$ represents the circle with radius 4 centered at $(4,0)$.

Converting rectangular equations

Now that we can convert polar equations into rectangular equations, let's work backwards and convert rectangular equations into polar equations.

Converting this way it a little easier, because we can always use the conversion equations $x = r\cos\theta$ and $y = r\sin\theta$ to make substitutions for x and y. Every time we see an x, we'll replace it with $r\cos\theta$, and every time we see a y, we'll replace it with $r\sin\theta$.

For some rectangular equations, the conversion formula $x^2 + y^2 = r^2$ can come in handy as well. Let's do an example.

Example

Convert the rectangular equation $x^2 + y^2 = 64$ into polar coordinates.

If we use the conversion equation $x^2 + y^2 = r^2$, we can see right away that we get the polar curve $r^2 = 64$, or $r = 8$.

But we could have also used the conversion equations $x = r\cos\theta$ and $y = r\sin\theta$ to arrive at the same answer.

$$(r\cos\theta)^2 + (r\sin\theta)^2 = 64$$

$$r^2\cos^2\theta + r^2\sin^2\theta = 64$$

$$r^2(\cos^2\theta + \sin^2\theta) = 64$$

If we remember the Pythagorean identity $\sin^2\theta + \cos^2\theta = 1$ from Trigonometry, then we get

$$r^2(1) = 64$$

$$r^2 = 64$$

$$r = 8$$

Let's do another example where we can't use $x^2 + y^2 = r^2$.

Example

Convert the rectangular equation $y = 25x$ into polar coordinates.

Using the conversion equations $x = r\cos\theta$ and $y = r\sin\theta$, we get

$$r\sin\theta = 25r\cos\theta$$

$$\frac{r\sin\theta}{r\cos\theta} = 25$$

$$\frac{\sin\theta}{\cos\theta} = 25$$

From the quotient identity for tangent, we can rewrite the left side of this equation to get

$$\tan\theta = 25$$

We'll do another one to get a little more practice.

Example

Convert the rectangular equation $y = -6x^2$ into polar coordinates.

Using the conversion equations $x = r\cos\theta$ and $y = r\sin\theta$, we get

$$r\sin\theta = -6(r\cos\theta)^2$$

$$\frac{r\sin\theta}{(r\cos\theta)(r\cos\theta)} = -6$$

$$\frac{r\sin\theta}{r\cos\theta}\left(\frac{1}{r\cos\theta}\right) = -6$$

$$\frac{\sin\theta}{\cos\theta}\left(\frac{1}{r\cos\theta}\right) = -6$$

The quotient identity for tangent from Trigonometry gives us

$$\tan\theta\left(\frac{1}{r\cos\theta}\right) = -6$$

$$\frac{\tan\theta}{r\cos\theta} = -6$$

$$r = -\frac{\tan\theta}{6\cos\theta}$$

Choosing the simpler system

This is a great time to point out that some equations are better expressed in polar coordinates, while others are better expressed in rectangular coordinates.

In this previous example, the rectangular $y = -6x^2$ is much easier to understand than its equivalent polar equation,

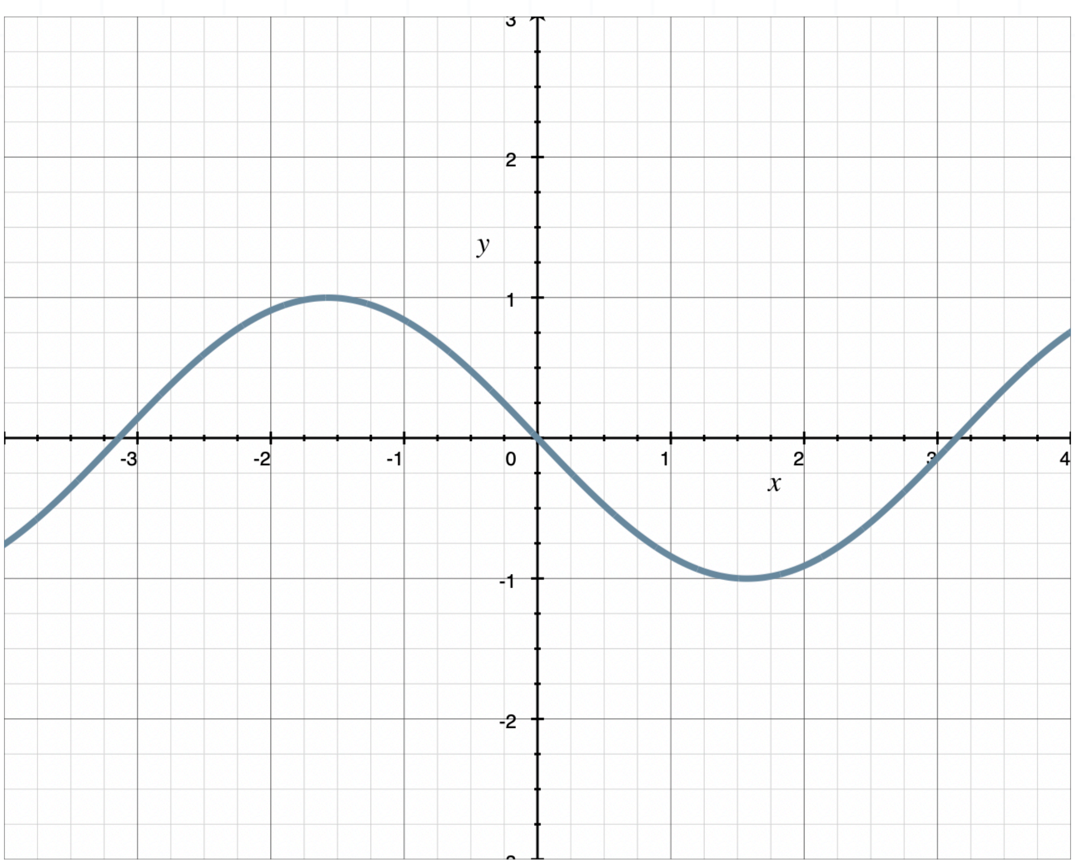
$$r = -\frac{\tan\theta}{6\cos\theta}$$

So this equation is probably best expressed in rectangular coordinates, if we get to choose which coordinate system to use. On the other hand, if we look back at an earlier example, the equation $r = 8$ is a simpler way to express $x^2 + y^2 = 64$, so polar coordinates might be better for that equation.

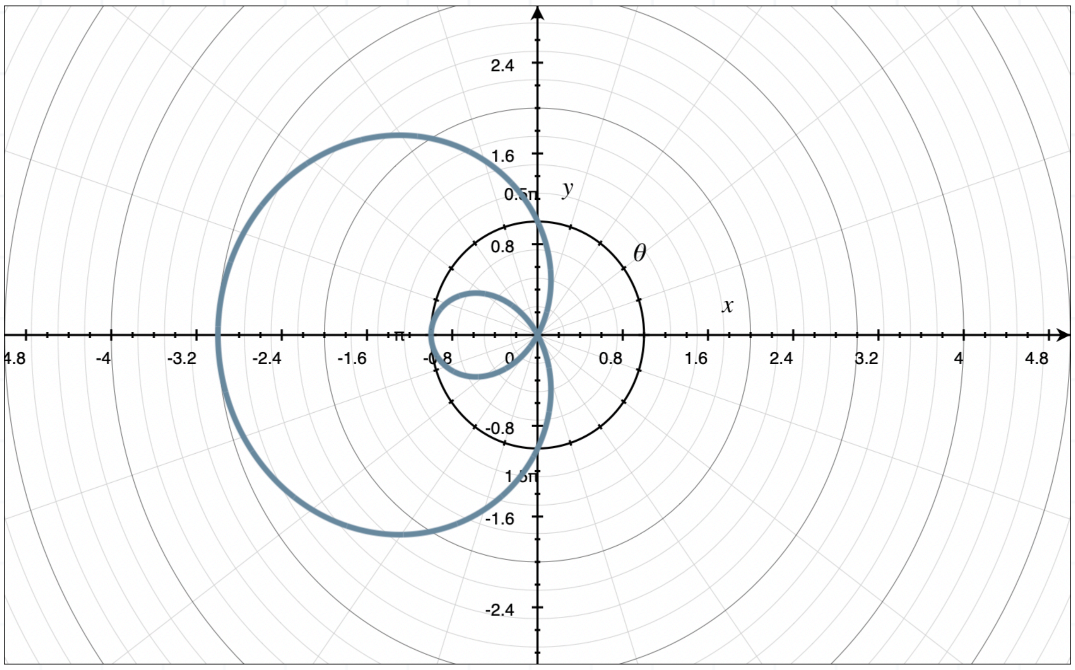
Graphing polar curves in a rectangular system

We want to start looking at how to sketch the graphs of polar curves. In the same way that we learned in Algebra and Trigonometry how to sketch the graphs of lines, parabolas, circles, trig functions, etc. in rectangular coordinates, we can also sketch the graphs of equations in polar coordinates.

For instance, here's the graph of $y=\sin(x-\pi)$, sketched in the rectangular coordinate system,



and here's the graph of $r=1-2\cos\theta$, sketched in the polar coordinate system.



But even though the polar coordinate system is completely different than the rectangular coordinate system, we can still use the rectangular coordinate system to help us sketch polar curves.

The method that uses the rectangular coordinate system has us plot polar points in the xy-plane, and then translate those points into the $r\theta$-plane.

In fact, this will be the primary method we'll use (sketching polar curves first in the xy-plane, and then translating them over into the $r\theta$-plane) as we're learning to sketch polar curves. As we get more comfortable, we'll start to cut out this intermediate step and move to sketching curves directly in the $r\theta$-plane.

Sketching polar curves in the xy-plane

To sketch a polar curve in the rectangular coordinate system, we'll

1. Set the argument of the trig function equal to $\pi/2$, and then solve this new equation for θ

2. Evaluate the polar curve at multiples of θ, starting with $\theta = 0$

3. Plot the resulting points in the xy-plane, treating the horizontal axis as the θ-axis (instead of the x-axis), and treating the vertical axis as the r-axis (instead of the y-axis).

4. Connect the plotted points with a smooth curve.

Once we've sketched the polar curve into the rectangular coordinate system, then we'll be ready to transfer the graph from the rectangular system into the polar system.

We'll walk through these steps in detail in an example, but first let's talk about why we set the argument equal to $\pi/2$. The reason we choose $\pi/2$, and not any other value, is because the sine and cosine functions evaluate to really simple values at multiples of $\pi/2$.

If we think back to what we learned about the unit circle in Trigonometry, we should remember that the values of the basic sine and cosine functions are 0, 1, or -1 at all multiples of $\pi/2$. For instance, if $\theta = \pi/2$, then $\sin\theta = \sin(\pi/2) = 1$ and $\cos\theta = \cos(\pi/2) = 0$.

So even though we could choose any value, choosing $\pi/2$ means that we'll get lots of results that are 0, 1, and -1, and these values are going to be really easy to work with.

Let's do an example where we sketch a polar curve in the rectangular coordinate system.

Example

Sketch the graph of $r = 6\cos\theta$.

The trigonometric function in this polar equation is $\cos\theta$, and its argument (the angle at which cosine is evaluated) is θ. So we'll set

$$\theta = \frac{\pi}{2}$$

This equation is already solved for θ, so we're ready to make a table of values. We'll list out the first few multiples of $\pi/2$, starting with 0, and find their corresponding r-values.

$\theta = 0$ $\qquad$ $r = 6\cos(0) = 6(1) = 6$

$\theta = \pi/2$ $\qquad$ $r = 6\cos(\pi/2) = 6(0) = 0$

$\theta = \pi$ $\qquad$ $r = 6\cos(\pi) = 6(-1) = -6$

$\theta = 3\pi/2$ $\qquad$ $r = 6\cos(3\pi/2) = 6(0) = 0$

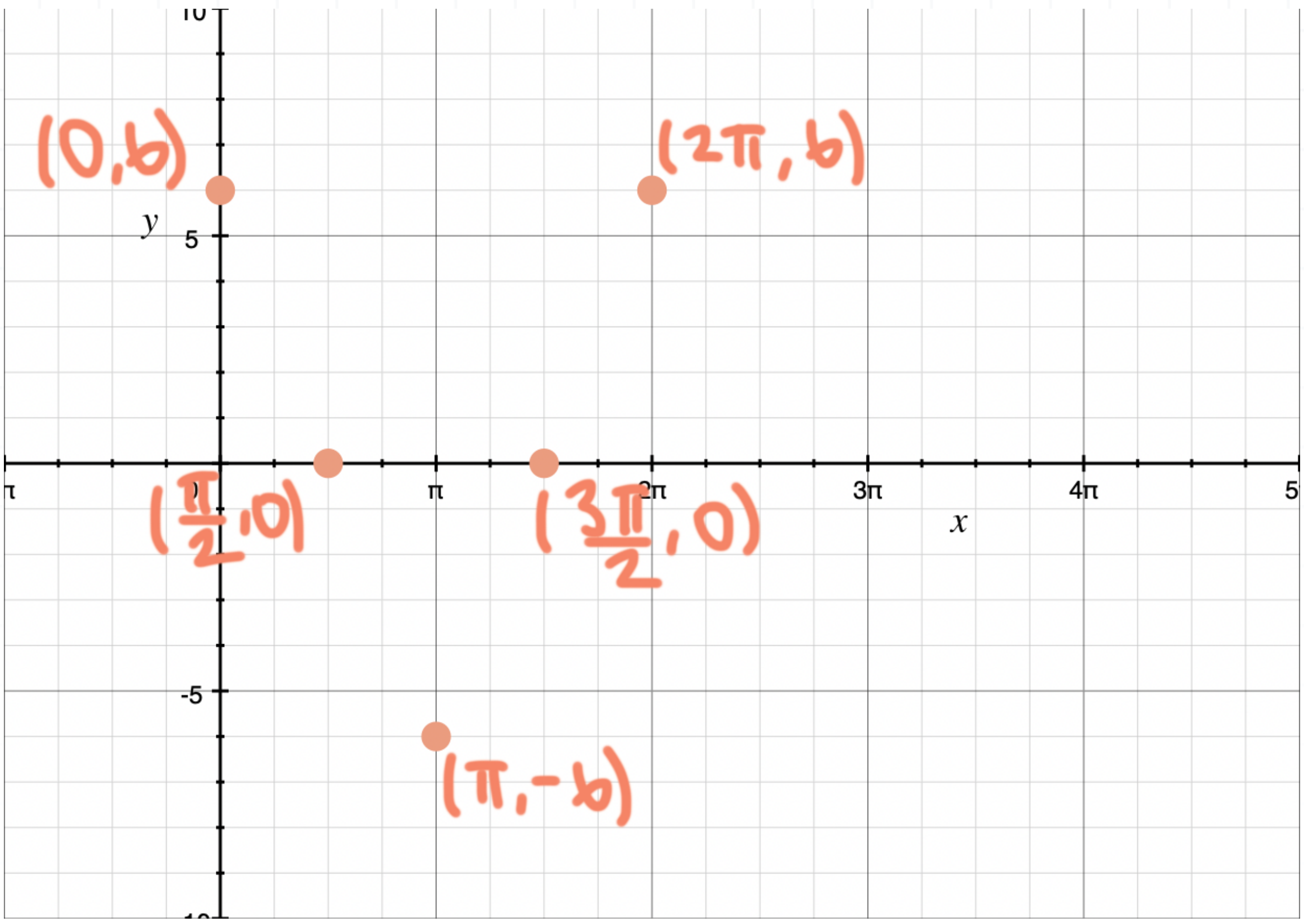
$\theta = 2\pi$ $\qquad$ $r = 6\cos(2\pi) = 6(1) = 6$

... $\qquad$ ...

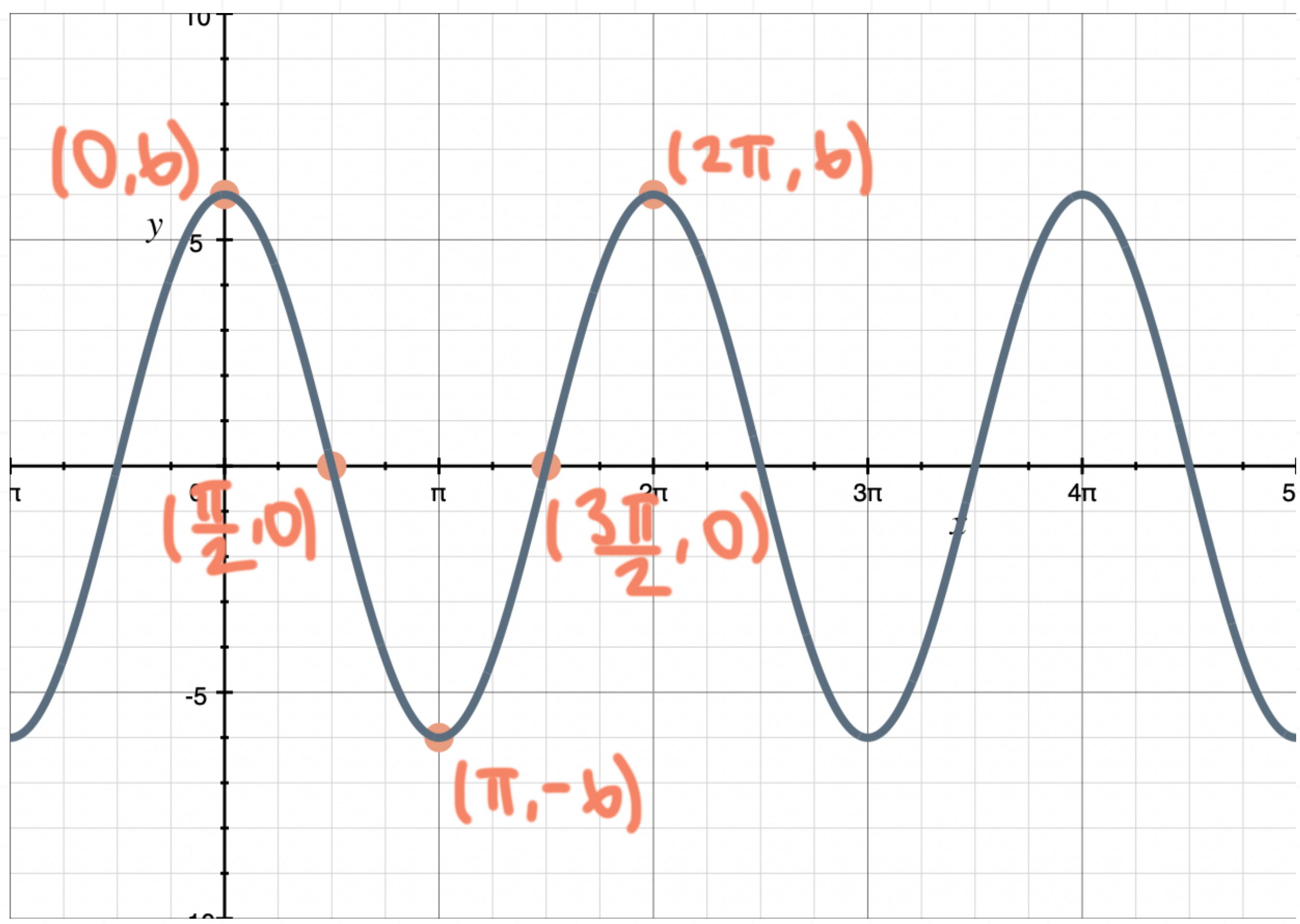
If we summarize these points in a table, we get

theta	0	π/2	π	3π/2	2π
r	6	0	-6	0	6

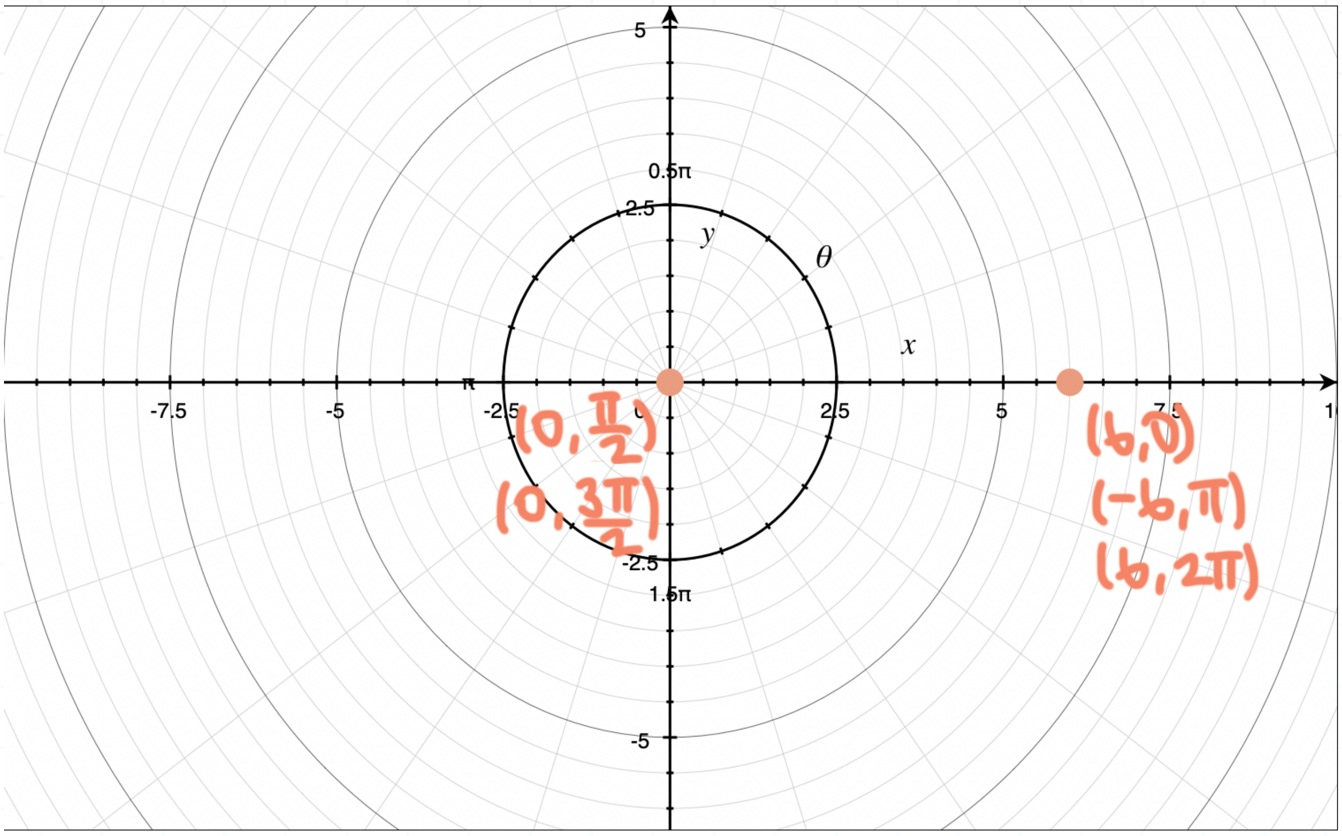
Now we'll plot them in the rectangular coordinate system.



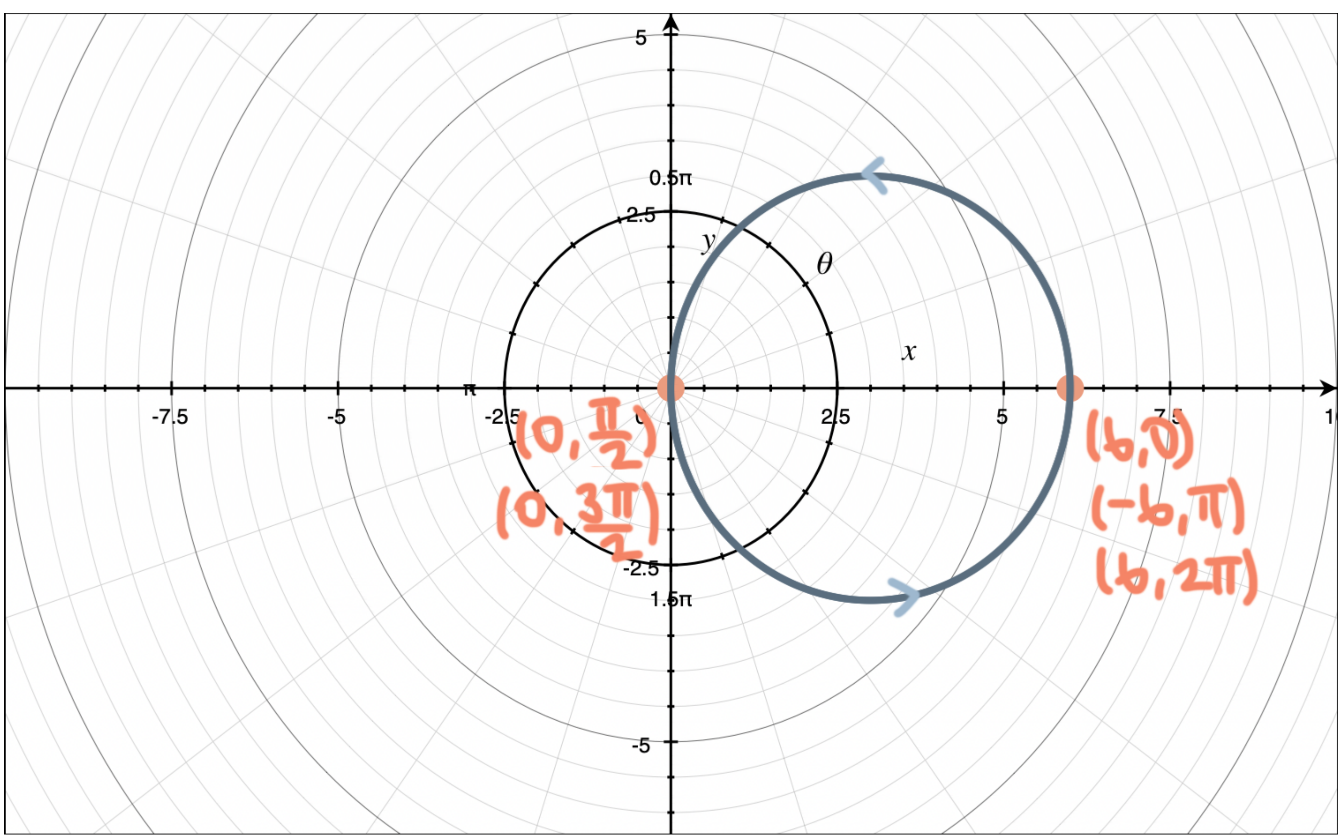
And if we connect these points with a smooth curve, we get



Every time this curve crosses the horizontal axis corresponds to a point at which the polar curve will pass through the pole in the polar coordinate system. If we transfer the coordinate points from the rectangular system into the polar system, in order, starting with $(r, \theta) = (6,0)$ and working our way up toward $(r, \theta) = (6,2\pi)$, the points get plotted as



And if we connect these points with a smooth curve, in order, we see the graph of the circle. We start at $(6,0)$, loop up around to the origin at $(0,\pi/2)$, then loop back down around to $(-6,\pi)$, which is actually the same point as $(6,0)$. From then on, we're retracing the same pieces of the circle over and over.



Let's do another example where we work through this same process.

Example

Sketch the graph of $r = 2.5\cos(3\theta)$.

The trigonometric function in this polar equation is $\cos(3\theta)$, and its argument (the angle at which cosine is evaluated) is 3θ. So we'll set

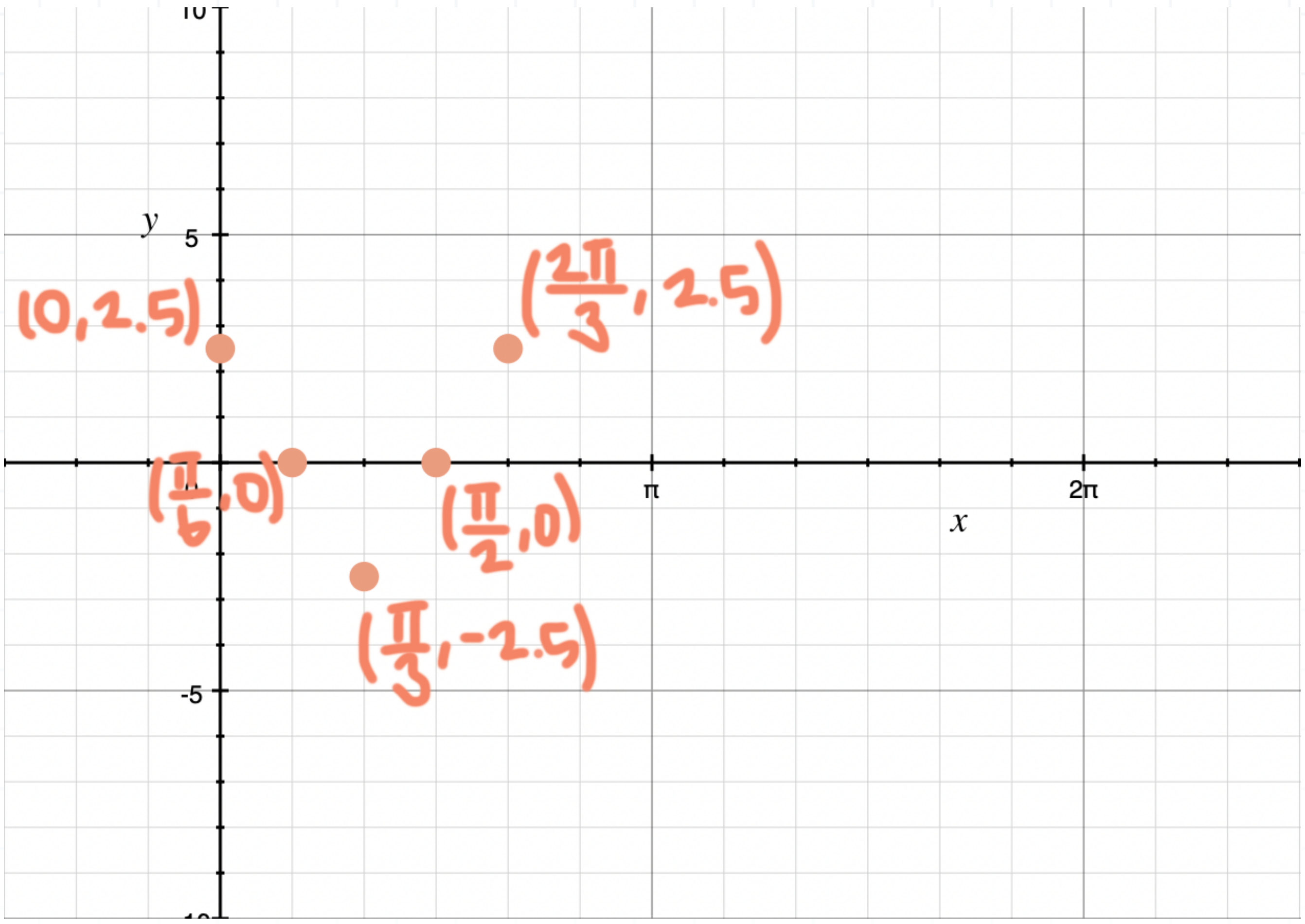
$$3\theta = \frac{\pi}{2}$$

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

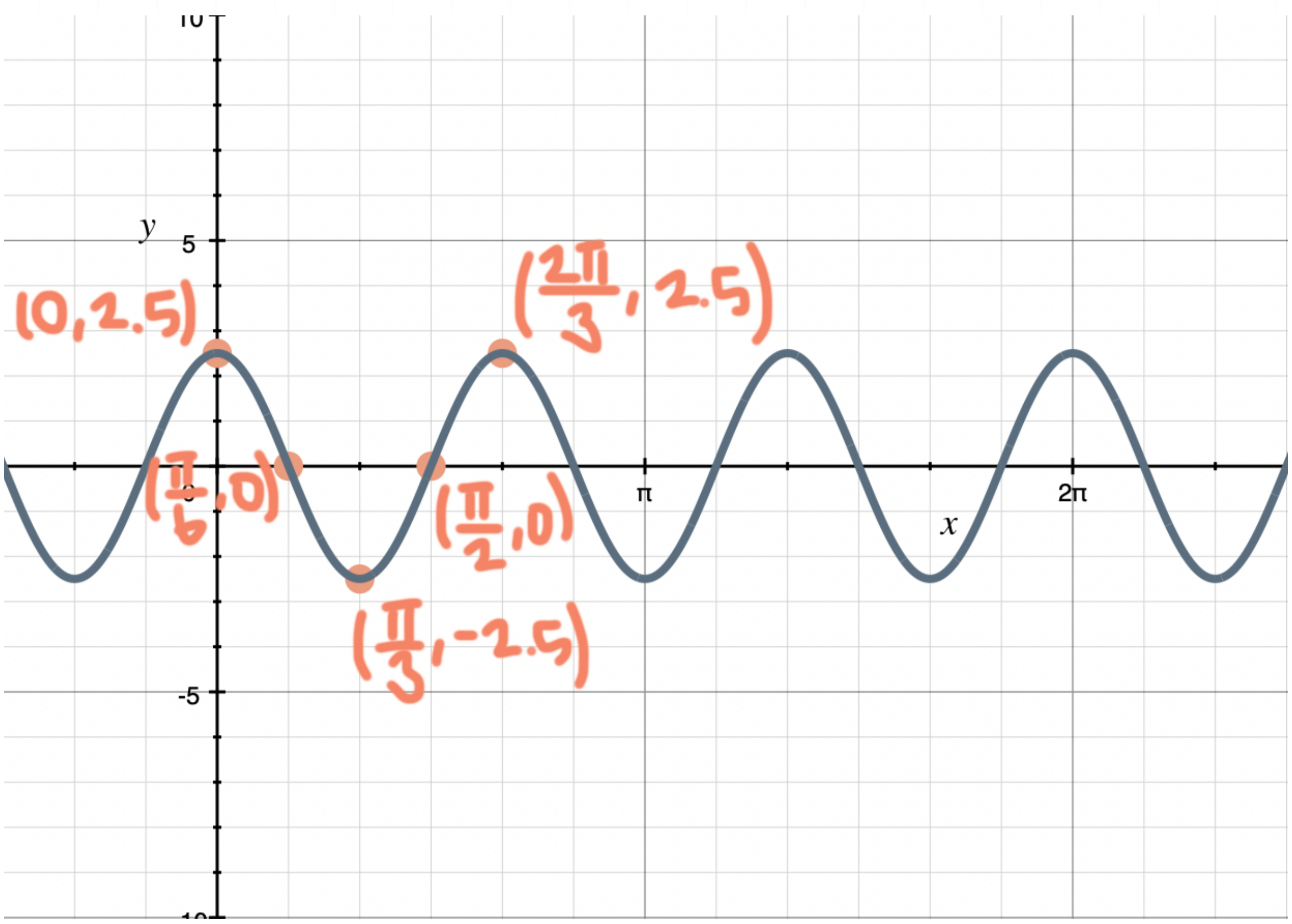
If we evaluate $r = 2.5\cos(3\theta)$ at multiples of $\pi/6$, like $\theta = 0$, $\pi/6$, $\pi/3$, $\pi/2$, $2\pi/3$, etc., then we can make a table of polar coordinate points that satisfy the equation.

theta	0	π/6	π/3	π/2	2π/3
r	2.5	0	-2.5	0	2.5

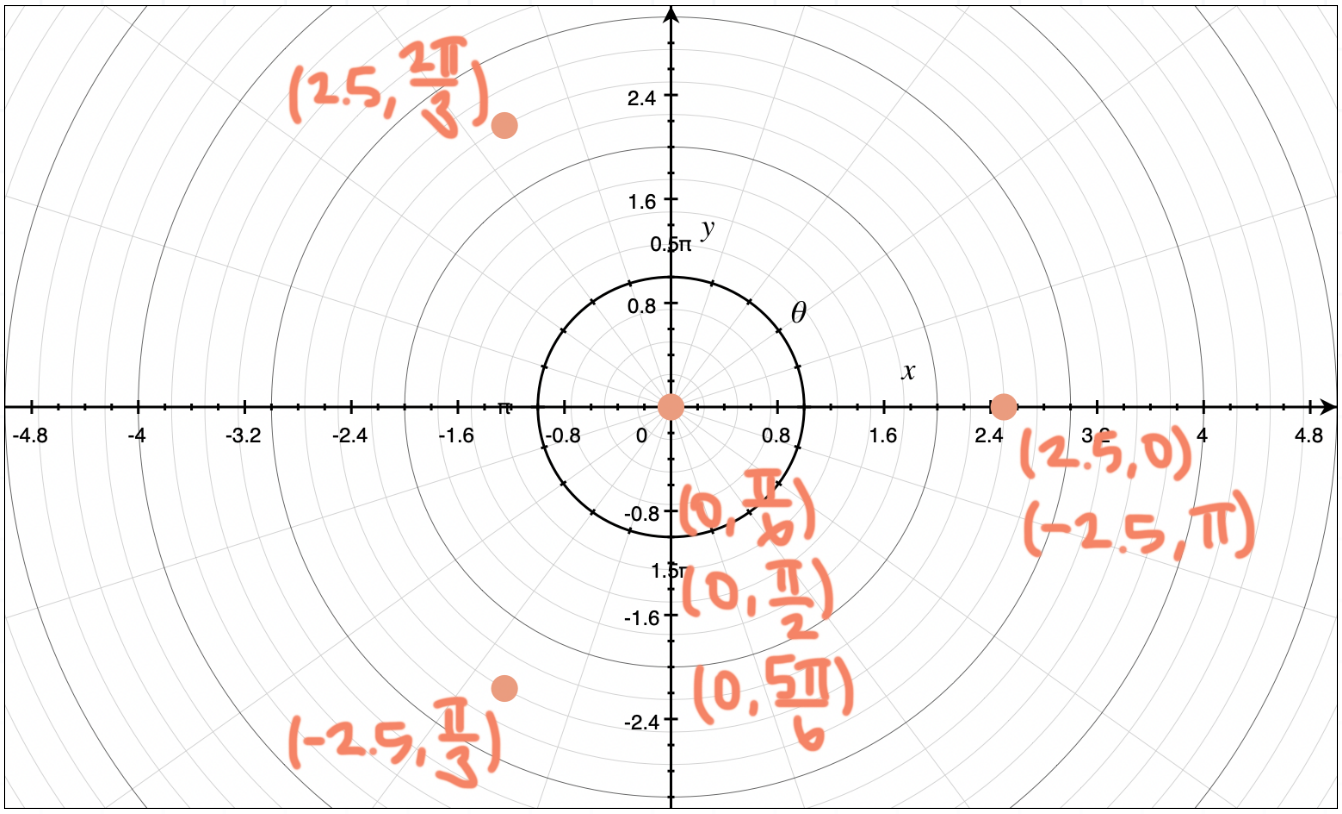
Plotting these points in the rectangular coordinate system gives



And if we connect these points with a smooth curve, we get



If we transfer the points from the rectangular system into the polar system, we get



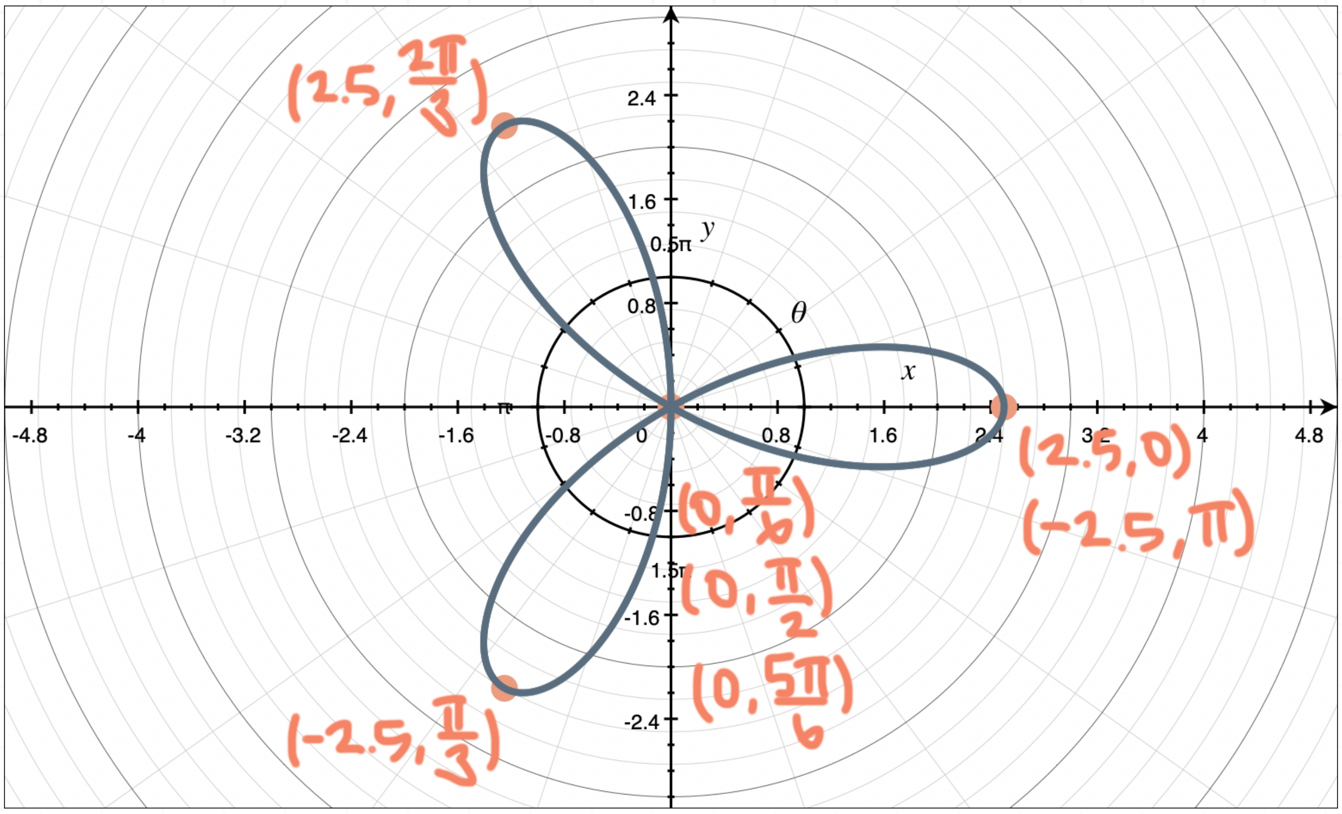
And if we then connect the points, in order, we see the graph. We start at $(2.5,0)$, and then

loop back to the pole at $(0,\pi/6)$ then out to $(-2.5,\pi/3)$,

loop back to the pole at $(0,\pi/2)$ then out to $(2.5,2\pi/3)$,

loop back to the pole at $(0,5\pi/6)$ then out to $(-2.5,2\pi)$,

which is actually the same point as $(2.5,0)$. From then on, we're retracing the same pieces of the graph over and over.



Let's do one more example to show one more polar shape.

Example

Sketch the graph of $r = 1.8 - 1.8 \sin \theta$.

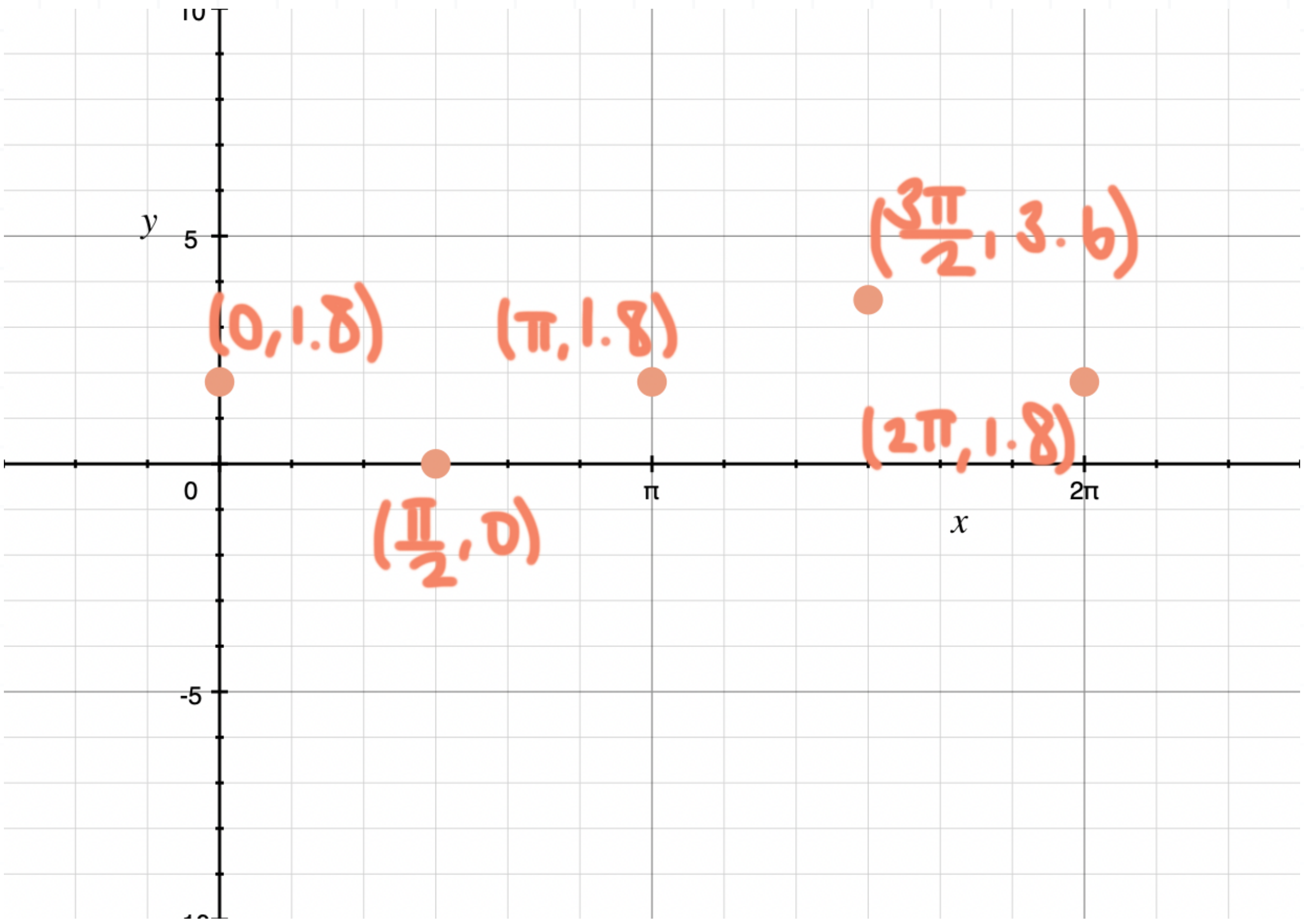
The trigonometric function in this polar equation is $\sin \theta$, and its argument (the angle at which cosine is evaluated) is θ. So we'll set

$$\theta = \frac{\pi}{2}$$

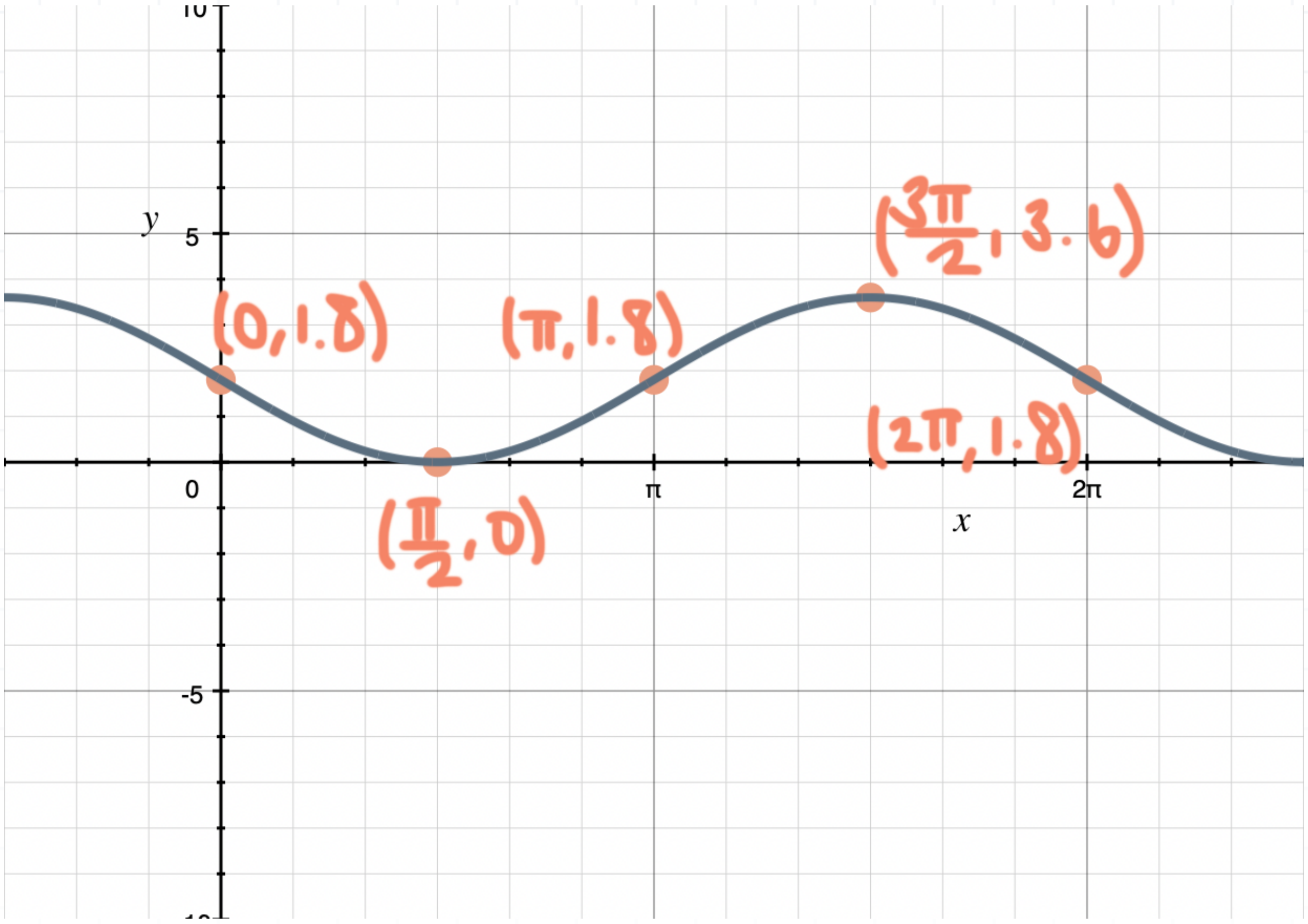
If we evaluate $r = 1.8 - 1.8 \sin \theta$ at multiples of $\pi/2$, like $\theta = 0$, $\pi/2$, π, $3\pi/2$, 2π, etc., then we can make a table of polar coordinate points that satisfy the equation.

theta	0	π/2	π	3π/2	2π
r	1.8	0	1.8	3.6	1.8

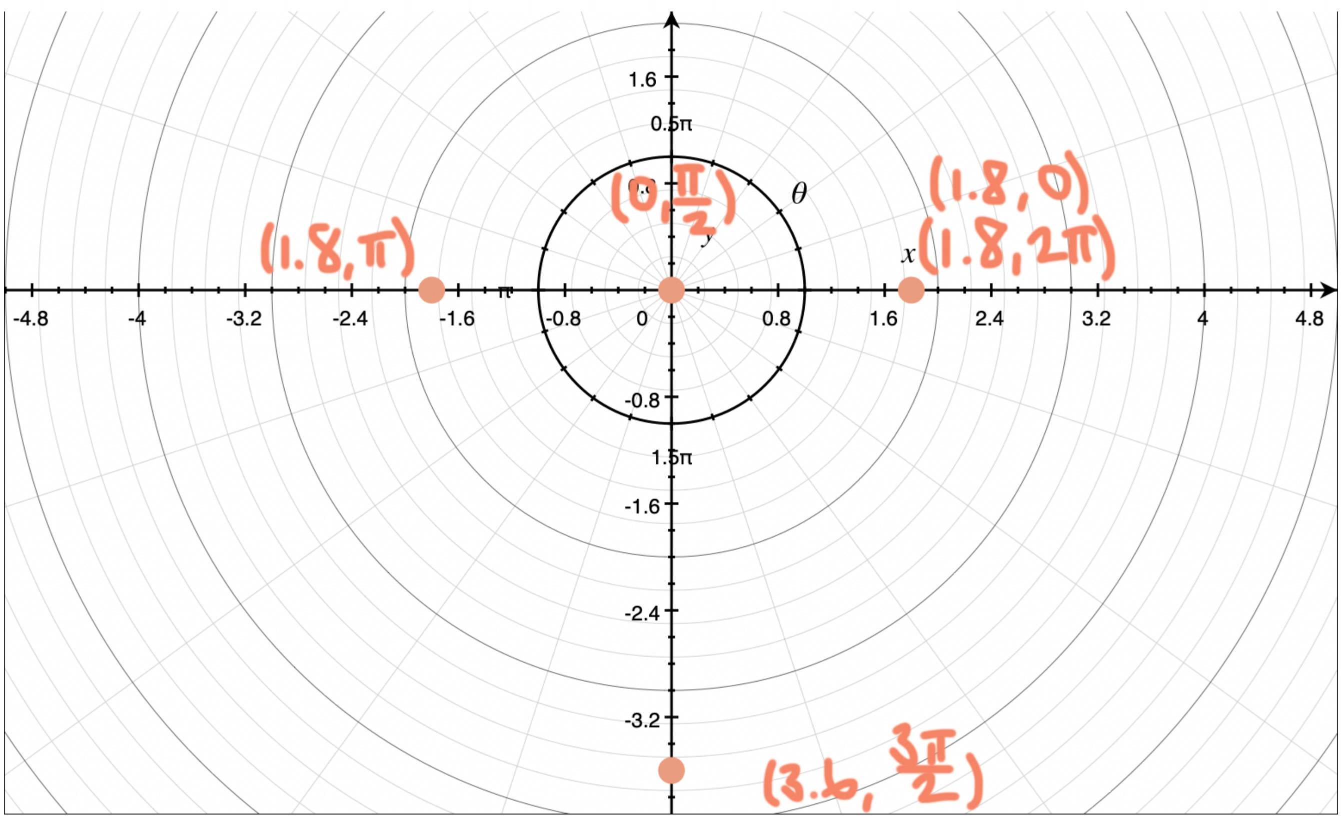
Plotting these points in the rectangular coordinate system gives



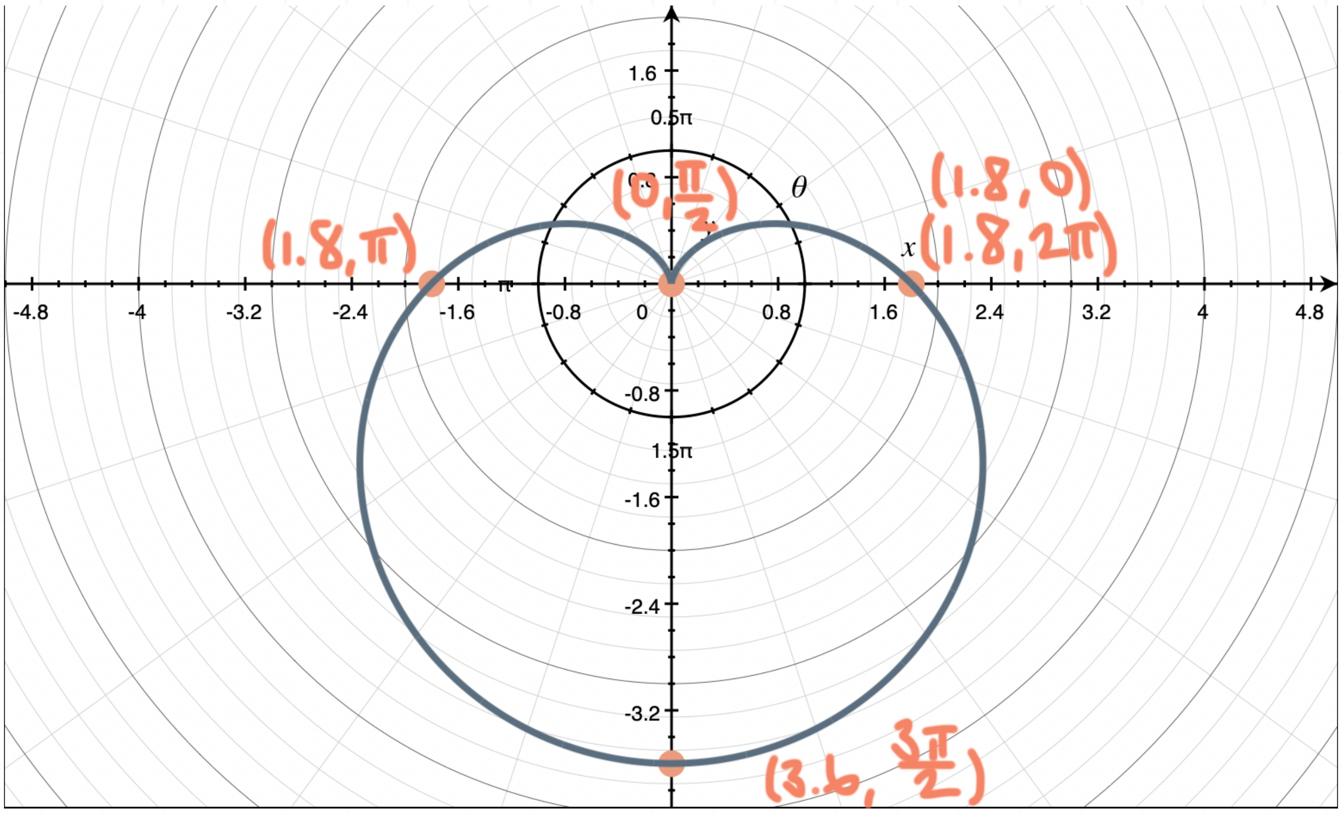
And if we connect these points with a smooth curve, we get



If we transfer the points from the rectangular system into the polar system, we get



And if we then connect the points, in order, we see the graph of the polar curve. We start at $(1.8,0)$, loop up around to the origin at $(0,\pi/2)$, then loop down around to $(1.8,\pi)$, loop down to $(3.6,3\pi/2)$, and then back to $(1.8,2\pi)$, we arrive back at $(1.8,0)$. From then on, we're retracing the same pieces of the polar curve over and over.



Graphing circles

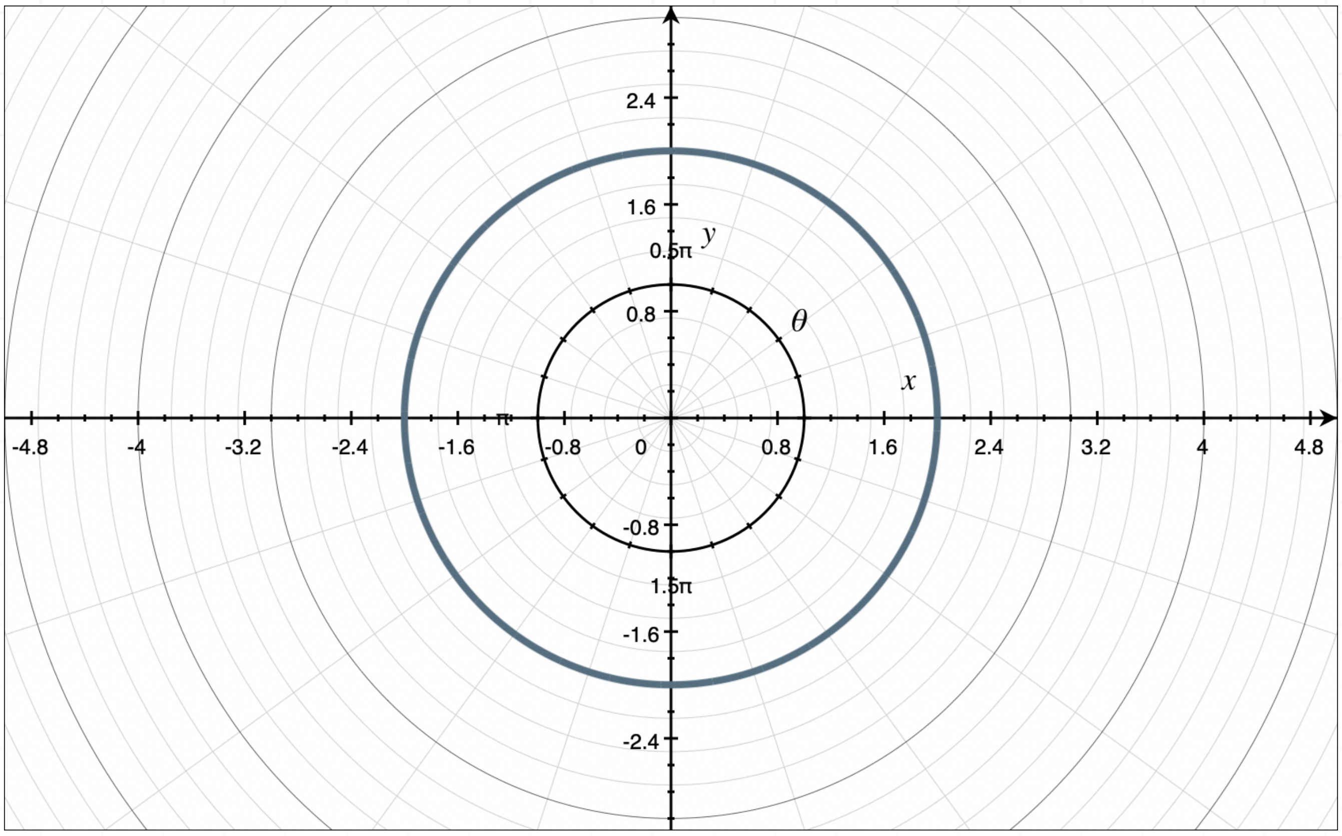
In the previous lesson, we talked about how to sketch polar curves in the rectangular coordinate system, and then transfer those graphs from the rectangular system into the polar system.

We'll continue to use this technique when it's helpful, but now we want to talk in more detail about sketching polar curves directly into the polar coordinate system. At the same time, we'll investigate polar curves of all different shapes, starting with circles.

Circles centered at the origin

We'll start by looking at the graphs of circles. In polar coordinates, the equation of a circle centered at the origin has the form $r = a$, where a is the radius of the circle. We can tell that $r = a$ represents a circle because it shows us that, regardless of the value of θ, the distance from the origin will always be the same.

For instance, $r = 2$ says that, regardless of the value of θ, the distance from the origin will always be 2, so its graph is



The smaller the value of a, the smaller the circle will be, and the larger the value of a, the larger the circle will be. The equation $r = 0$ represents a "circle" with radius 0, which is just the single point at the pole, $(r, \theta) = (0,0)$.

Realize that the value of r can also be negative, but changing r's value from positive to negative doesn't change the graph. If we're graphing $r = -2$, then for every angle θ, we plot a point on the opposite side of the pole from θ, which will still just give us the same circle with radius 2.

Circles centered away from the pole

Circles that are centered away from the pole, but still on the horizontal or vertical axis, have equations that take the form

$$r = c \sin \theta$$

$$r = c \cos \theta$$

where c is a non-zero constant. Circles in this form intersect the pole and extend out to a distance of $r = |c|$ from the pole. Let's work through an example with a cosine function so that we can see how to sketch circles in the polar coordinate system.

Example

Sketch the graph of $r = -7\cos\theta$.

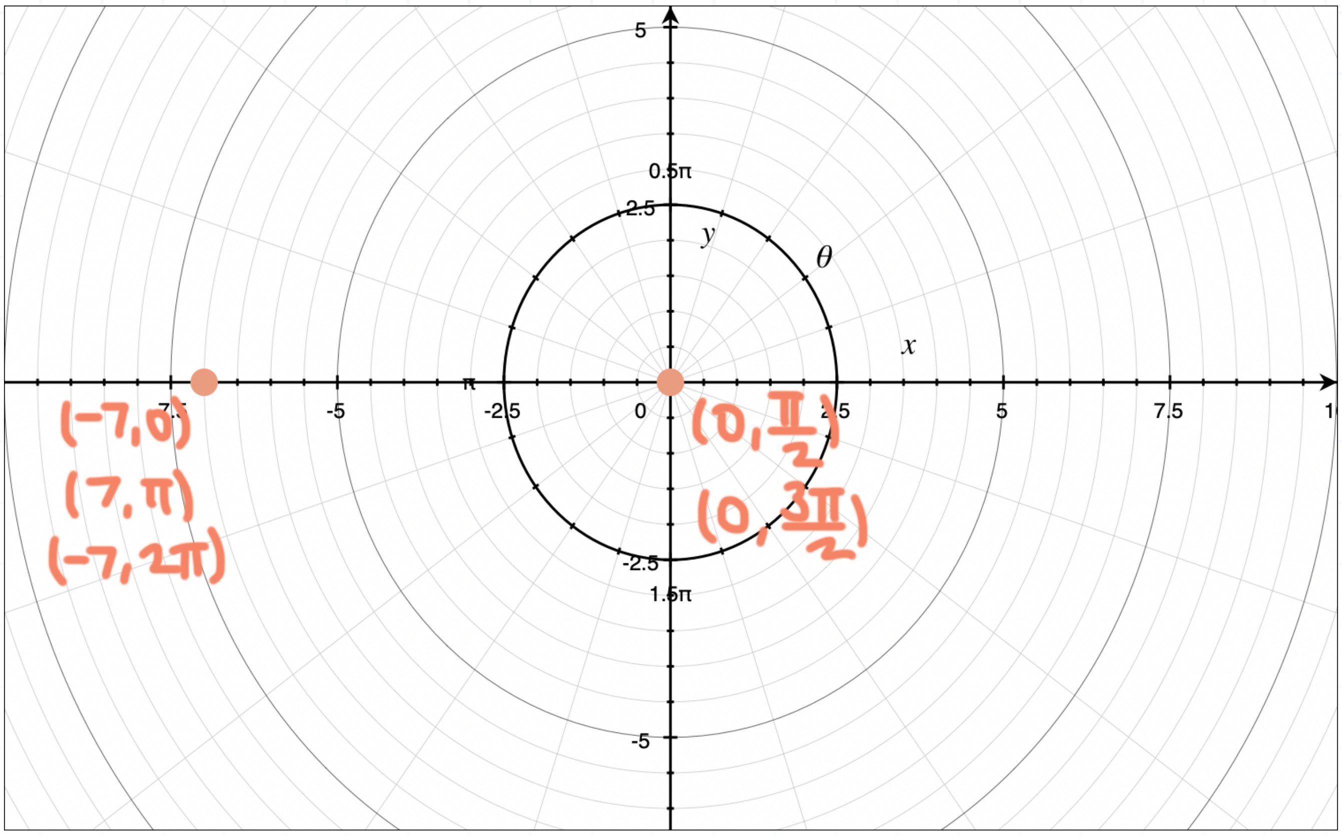
The trigonometric function in this polar equation is $\cos\theta$, and its argument is θ. So we'll set

$$\theta = \frac{\pi}{2}$$

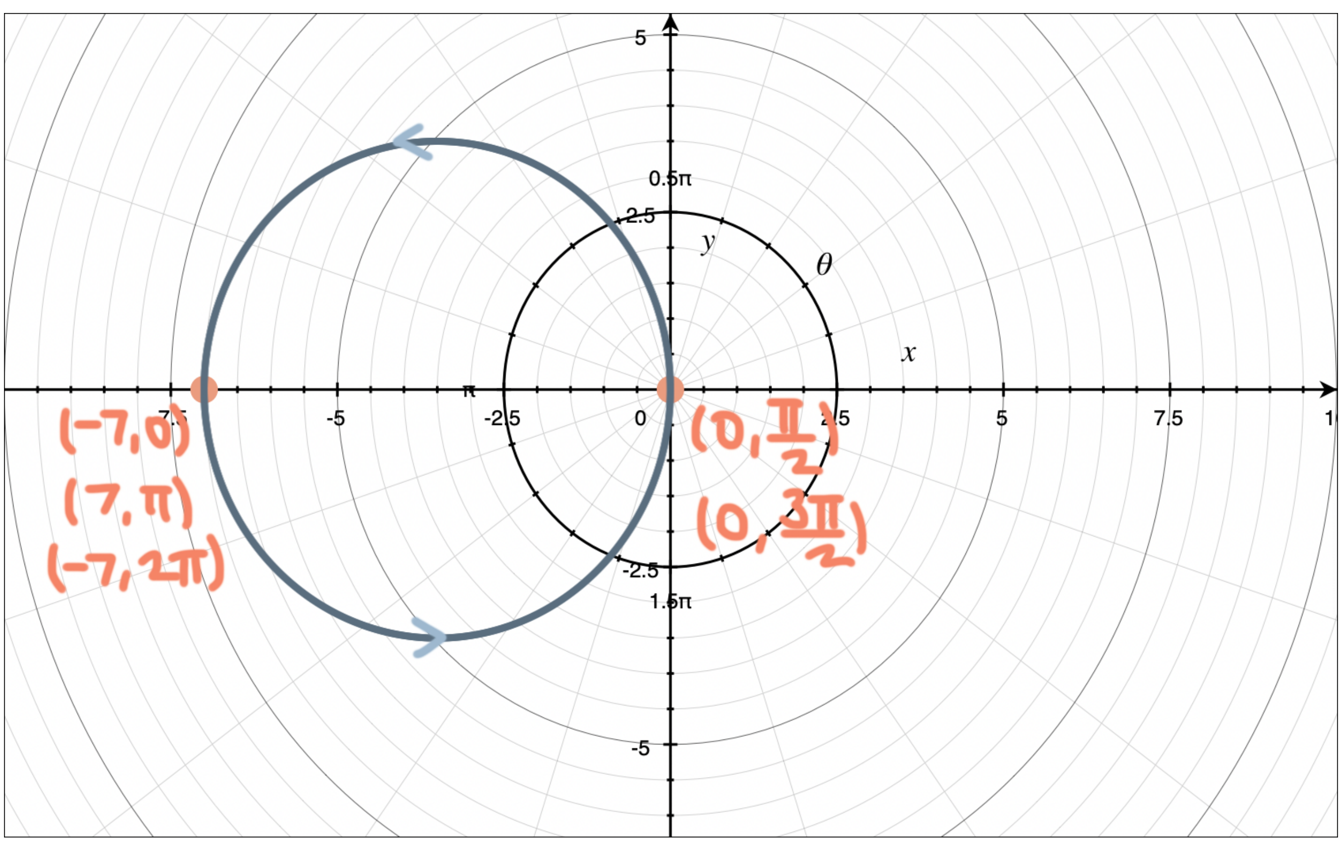
Now we'll make a table with multiples of $\pi/2$, like $\theta = 0$, $\pi/2$, π, $3\pi/2$, 2π, etc., and include the values of r that correspond to each of these θ-values.

theta	0	π/2	π	3π/2	2π
r	-7	0	7	0	-7

Plotting these points on the polar graph gives



And if we connect these points with a smooth curve, in order, we see the graph of the circle. We start at $(-7,0)$, loop down around to the pole at $(0,\pi/2)$, then loop back up around to $(7,\pi)$, which is actually the same point as $(-7,0)$. From then on, we're retracing the same pieces of the circle over and over.



Now let's look at an example with a sine function.

Example

Sketch the graph of $r = -10\sin\theta$.

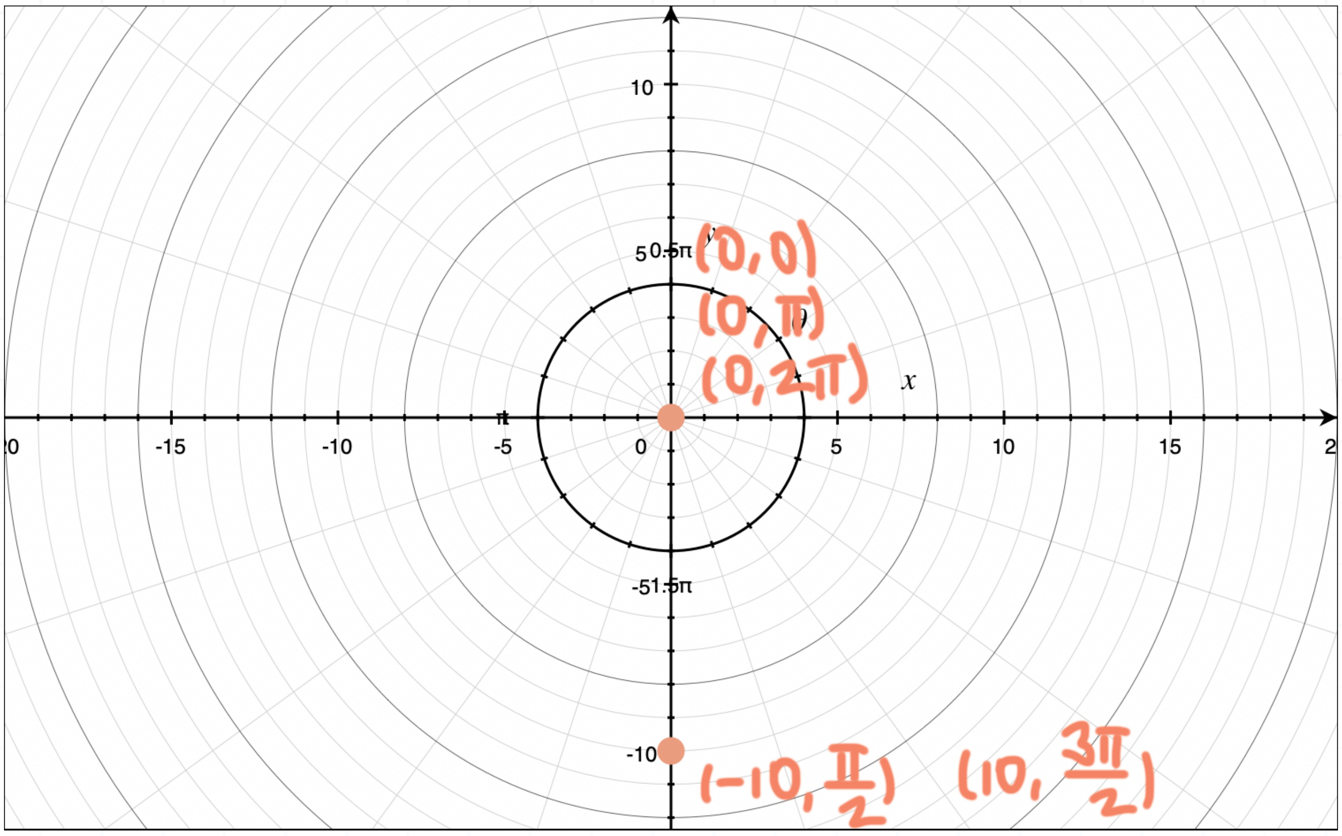
The trigonometric function in this polar equation is $\sin\theta$, and its argument is θ. So we'll set

$$\theta = \frac{\pi}{2}$$

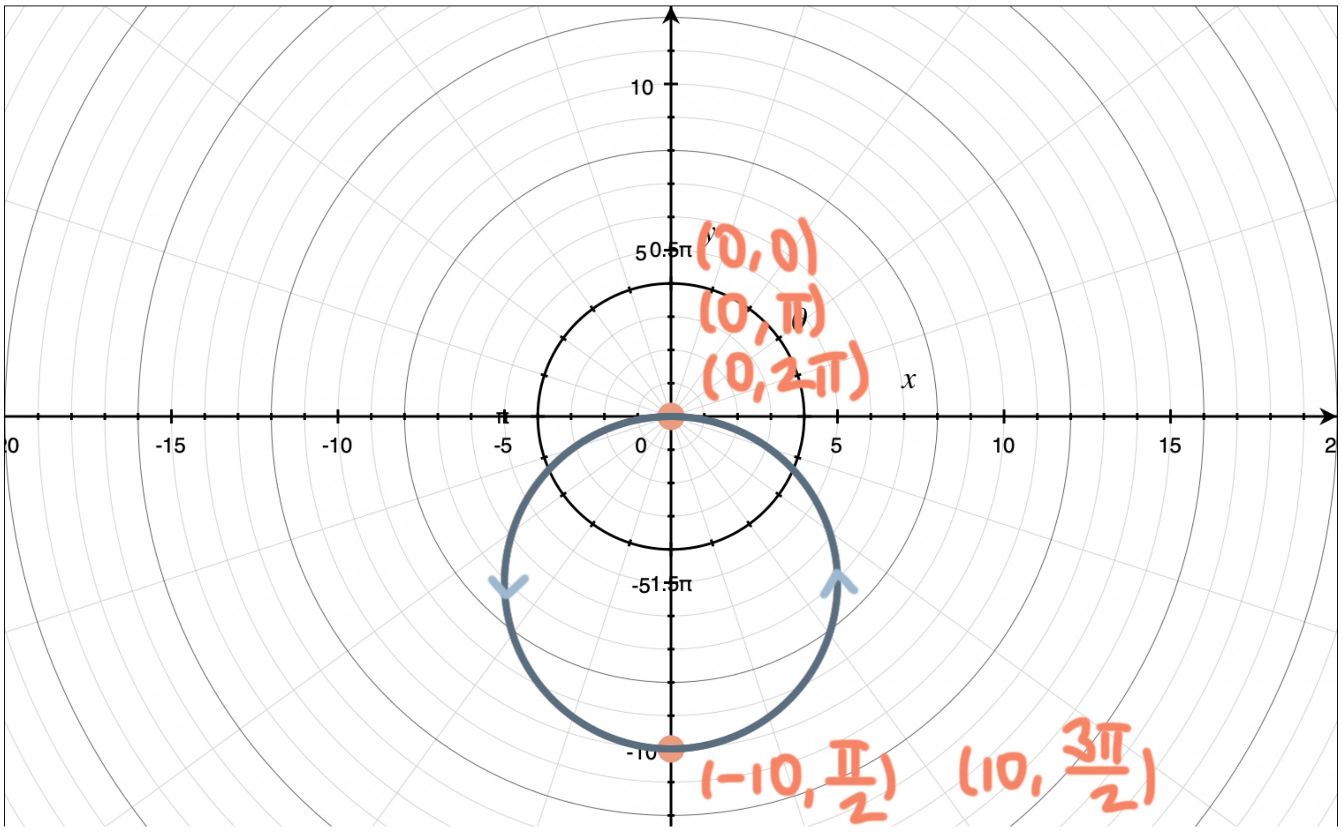
Now we'll make a table with multiples of $\pi/2$, like $\theta = 0$, $\pi/2$, π, $3\pi/2$, 2π, etc., and include the values of r that correspond to each of these θ-values.

theta	0	π/2	π	3π/2	2π
r	0	-10	0	10	0

Plotting these points on the polar graph gives



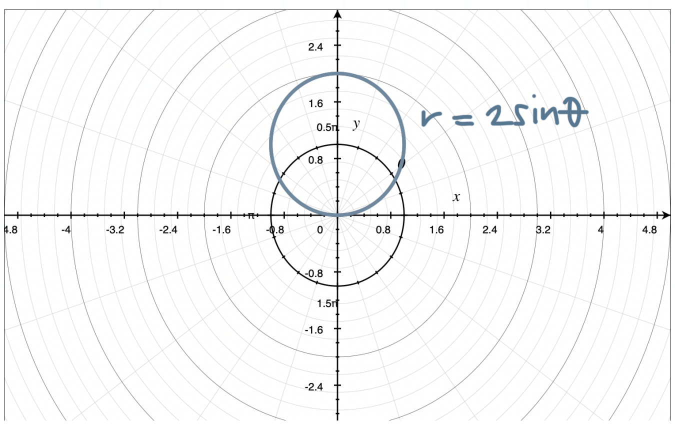
And if we connect these points with a smooth curve, in order, we see the graph of the circle. We start at $(0,0)$, loop down around to $(-10,\pi/2)$, then loop back up around to $(0,\pi)$, which is actually the same point as $(0,0)$. From then on, we're retracing the same pieces of the circle over and over.

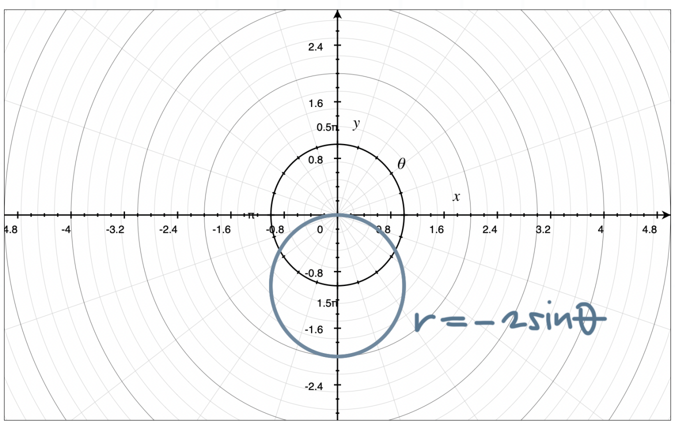


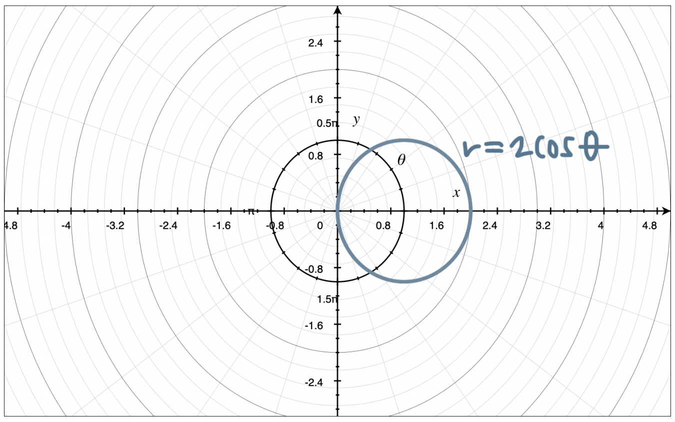
Properties of circles

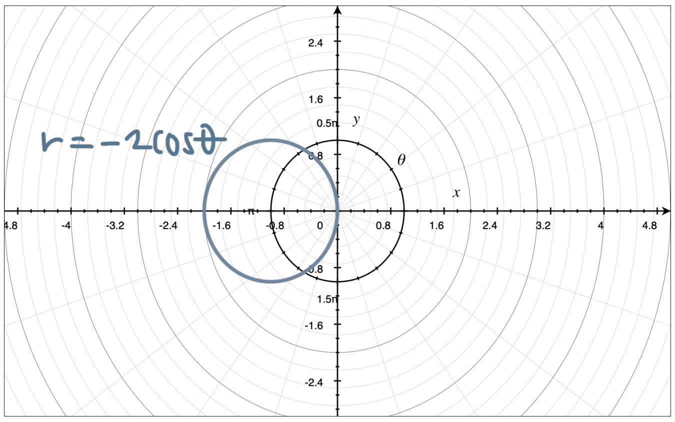
The examples we just worked through should hint at the properties of circles given in the form $r = c\sin\theta$ or $r = c\cos\theta$.

For instance, sine circles are symmetric around the vertical axis, whereas cosine circles are symmetric around the horizontal axis. And for equations with a positive constant $c > 0$, the circle will sit along the positive side of the horizontal or vertical axis, whereas for equations with a negative constant $c < 0$, the circle will sit along the negative side of the horizontal or vertical axis.









Graphing roses

Now we'll look at how to sketch the graphs of polar **roses**, which are polar equations in either of these forms:

$$r = c\cos(n\theta)$$

$$r = c\sin(n\theta)$$

where c is a non-zero constant, and n is an integer greater than 1 (or less than -1).

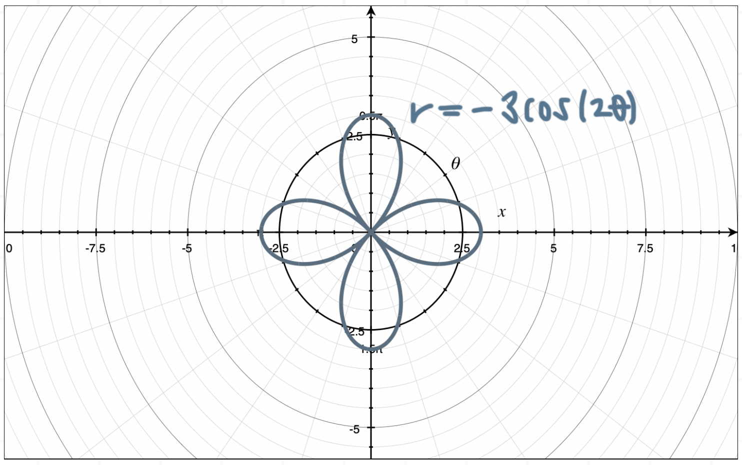
We call these kinds of curves roses because they look a little bit like flowers, with the "petals" of the flower extending out from the origin.

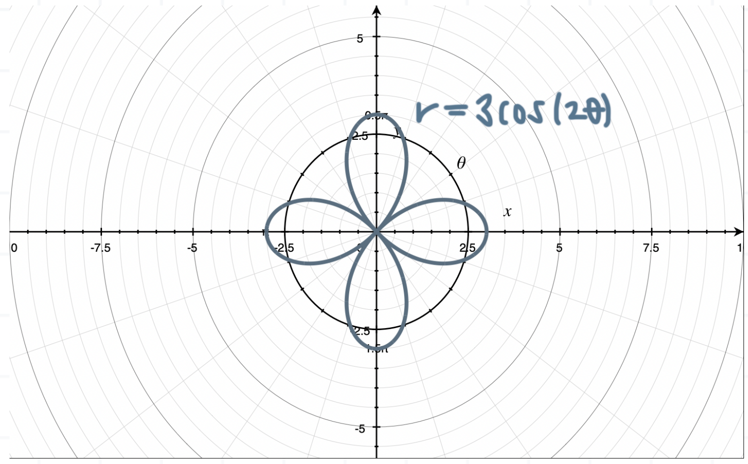
Properties of roses

As we learn to graph these kinds of curves, it's helpful to keep a few properties of roses in mind.

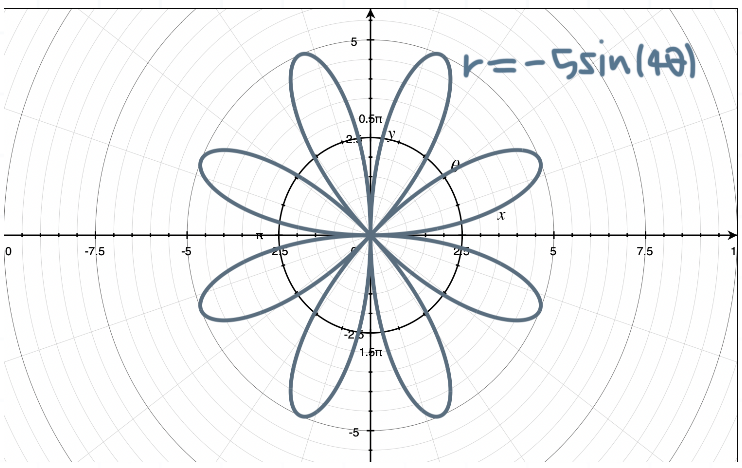
First, the tips of the petals will extend out to a distance of $r = |c|$ away from the pole. So if $c = \pm 3$, for example, then the petals will extend out to a distance of $r = 3$ from the pole.

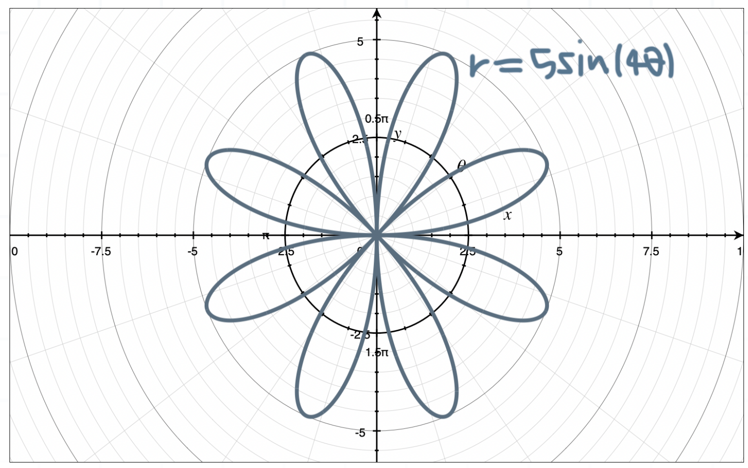
If n is even, the graph of the rose doesn't change when the sign of c changes, so the graph of $r = -3\cos(2\theta)$ will be identical to $r = 3\cos(2\theta)$,



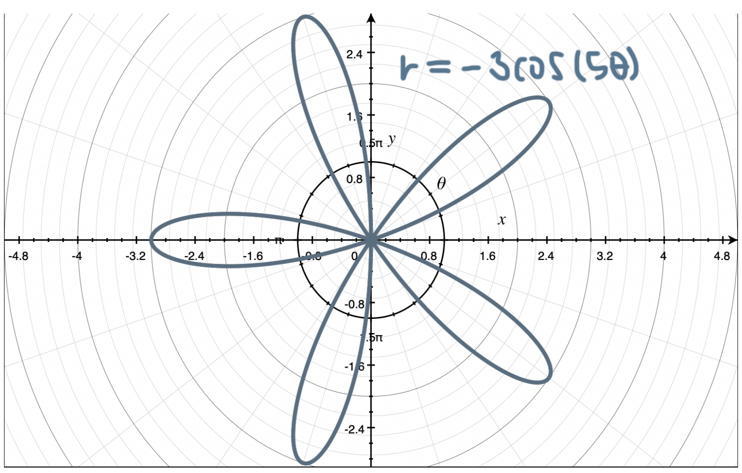


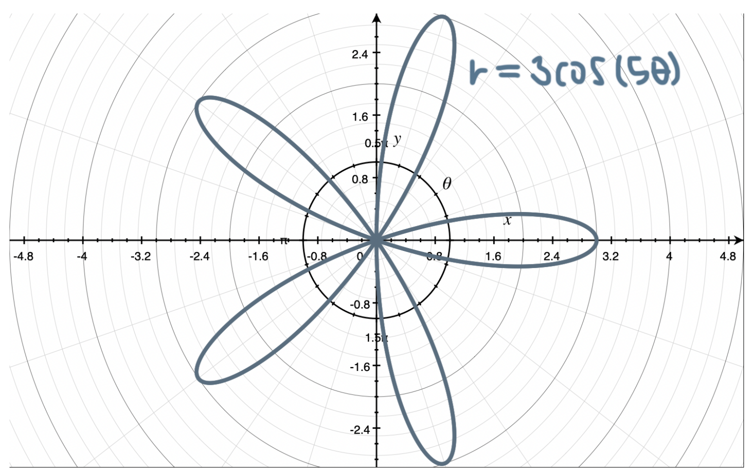
and the graph of $r = -5\sin(4\theta)$ will be identical to $r = 5\sin(4\theta)$.



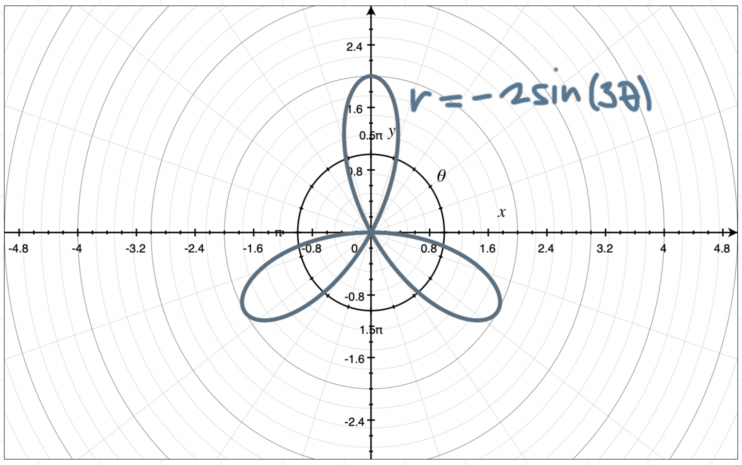


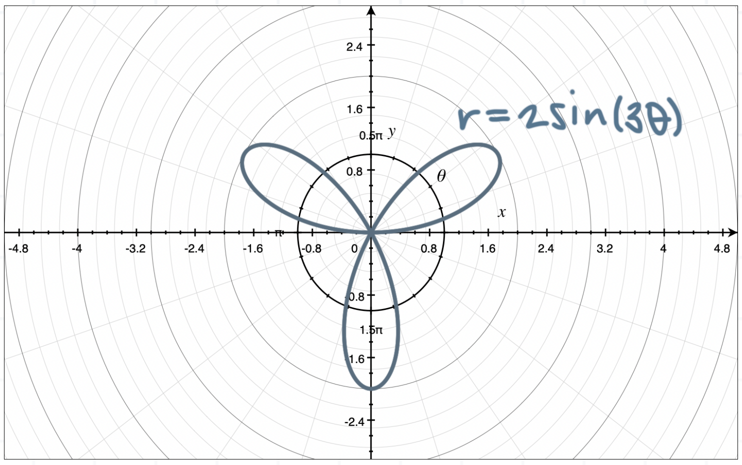
If n is odd, the graph of the rose rotates when the sign of c changes. The graphs of $r = -3\cos(5\theta)$ and $r = 3\cos(5\theta)$ are





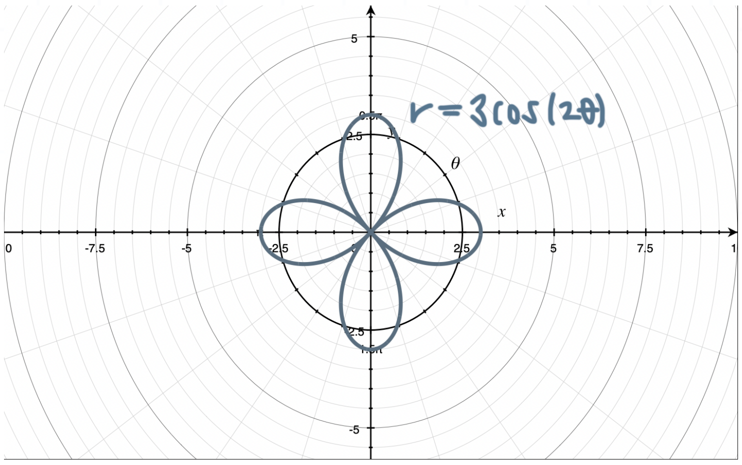
and the graphs of $r = -2\sin(3\theta)$ and $r = 2\sin(3\theta)$ are

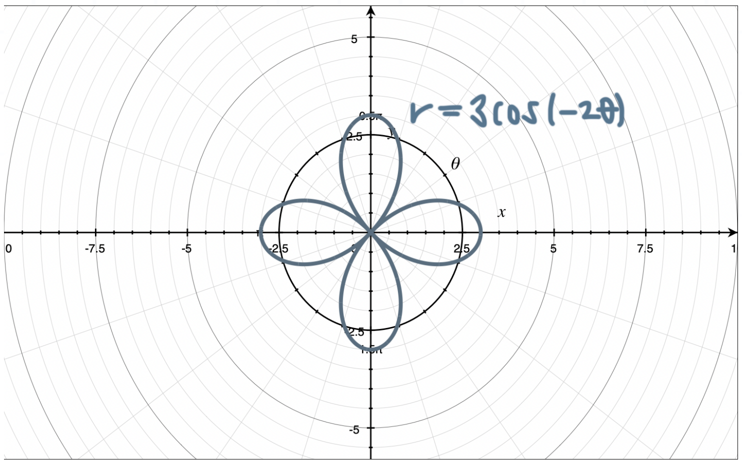




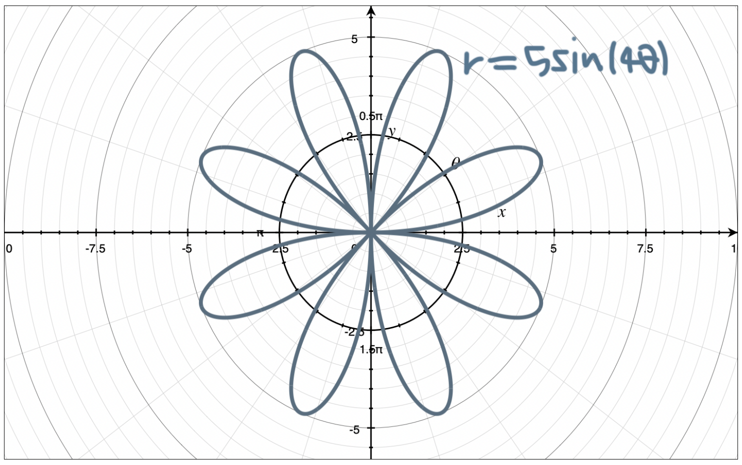
Second, the rose will have $|2n|$ petals when n is an even integer, but $|n|$ petals when n is an odd integer.

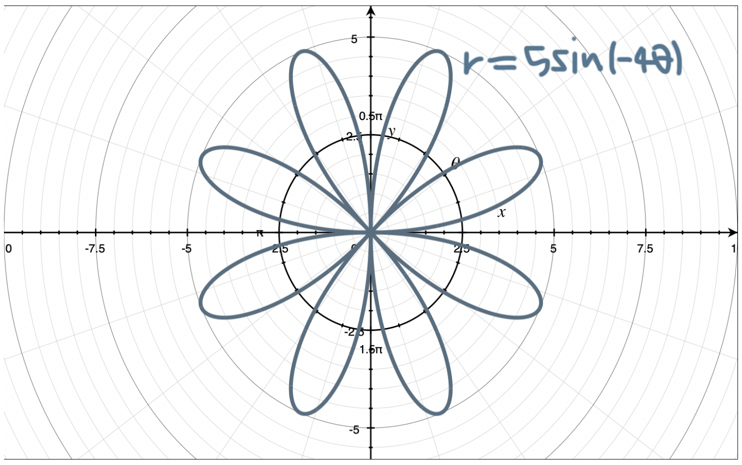
If n is even, the graph of the rose doesn't change when the sign of n changes, so the graphs of $r = 3\cos(2\theta)$ and $r = 3\cos(-2\theta)$ will be identical with $|2n| = |2(2)| = 4$ petals,



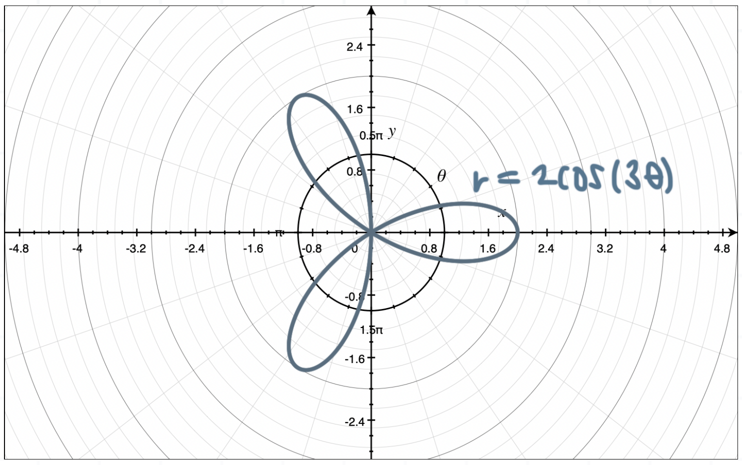


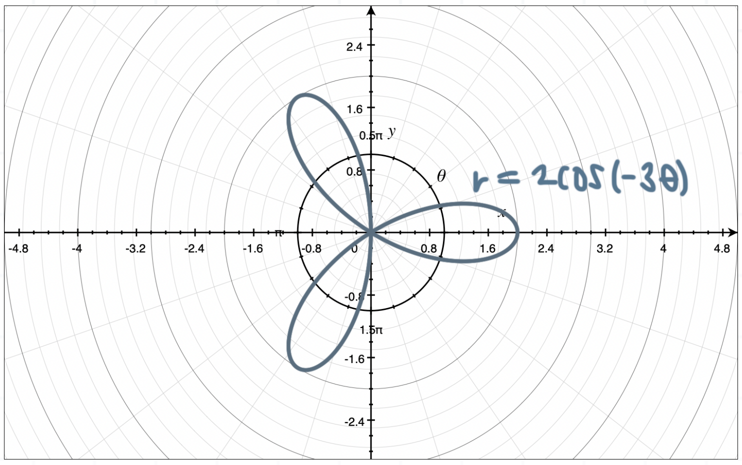
and the graphs of $r = 5\sin(4\theta)$ and $r = 5\sin(-4\theta)$ will be identical with $|2n| = |2(4)| = 8$ petals.



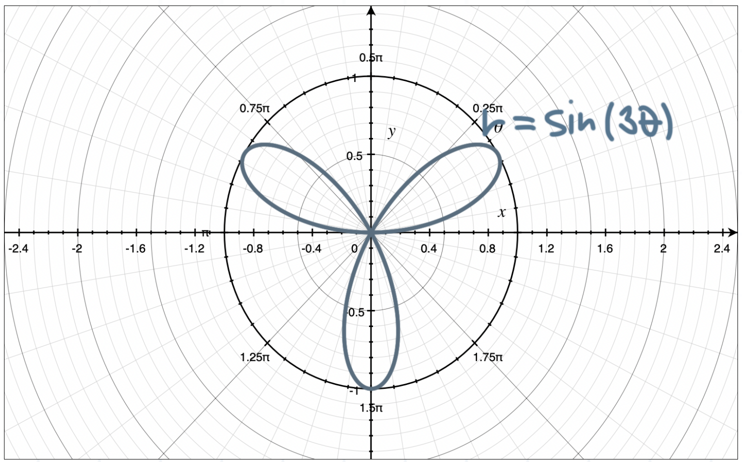


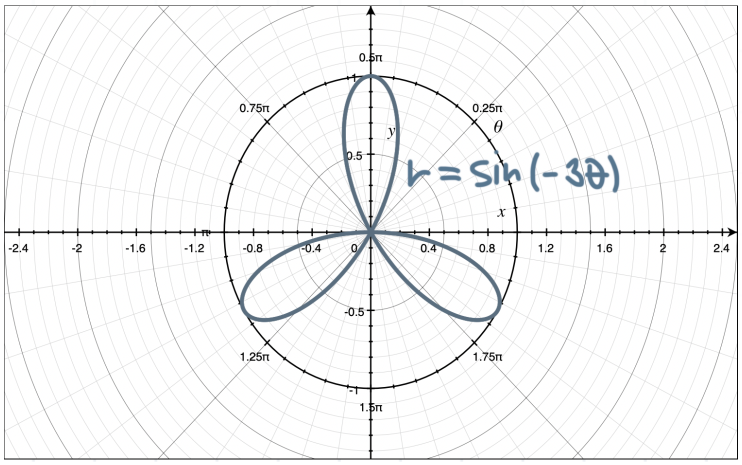
If n is odd, the graphs of cosine roses are identical when the sign of n changes, but the graphs of sine roses rotate. The graphs of $r = 2\cos(3\theta)$ and $r = 2\cos(-3\theta)$ will have $|n| = |3| = 3$ petals,



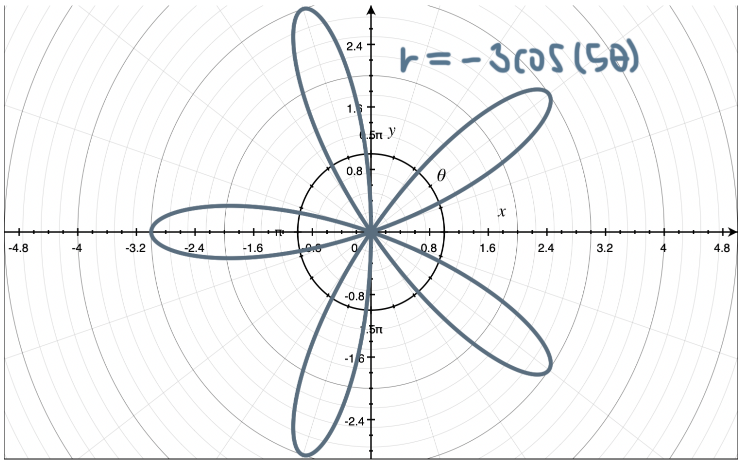


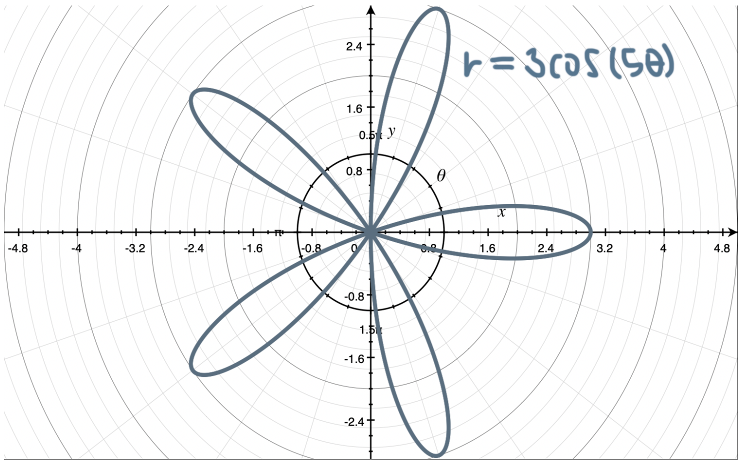
and the graphs of $r = \sin(3\theta)$ and $r = \sin(-3\theta)$ will have $|n| = |3| = 3$ petals.



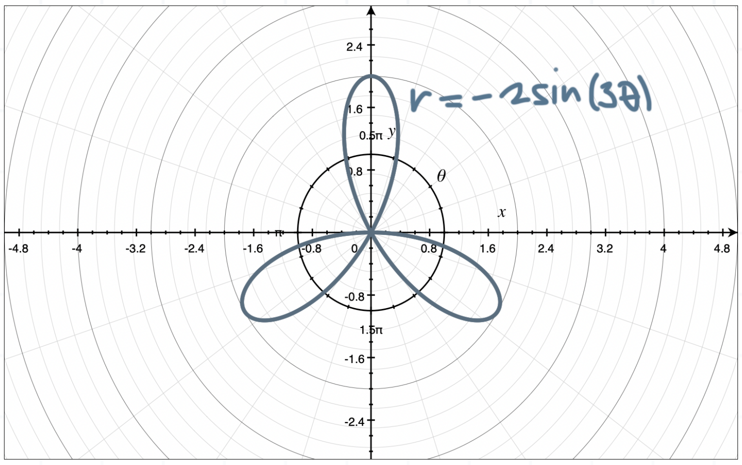


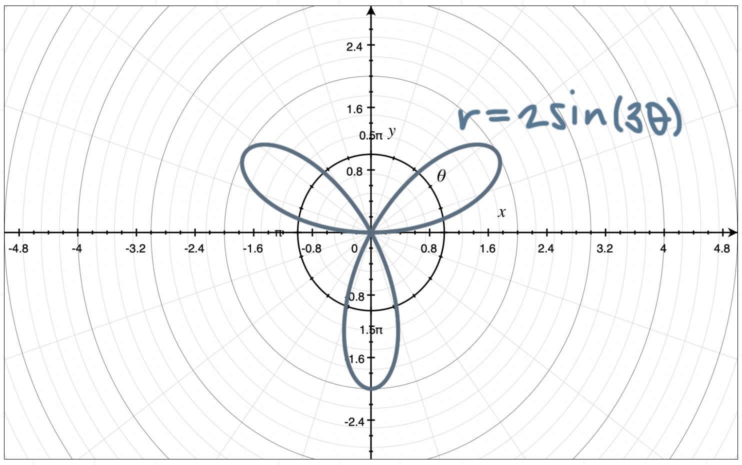
Third, if the rose has one petal that straddles the horizontal axis and none that straddle the vertical axis, it must be a cosine rose,

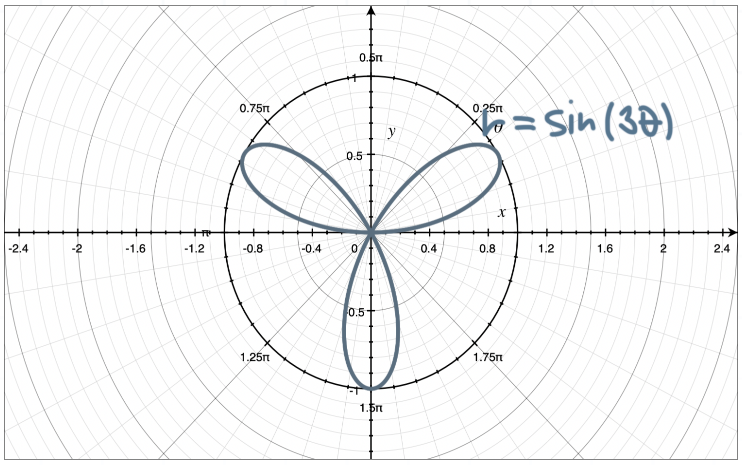


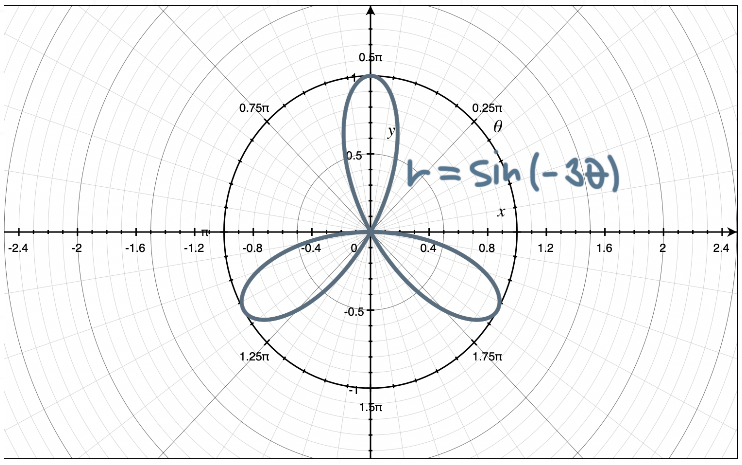


whereas if the rose has one petal that straddles the vertical axis and none that straddle the horizontal axis, it must be a sine rose.









How to sketch roses

We'll use the same approach to sketch roses that we used in the previous lesson for sketching circles:

1. Set the argument of the trigonometric function equal to $\pi/2$, and then solve the equation for θ.

2. Evaluate the polar curve at multiples of the θ-value we solved for in Step 1, starting with $\theta = 0$, and plot the resulting points on the polar graph.

3. Connect the points on the polar graph with a smooth curve.

Let's do an example where we work through these steps in order to sketch a rose in the form $r = c\cos(n\theta)$.

Example

Sketch the graph of $r = 3\cos(2\theta)$.

Because $c = 3$ and $n = 2$, this rose will have $|2n| = |2(2)| = 4$ petals that extend out a distance of 3 from the origin. To sketch the graph, we recognize that the trigonometric function in this polar equation is $\cos(2\theta)$, and its argument is 2θ. So we'll set

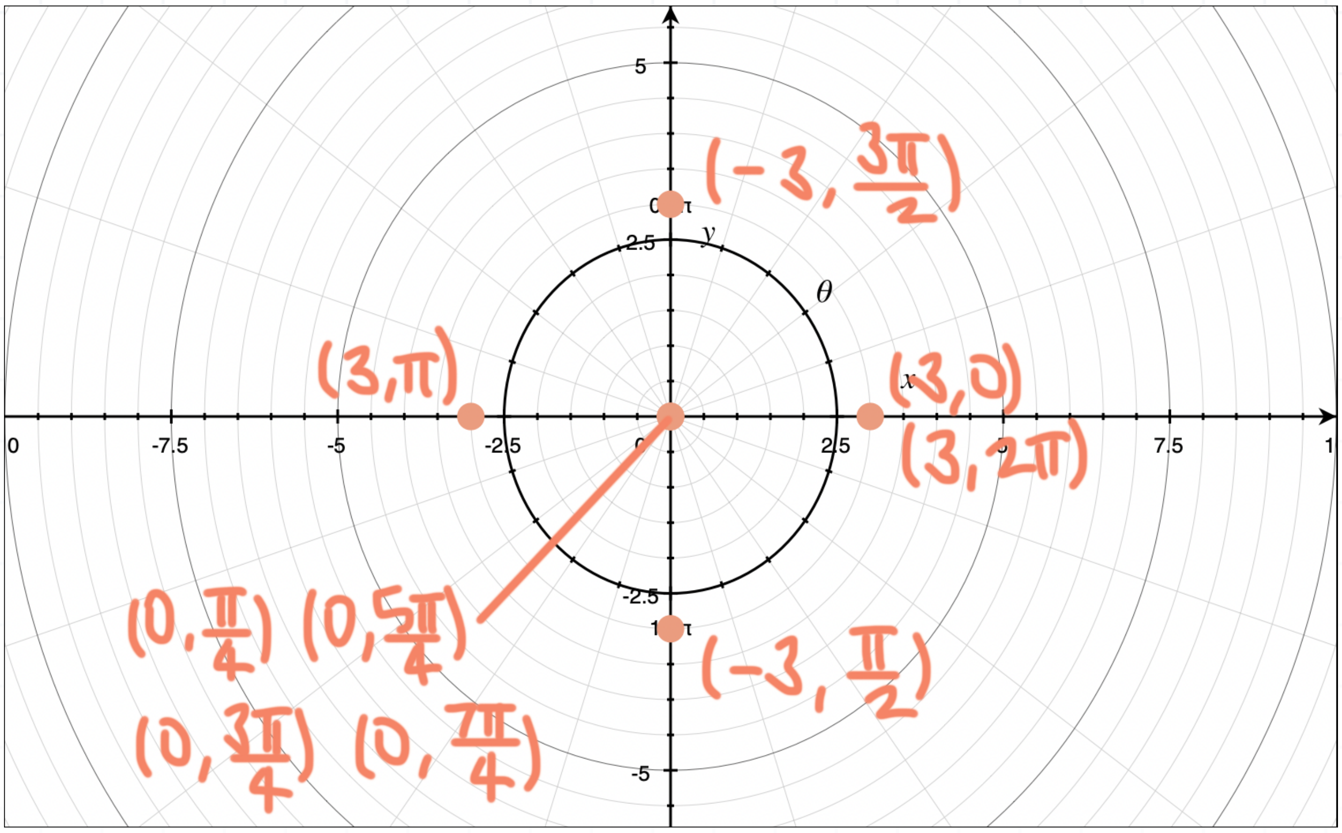
$$2\theta = \frac{\pi}{2}$$

$$\theta = \frac{\pi}{4}$$

Now we'll make a table with multiples of $\pi/4$, like $\theta = 0$, $\pi/4$, $\pi/2$, $3\pi/4$, π, $5\pi/4$, 3π, etc., and include the values of r that correspond to each of these θ -values.

theta	0	π/4	π/2	3π/4	π	5π/4	3π/2	7π/4	2π
r	3	0	-3	0	3	0	-3	0	3

Plotting these points on the polar graph gives



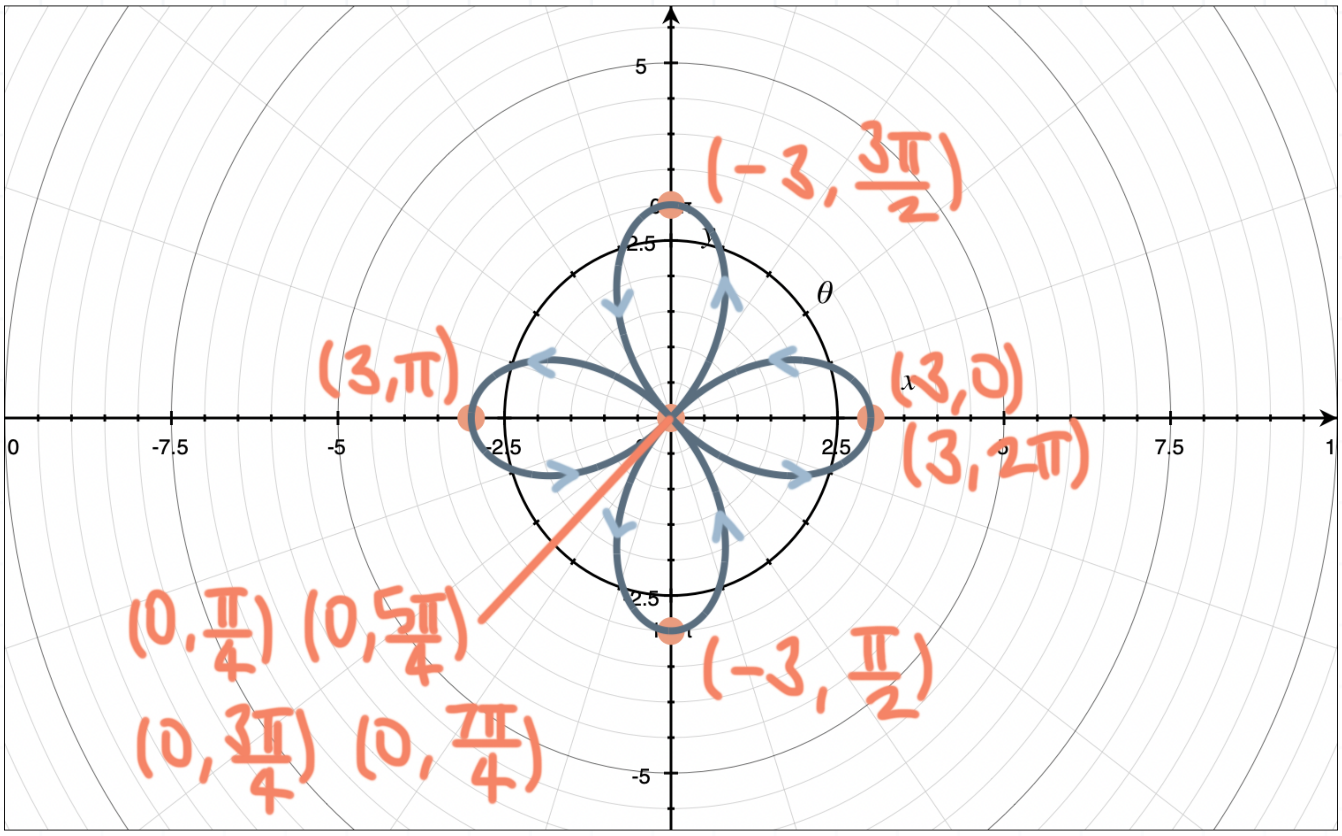
And if we connect these points with a smooth curve, in order, we see the graph of the rose. We start at $(3,0)$, and then loop to the pole at $(0,\pi/4)$,

loop to $(-3,\pi/2)$ then back to the pole at $(0,3\pi/4)$,

loop to $(3,\pi)$ then back to the pole at $(0,5\pi/4)$,

loop to $(-3,3\pi/2)$ then back to the pole at $(0,7\pi/4)$,

then finally loop back to $(3,2\pi)$, which is actually the same point as $(3,0)$. From then on, we're retracing the same pieces of the rose over and over.



Let's do another example with a cosine rose.

Example

Sketch the graph of $r = -5\cos(4\theta)$.

Because $c = -5$ and $n = 4$, this rose will have $|2n| = |2(4)| = 8$ petals that extend out a distance of 5 from the origin. To sketch the graph, we recognize that the trigonometric function in this polar equation is $\cos(4\theta)$, and its argument is 4θ. So we'll set

$$4\theta = \frac{\pi}{2}$$

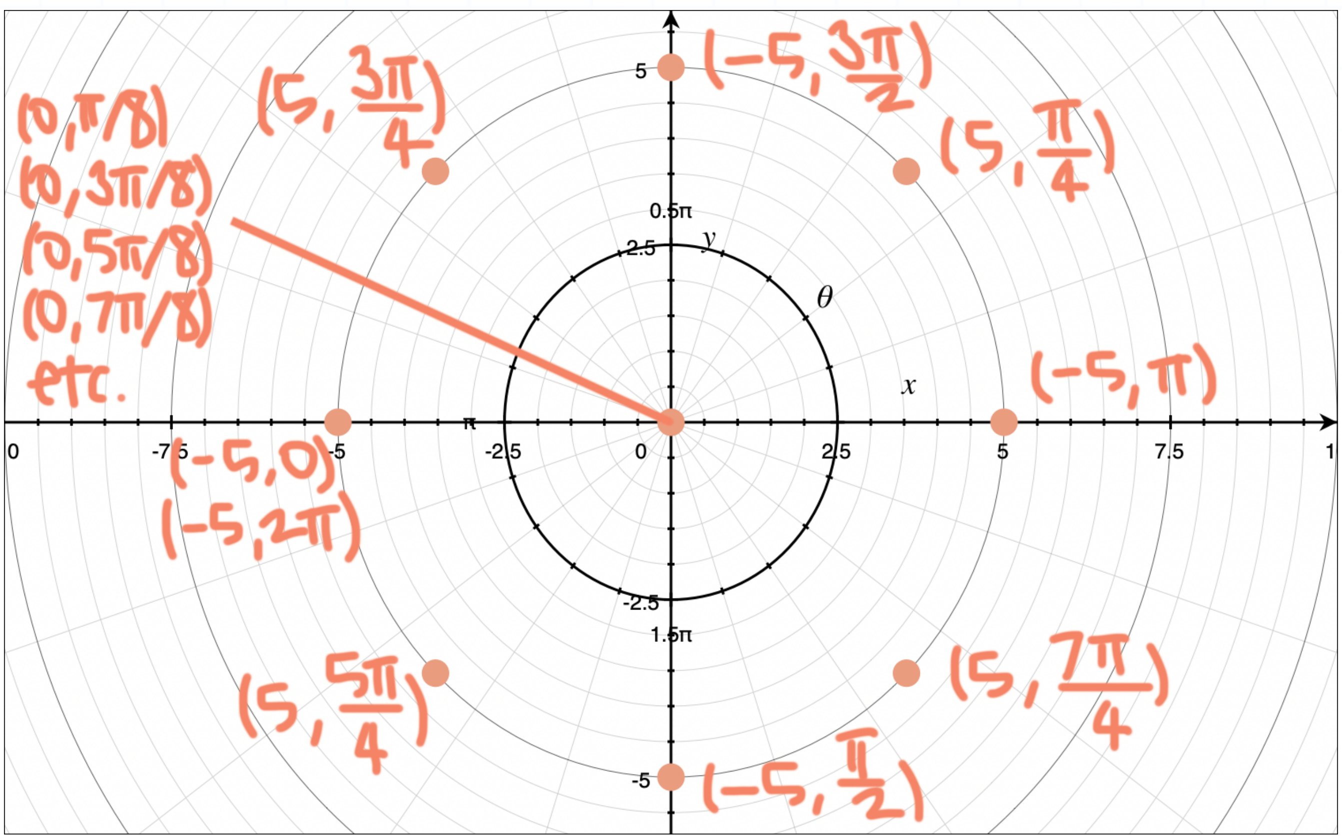
$$\theta = \frac{\pi}{8}$$

Now we'll make a table with multiples of $\pi/8$, like $\theta = 0$, $\pi/8$, $\pi/4$, $3\pi/8$, $\pi/2$, $5\pi/8$, $3\pi/4$, etc., and include the values of r that correspond to each of these θ-values.

theta	0	π/8	π/4	3π/8	π/2	5π/8	3π/4	7π/8	π	9π/8
r	-5	0	5	0	-5	0	5	0	-5	0

5π/4	11π/8	3π/2	13π/8	7π/4	15π/8	2π
5	0	-5	0	5	0	-5

Plotting these points on the polar graph gives



And if we connect these points with a smooth curve, in order, we see the graph of the rose. We start at $(-5,0)$ and then loop to the pole at $(0,\pi/8)$,

loop to $(5,\pi/4)$ then back to the pole at $(0,3\pi/8)$,

loop to $(-5,\pi/2)$ then back to the pole at $(0,5\pi/8)$, loop to $(5,3\pi/4)$ then back to the pole at $(0,7\pi/8)$,

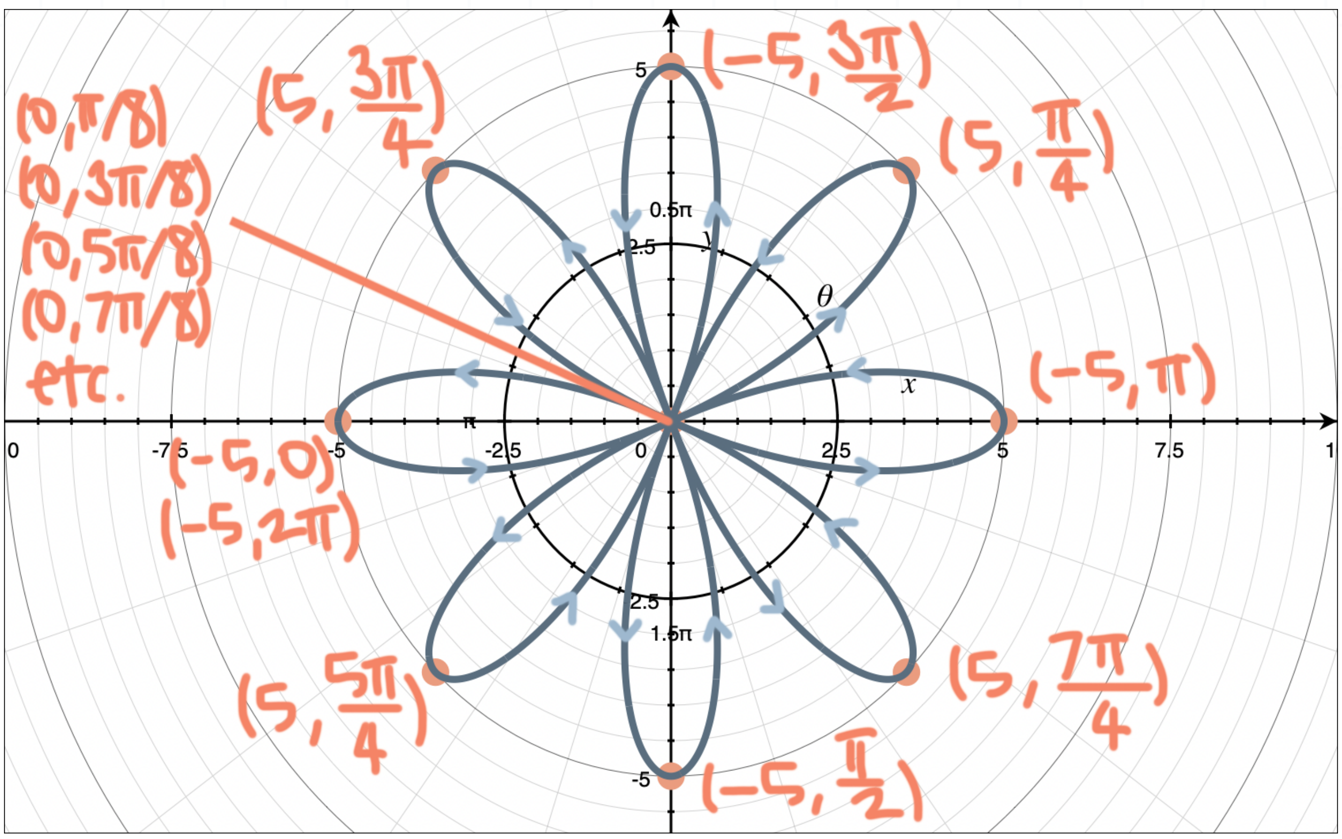
loop to $(-5,\pi)$ then back to the pole at $(0,9\pi/8)$,

loop to $(5,5\pi/4)$ then back to the pole at $(0,11\pi/8)$,

loop to $(-5,3\pi/2)$ then back to the pole at $(0,13\pi/8)$,

loop to $(5,7\pi/4)$ then to the pole at $(0,15\pi/8)$,

then finally loop back to $(-5,2\pi)$, which is actually the same point as $(-5,0)$. From then on, we're retracing the same pieces of the rose over and over.



Let's try an example with a sine rose.

Example

Sketch the graph of $r = 4\sin(6\theta)$.

Because $c = 4$ and $n = 6$, this rose will have $|2n| = |2(6)| = 12$ petals that extend out a distance of 4 from the origin. To sketch the graph, we recognize that the trigonometric function in this polar equation is $\sin(6\theta)$, and its argument is 6θ. So we'll set

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

$$\theta = \frac{\pi}{12}$$

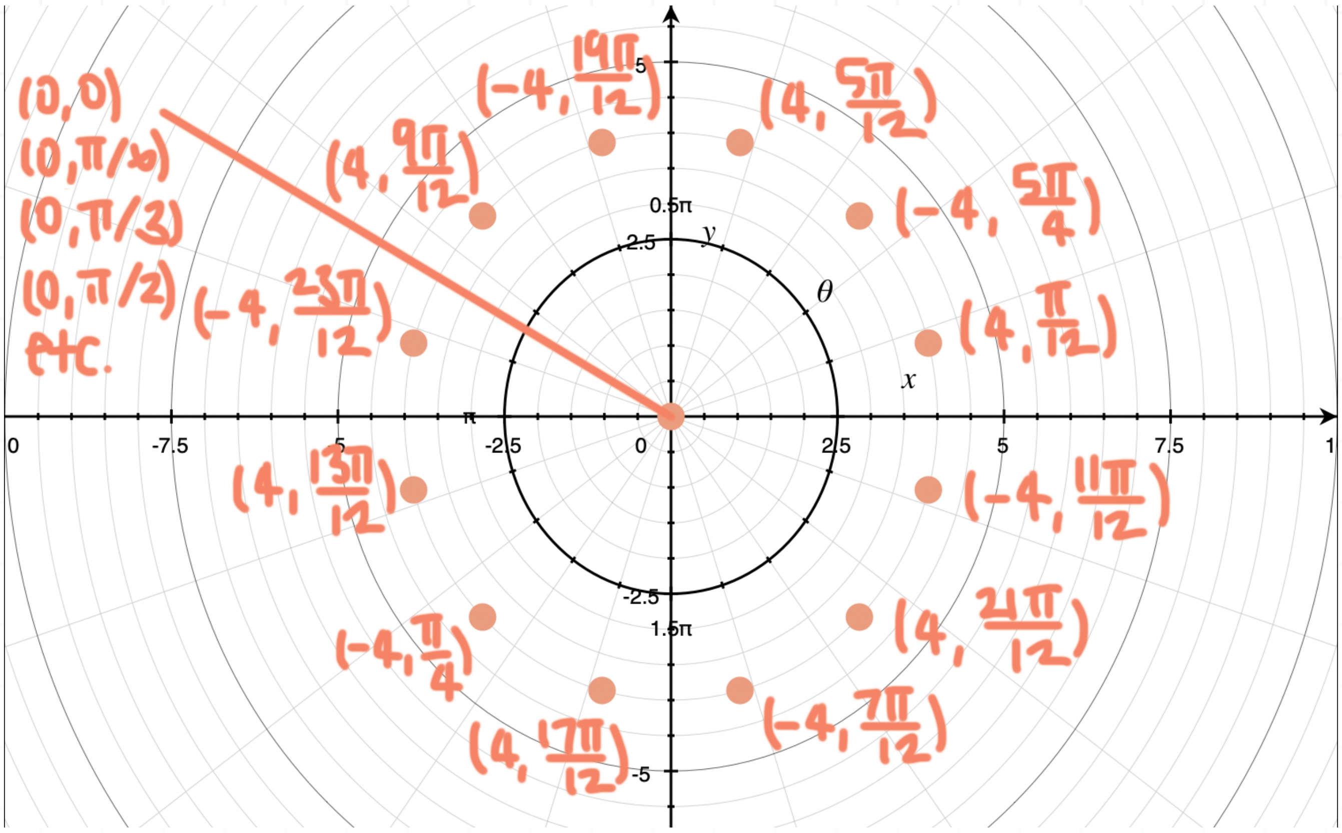
Now we'll make a table with multiples of $\pi/12$, like $\theta = 0$, $\pi/12$, $\pi/6$, $\pi/4$, $\pi/3$, $5\pi/12$, $\pi/2$, etc., and include the values of r that correspond to each of these θ-values.

theta	0	π/12	π/6	π/4	π/3	5π/12	π/2	7π/12	2π/3	3π/4
r	0	4	0	-4	0	4	0	-4	0	4

5π/6	11π/12	π	13π/12	7π/6	5π/4	4π/3	17π/12	3π/2	19π/12
0	-4	0	4	0	-4	0	4	0	-4

5π/3	7π/4	11π/6	23π/12	2π
0	4	0	-4	0

Plotting these points on the polar graph gives



And if we connect these points with a smooth curve, in order, we see the graph of the rose. We start at $(0,0)$ and then loop out to $(4,\pi/12)$,

loop back to the pole at $(0,\pi/6)$ then out to $(-4,\pi/4)$,

loop back to the pole at $(0,\pi/3)$ then out to $(4,5\pi/12)$,

loop back to the pole at $(0,\pi/2)$ then out to $(-4,7\pi/12)$,

loop back to the pole at $(0,2\pi/3)$ then out to $(4,3\pi/4)$,

loop back to the pole at $(0,5\pi/6)$ then out to $(-4,11\pi/12)$,

loop back to the pole at $(0,\pi)$ then out to $(4,13\pi/12)$,

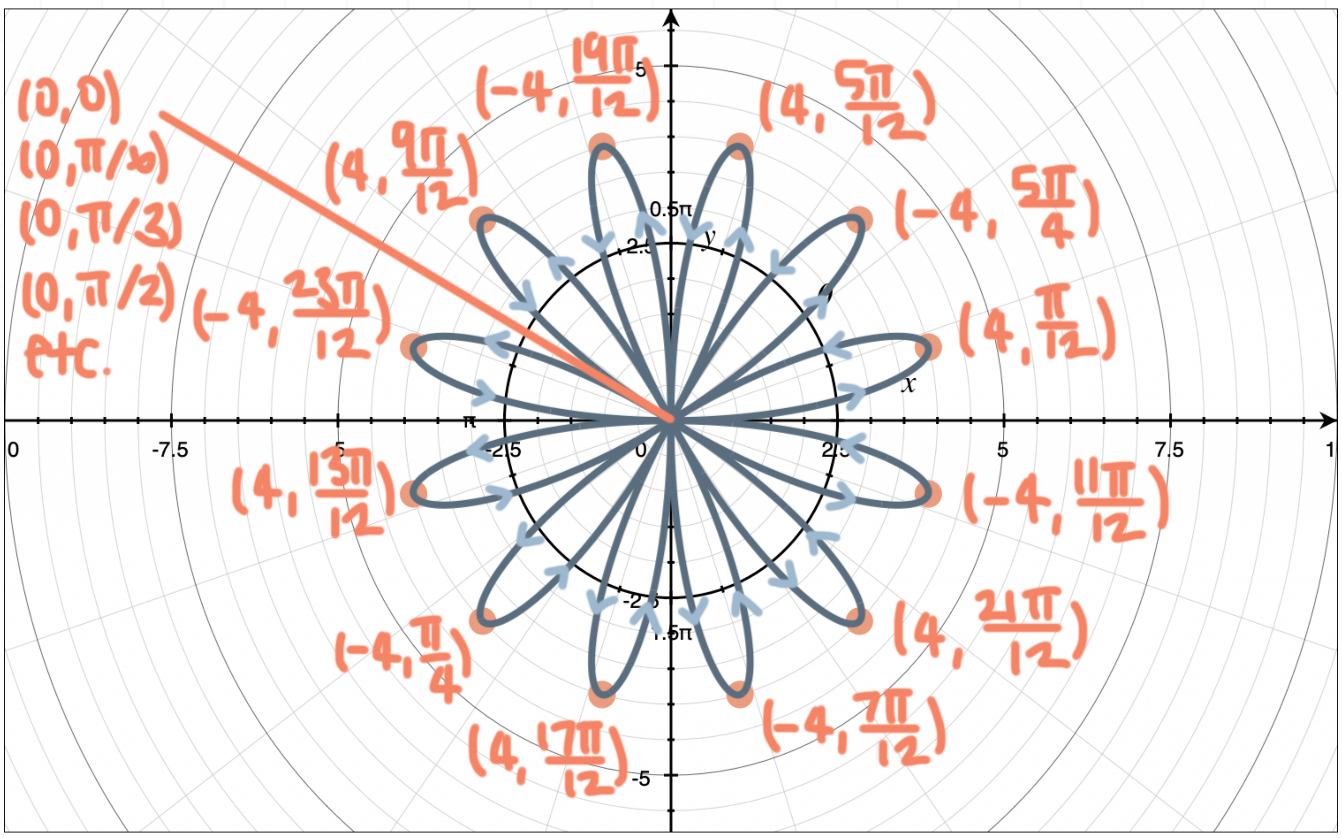
loop back to the pole at $(0,7\pi/6)$ then out to $(-4,5\pi/4)$,

loop back to the pole at $(0,4\pi/3)$ then out to $(4,17\pi/12)$,

loop back to the pole at $(0,3\pi/2)$ then out to $(-4,19\pi/12)$, loop back to the pole at $(0,5\pi/3)$ then out to $(4,7\pi/4)$,

loop back to the pole at $(0,11\pi/6)$ then out to $(-4,23\pi/12)$,

then finally back to $(0,2\pi)$, which is actually the same point as $(0,0)$. From then on, we're retracing the same pieces of the rose over and over.



Let's do one more example of a sine rose with a negative c-value and an odd n-value.

Example

Sketch the graph of $r = -6\sin(5\theta)$.

Because $c = -6$ and $n = 5$, this rose will have $|n| = |5| = 5$ petals that extend out a distance of 6 from the origin. To sketch the graph, we recognize that the trigonometric function in this polar equation is $\sin(5\theta)$, and its argument is 5θ. So we'll set

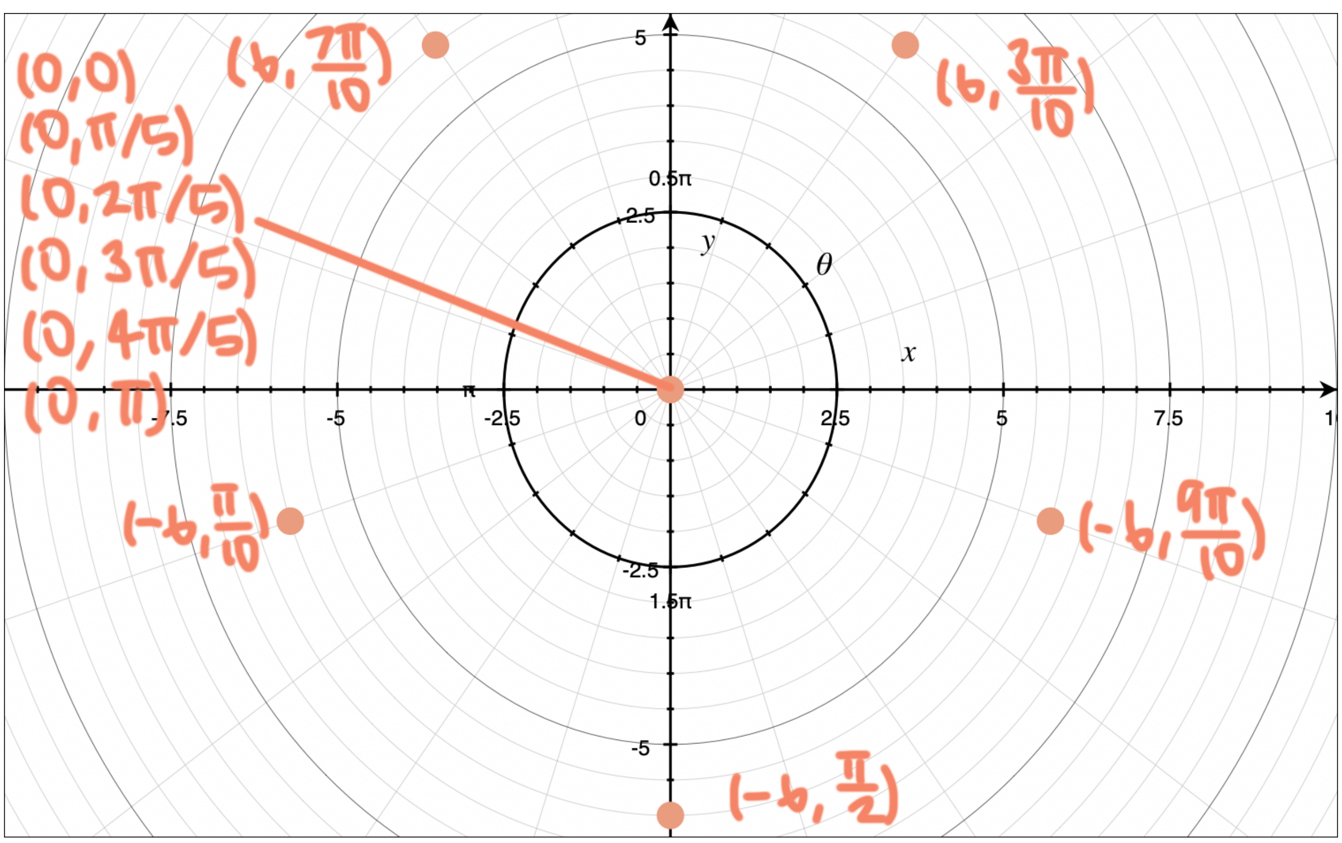
$$5\theta = \frac{\pi}{2}$$

$$\theta = \frac{\pi}{10}$$

Now we'll make a table with multiples of $\pi/10$, like $\theta = 0, \pi/10, \pi/5, 3\pi/10, 2\pi/5, \pi/2, 3\pi/5$, etc., and include the values of r that correspond to each of these θ-values.

theta	0	π/10	π/5	3π/10	2π/5	π/2	3π/5	7π/10	4π/5	9π/10	π
r	0	-6	0	6	0	-6	0	6	0	-6	0

Plotting these points on the polar graph gives



And if we connect these points with a smooth curve, in order, we see the graph of the rose. We start at $(0,0)$ and then loop out to $(-6,\pi/10)$,

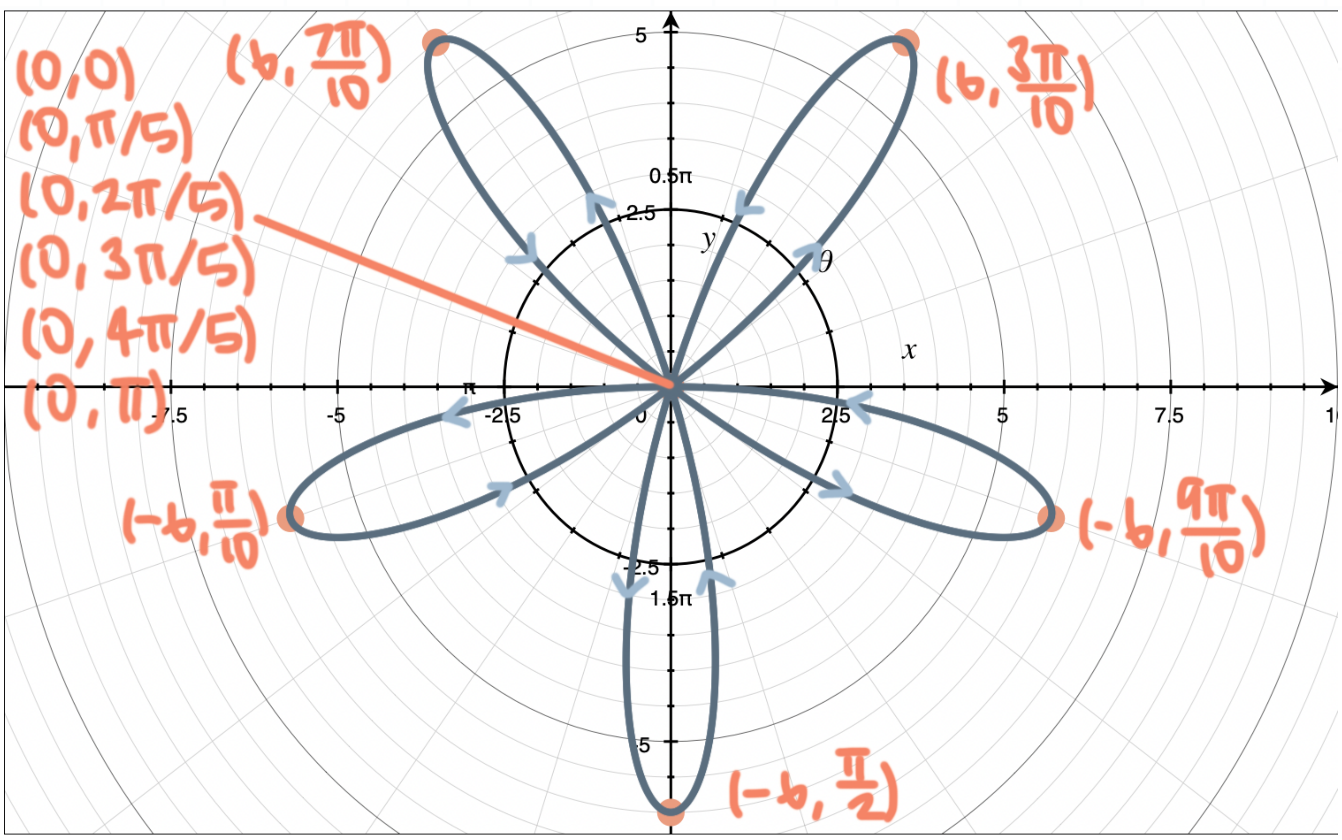
loop back to the pole at $(0,\pi/5)$ then out to $(6,3\pi/10)$,

loop back to the pole at $(0,2\pi/5)$ then out to $(-6,\pi/2)$,

loop back to the pole at $(0,3\pi/5)$ then out to $(6,7\pi/10)$,

loop back to the pole at $(0,4\pi/5)$ then out to $(-6,9\pi/10)$,

then finally back to $(0,\pi)$, which is actually the same point as $(0,0)$. From then on, we're retracing the same pieces of the rose over and over.



Graphing cardioids

Our next category of polar curves are called "**cardioids**" because of their heart-like shape, and their equations take the form

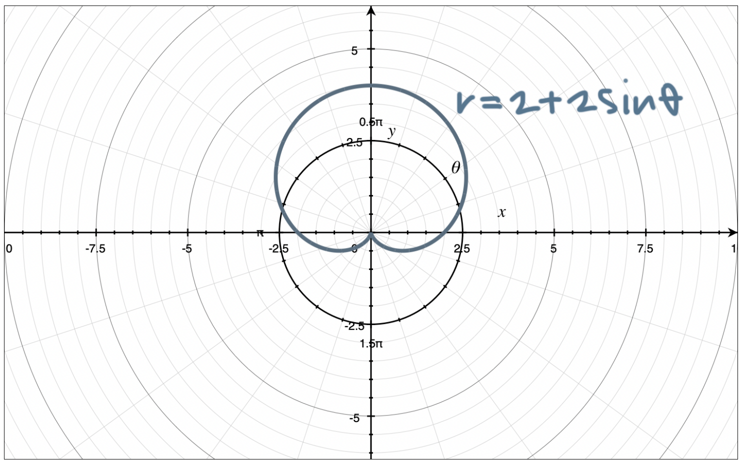
$$r = c + c\cos\theta \qquad r = c + c\sin\theta$$

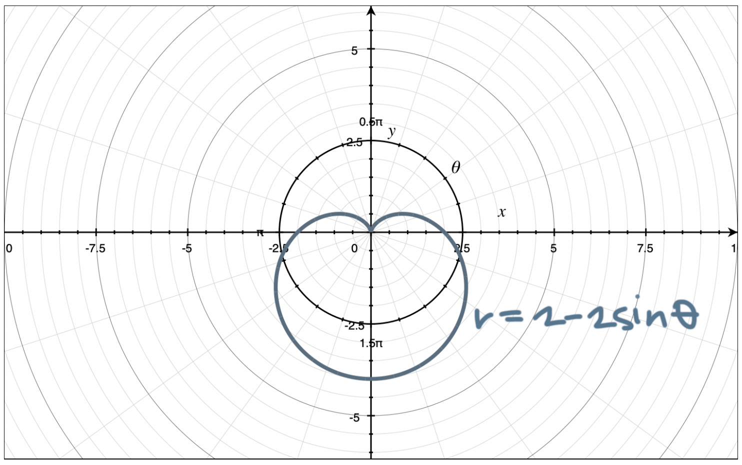
$$r = c - c\cos\theta \qquad r = c - c\sin\theta$$

where c is a positive constant. In the equations of cardioids, the coefficient on the argument θ is always 1.

Properties of cardioids

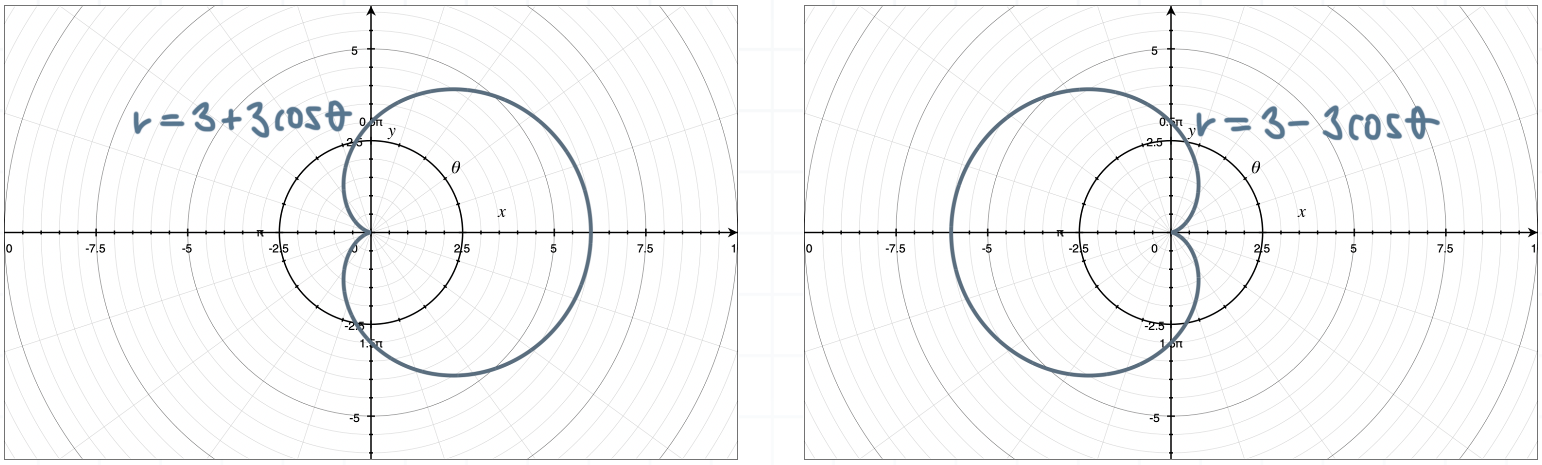
Sine cardioids will have symmetry around the vertical axis, and $r = c + c\sin\theta$ will sit mostly above the horizontal axis while $r = c - c\sin\theta$ will sit mostly below the horizontal axis. For example, the graphs of $r = 2 + 2\sin\theta$ and $r = 2 - 2\sin\theta$ are





Cosine cardioids will have symmetry around the horizontal axis, and $r = c + c\cos\theta$ will sit mostly to the right of the vertical axis while

$r = c - c\cos\theta$ will sit mostly to the left of the vertical axis. For example, the graphs of $r = 3 + 3\cos\theta$ and $r = 3 - 3\cos\theta$ are



The cardioid's furthest distance from the pole will always be at a distance of $2c$, which is why we see the two sine graphs above, $r = 2 + 2\sin\theta$ and $r = 2 - 2\sin\theta$, extend out to $2c = 2(2) = 4$, while the two cosines graphs above, $r = 3 + 3\cos\theta$ and $r = 3 - 3\cos\theta$, extend out to $2c = 2(3) = 6$.

How to sketch cardioids

We'll use the same approach to sketch cardioids that we've used previously to sketch circles and roses:

1. Set the argument of the trigonometric function equal to $\pi/2$, and then solve the equation for θ.

2. Evaluate the polar curve at multiples of the θ-value we solved for in Step 1, starting with $\theta = 0$, and plot the resulting points on the polar graph.

3. Connect the points on the polar graph with a smooth curve.

Let's work through an example of how to sketch a cosine cardioid.

Example

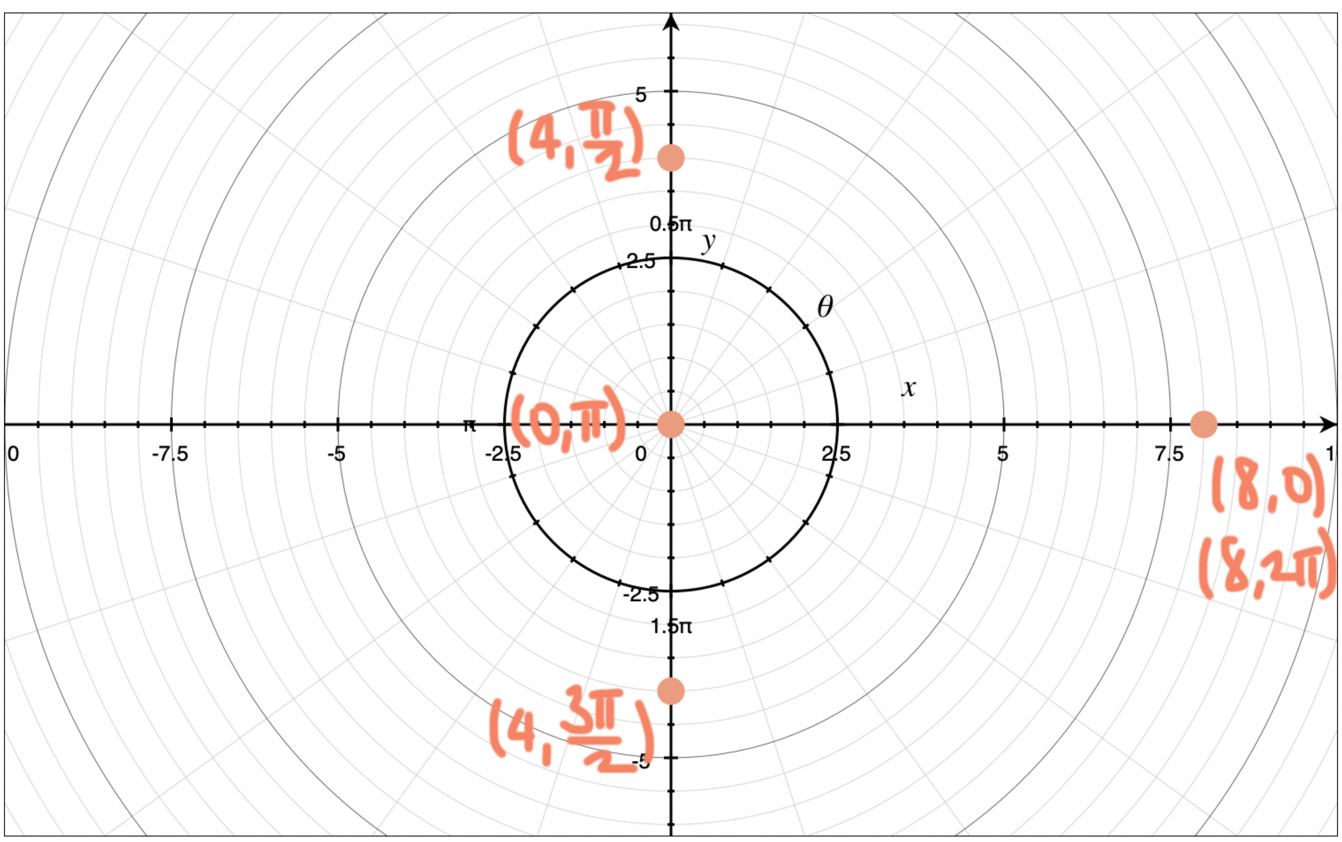
Sketch the graph of $r = 4 + 4\cos\theta$.

Because $c = 4$, this cardioid will extend out to a distance of $2c = 2(4) = 8$ from the pole. Because it's a cosine curve where the sign between the terms is positive, the graph will sit mostly to the right of the vertical axis, with symmetry across the horizontal axis.

Now we'll make a table with for $\theta = 0$, $\pi/2$, π, $3\pi/2$, and 2π, and include the values of r that correspond to each of these θ-values.

theta	0	π/2	π	3π/2	2π
r	8	4	0	4	8

Plotting these points on the polar graph gives



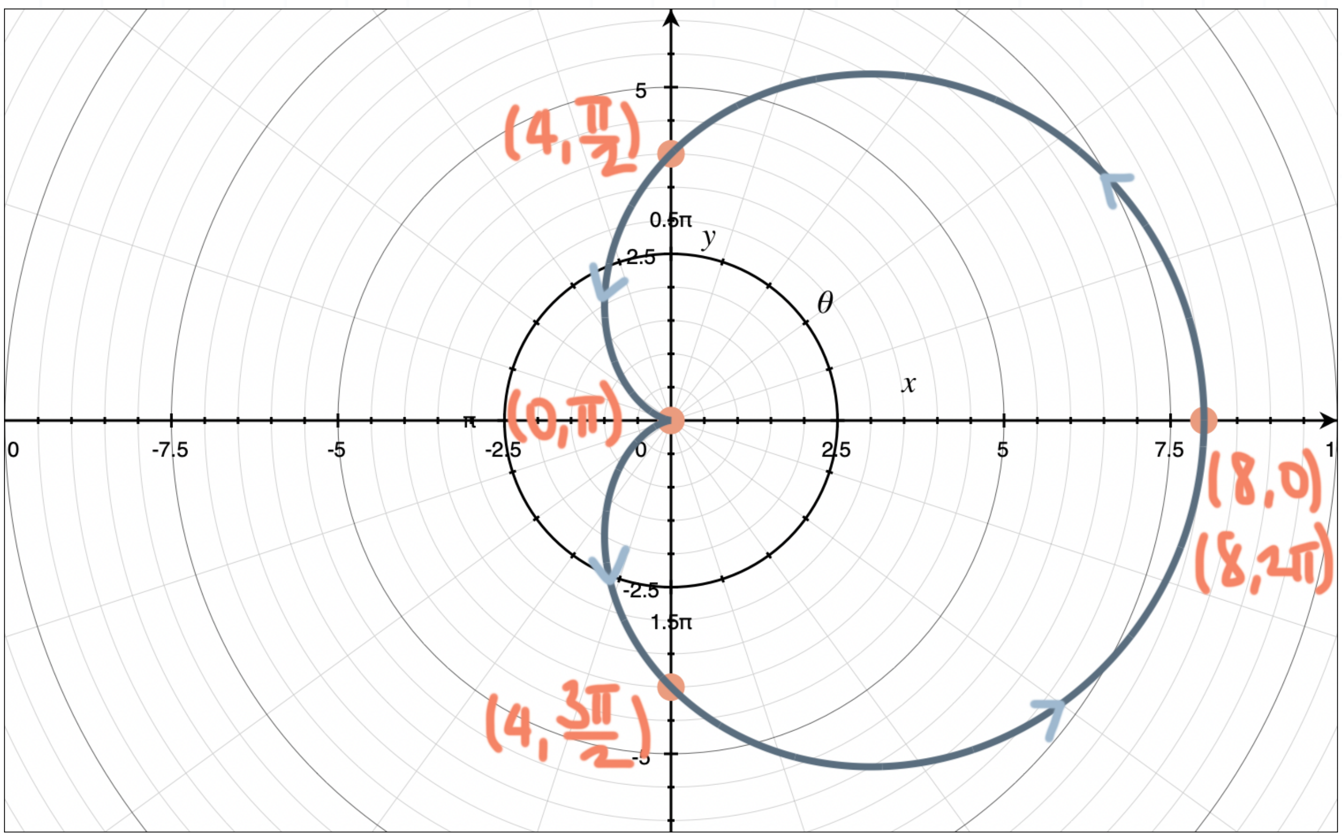
And if we connect these points with a smooth curve, in order, we see the graph of the cardioid. We start at $(8,0)$, and then

loop to $(4,\pi/2)$,

loop back to the pole at $(0,\pi)$,

loop to $(4,3\pi/2)$,

then finally loop back to $(8,2\pi)$, which is actually the same point as $(8,0)$. From then on, we're retracing the same pieces of the cardioid over and over.



Now let's do an example with a cardioid in the form $r = c - c\cos\theta$.

Example

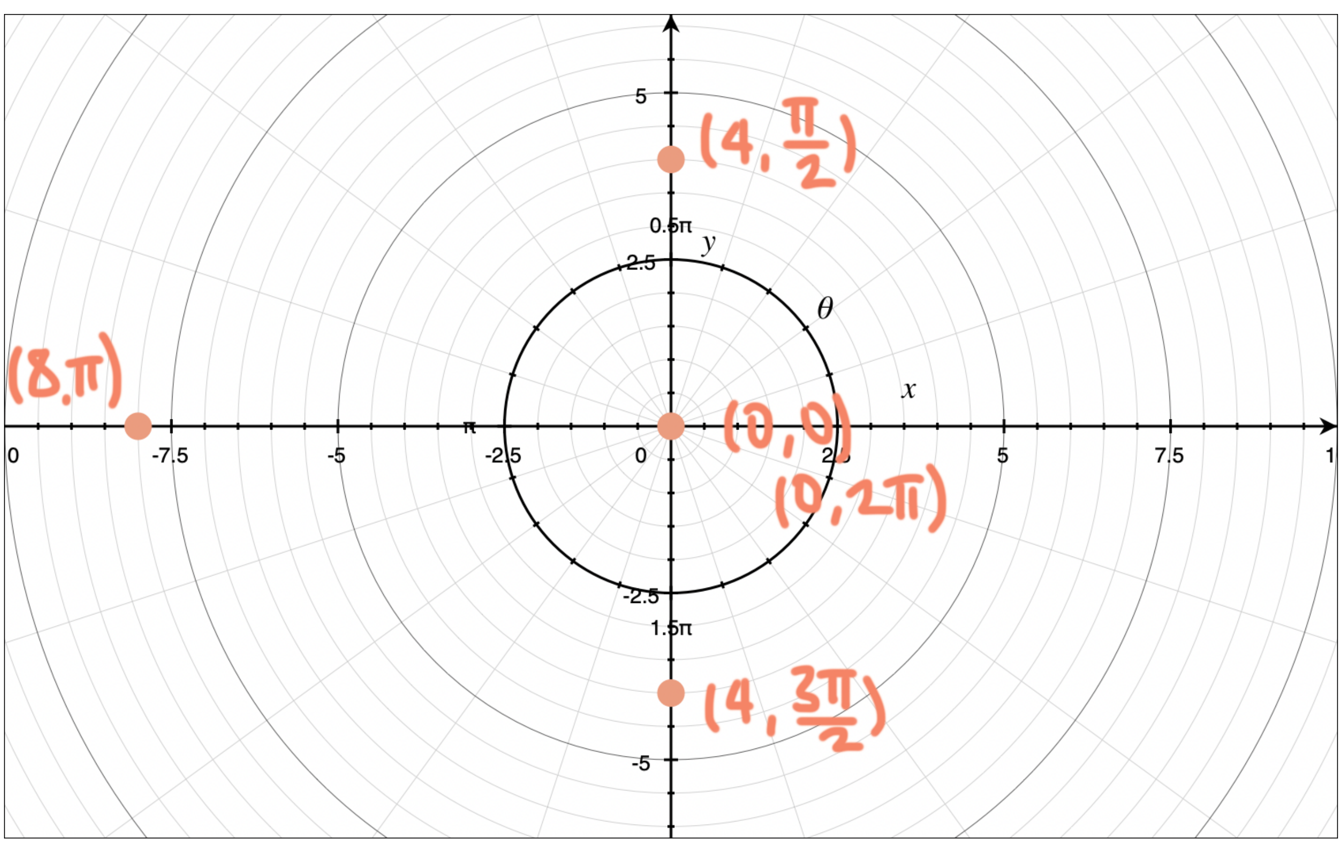
Sketch the graph of $r = 4 - 4\cos\theta$.

Because $c = 4$, this cardioid will extend out to a distance of $2c = 2(4) = 8$ from the pole. Because it's a cosine curve where the sign between the terms is negative, the graph will sit mostly to the left of the vertical axis, with symmetry across the horizontal axis.

Now we'll make a table with for $\theta = 0$, $\pi/2$, π, $3\pi/2$, and 2π, and include the values of r that correspond to each of these θ-values.

theta	0	π/2	π	3π/2	2π
r	0	4	8	4	0

Plotting these points on the polar graph gives



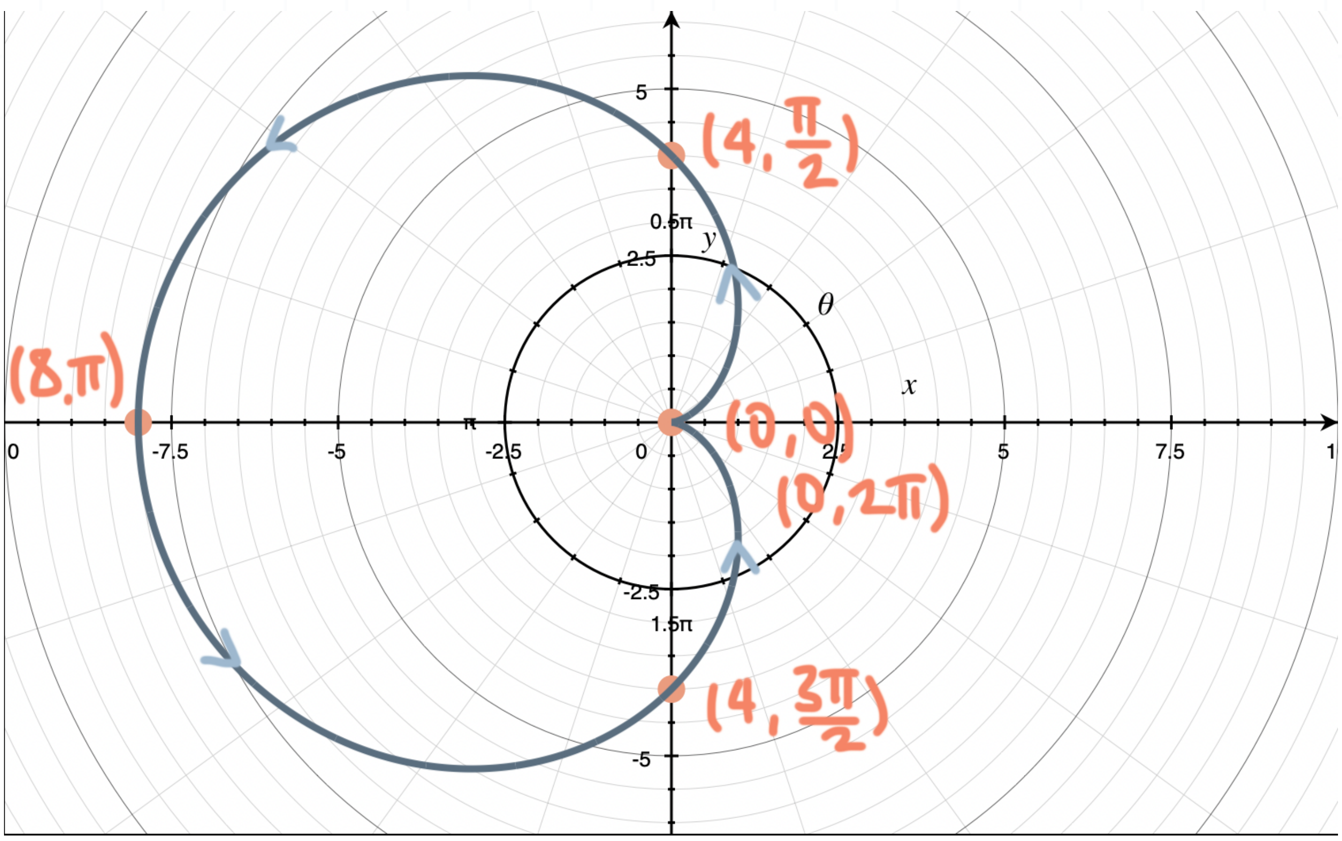
And if we connect these points with a smooth curve, in order, we see the graph of the cardioid. We start at $(0,0)$, and then

loop to $(4,\pi/2)$,

loop to $(8,\pi)$,

loop to $(4,3\pi/2)$,

then finally loop back to $(0,2\pi)$, which is actually the same point as $(0,0)$. From then on, we're retracing the same pieces of the cardioid over and over.



Now let's look at the graphs of sine cardioids.

Example

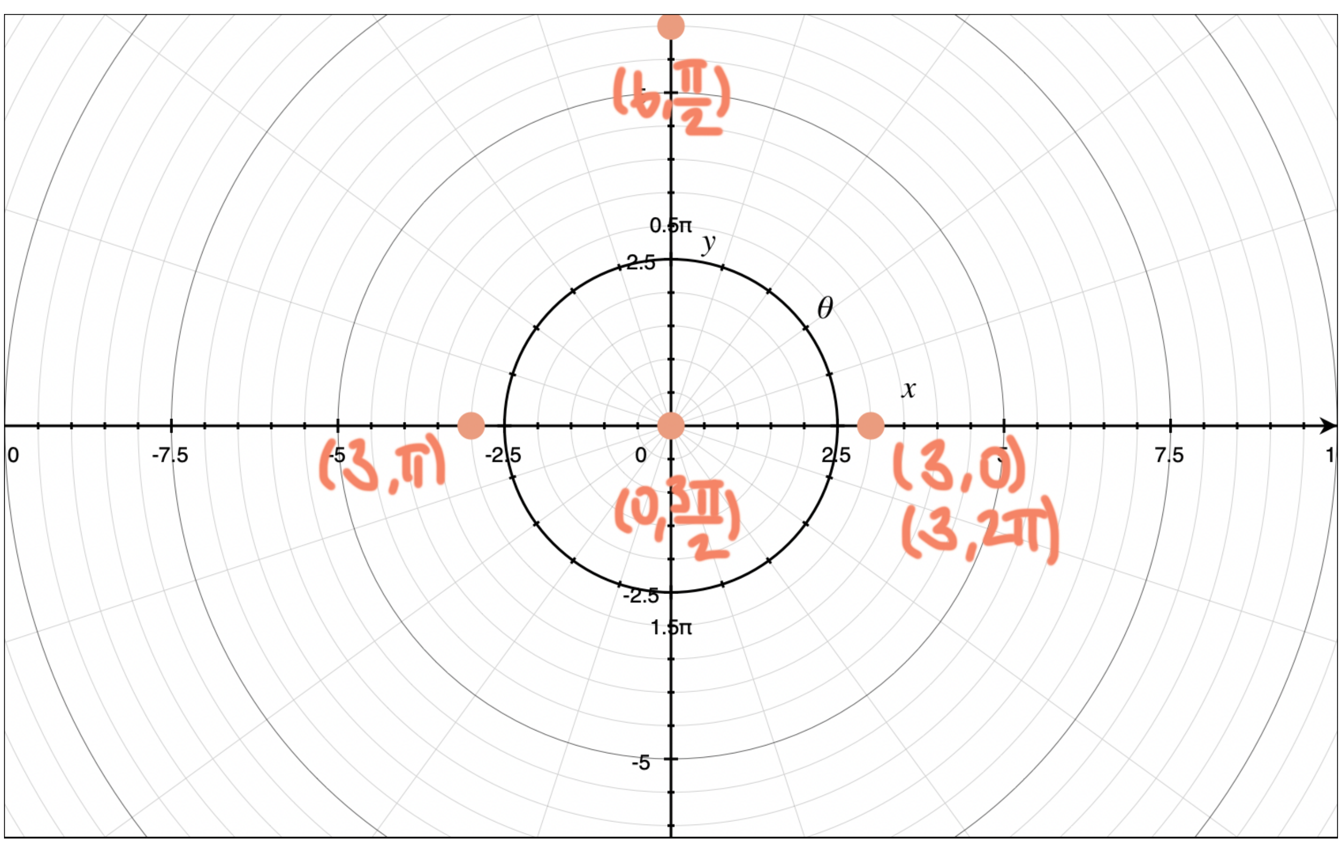
Sketch the graph of $r = 3 + 3\sin\theta$.

Because $c = 3$, this cardioid will extend out to a distance of $2c = 2(3) = 6$ from the pole. Because it's a sine curve where the sign between the terms if positive, the graph will sit mostly above the horizontal axis, with symmetry across the vertical axis.

Now we'll make a table with for $\theta = 0$, $\pi/2$, π, $3\pi/2$, and 2π, and include the values of r that correspond to each of these θ-values.

theta	0	π/2	π	3π/2	2π
r	3	6	3	0	3

Plotting these points on the polar graph gives



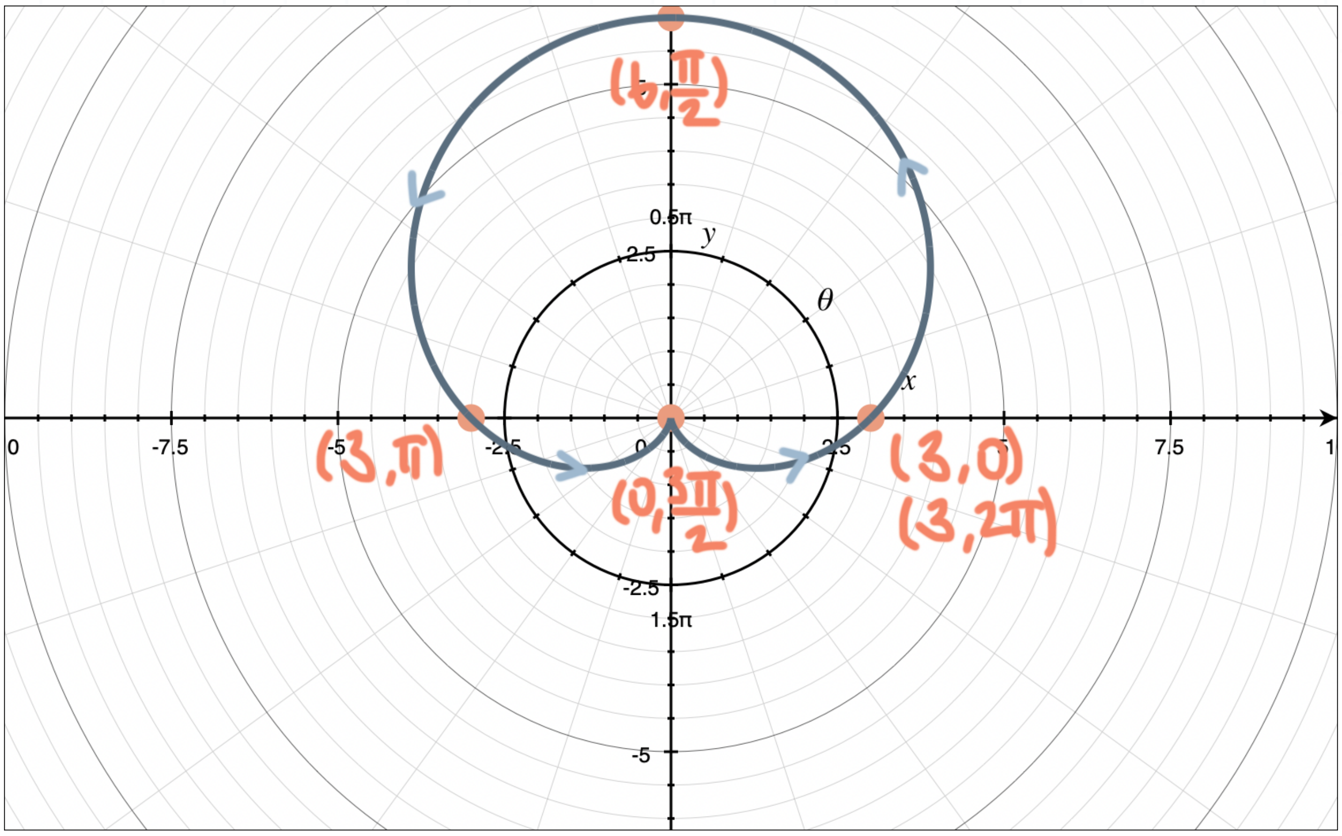
And if we connect these points with a smooth curve, in order, we see the graph of the cardioid. We start at $(3,0)$, and then

loop to $(6,\pi/2)$,

loop to $(3,\pi)$,

loop to $(0,3\pi/2)$,

then finally loop back to $(3,2\pi)$, which is actually the same point as $(3,0)$. From then on, we're retracing the same pieces of the cardioid over and over.



Let's do one more example to see what happens when $c < 0$.

Example

Sketch the graph of $r = -4 - 4\cos\theta$.

Remember we said earlier that the value of c needs to be positive. Because this equation begins with $c = -4$, we need to factor out a negative sign.

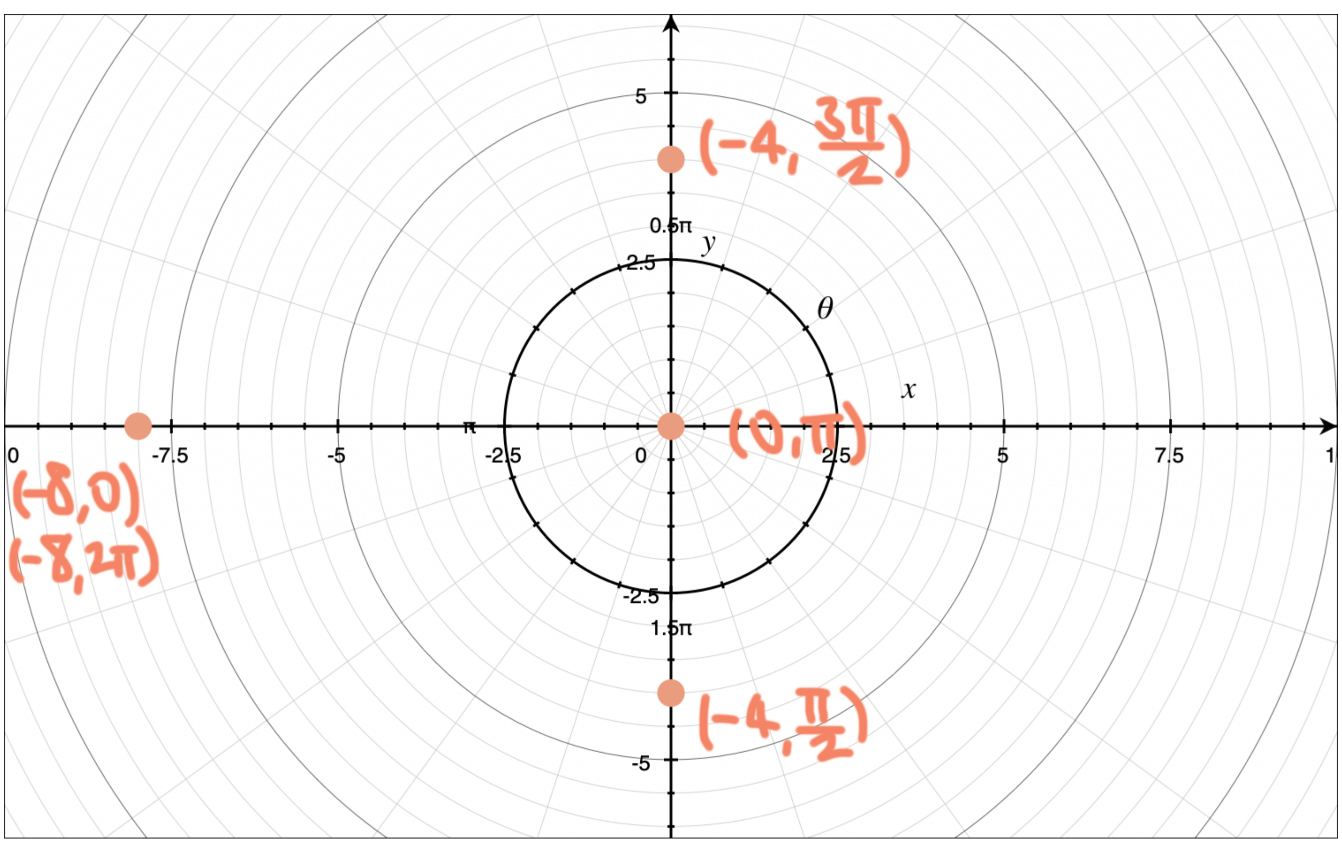
$$r = -(4 + 4\cos\theta)$$

Now we have the cardioid equation $r = 4 + 4\cos\theta$, and we'll just need to apply the negative sign to each of our r-values.

Now we'll make a table with for $\theta = 0$, $\pi/2$, π, $3\pi/2$, and 2π, and include the values of r that correspond to each of these θ-values.

theta	0	π/2	π	3π/2	2π
r	-8	-4	0	-4	-8

Plotting these points on the polar graph gives



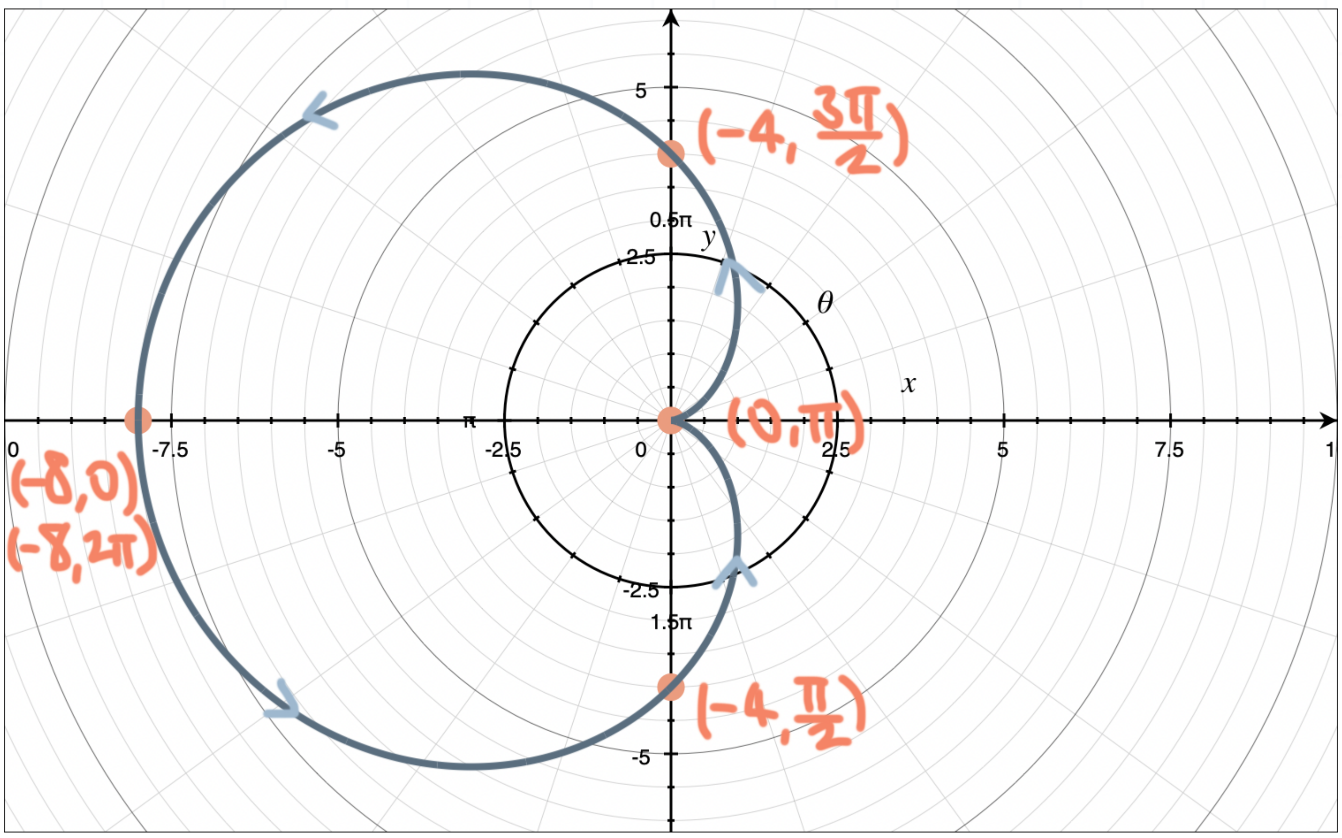
And if we connect these points with a smooth curve, in order, we see the graph of the cardioid. We start at $(-8,0)$, and then

loop to $(-4,\pi/2)$,

loop to $(0,\pi)$,

loop to $(-4,3\pi/2)$,

then finally loop back to $(-8,2\pi)$, which is actually the same point as $(-8,0)$. From then on, we're retracing the same pieces of the cardioid over and over.



In this last example, what we notice is that the graph of $r = -(4 + 4\cos\theta)$ is simply the graph of $r = 4 + 4\cos\theta$ reflected over the vertical axis. Which means $r = -(4 + 4\cos\theta)$ is actually equivalent to the cosine curve

$$r = 4 - 4\cos\theta$$

That's why we include the condition that c should be positive in these cardioid equations, because leading the cardioid equation with a negative value of c never actually gives us a new curve.

$r = -c + c\cos\theta$ turns out to be equivalent to $r = c + c\cos\theta$

$r = -c - c\cos\theta$ turns out to be equivalent to $r = c - c\cos\theta$

$r = -c + c\sin\theta$ turns out to be equivalent to $r = c + c\sin\theta$

$r = -c - c\sin\theta$ turns out to be equivalent to $r = c - c\sin\theta$

Graphing limacons

In the previous lesson, we looked at how to graph cardioids, which are a specific category of limaçon (limaçon is the French word for "snail," which is a nod to the shape of the curve). The equations of **limaçons** take the form

$$r = a + b\cos\theta \qquad\qquad r = a + b\sin\theta$$

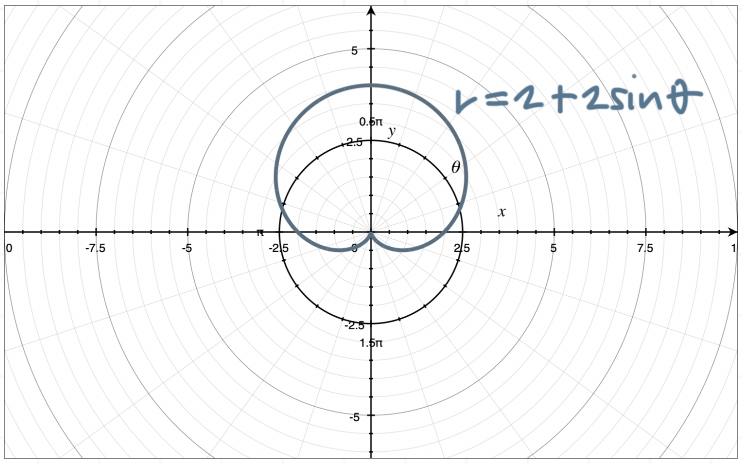
$$r = a - b\cos\theta \qquad\qquad r = a - b\sin\theta$$

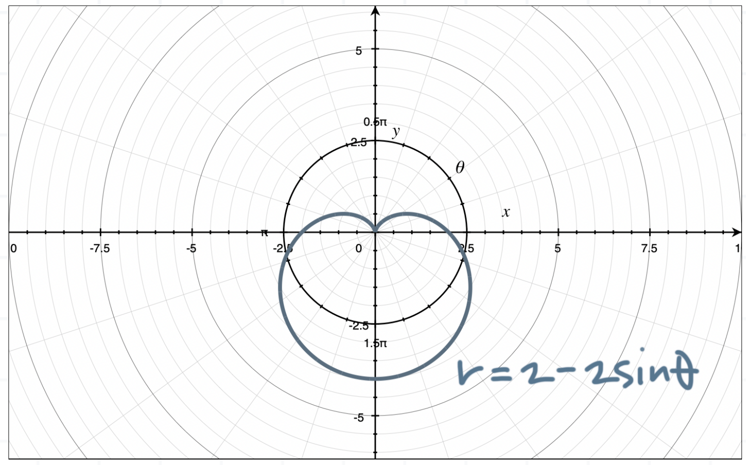
where a and b are positive numbers and where $a \neq b$. A cardioid is just a category of limaçon where $a = b$, so in this lesson we'll focus just on limaçons where $a \neq b$.

Properties of limaçons

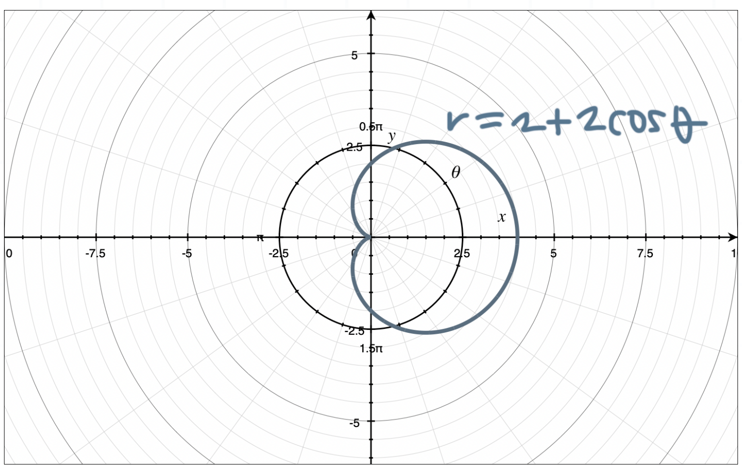
Just like for cardioids specifically, the location of the graphs of limaçons depends on whether the curve is a sine limaçon or a cosine limaçon, and whether the sign between the terms in the equation is positive or negative.

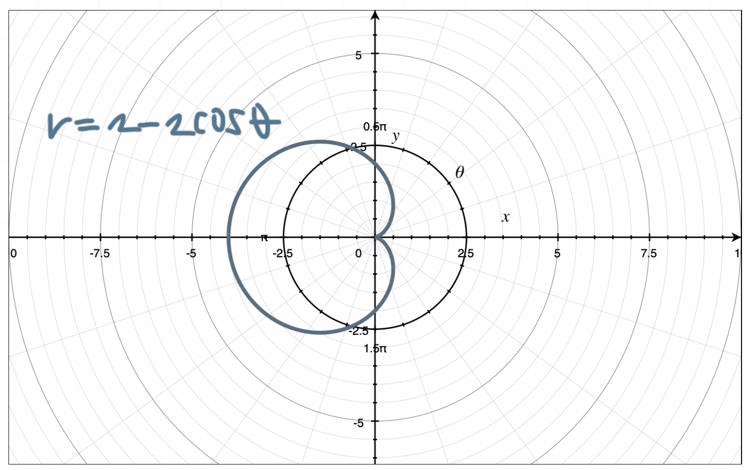
Sine limaçons are symmetric about the vertical axis, and sine limaçon equations with a positive sign will sit mostly above the horizontal axis, while sine limaçon equations with a negative sign will sit mostly below the horizontal axis.





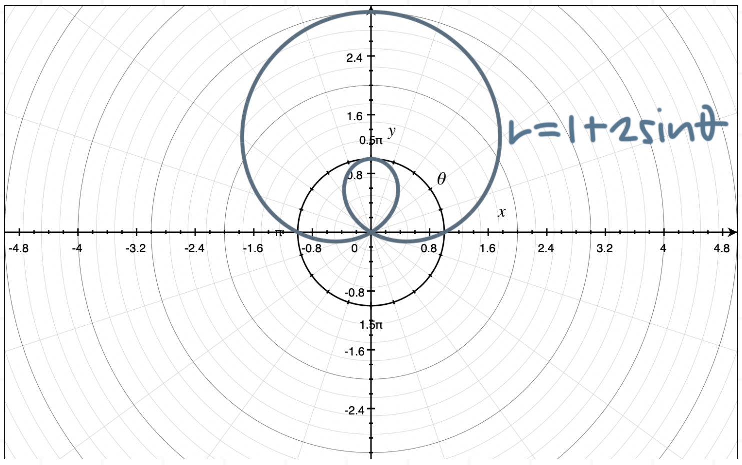
Cosine limaçons are symmetric about the horizontal axis, and cosine limaçon equations with a positive sign will sit mostly to the right of the vertical axis, while cosine limaçon equations with a negative sign will sit mostly to the left of the vertical axis.

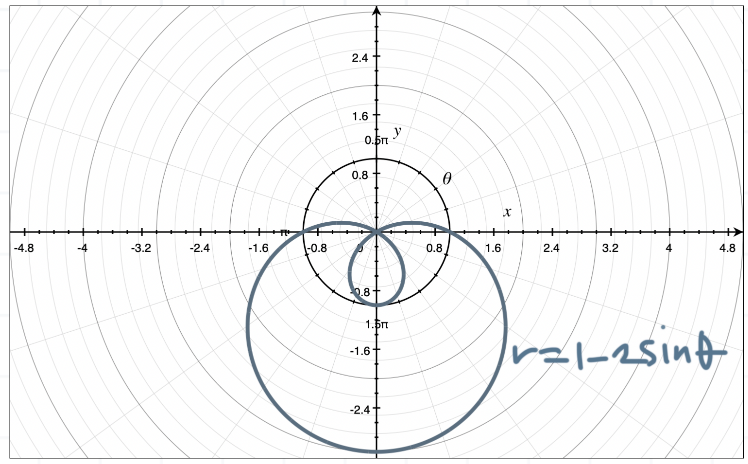


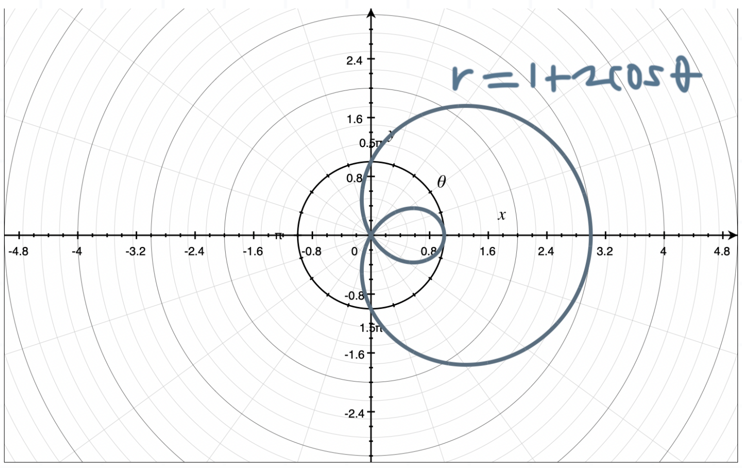


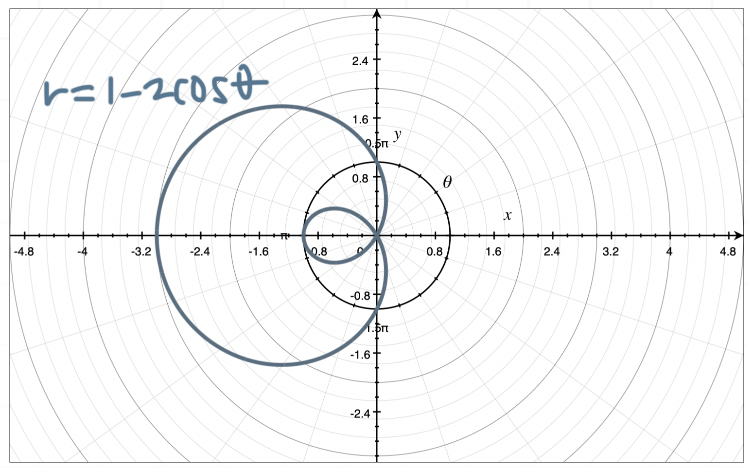
The relationship between a and b will determine the exact shape of the limaçon. We've already seen that the limaçon will be a cardioid when $a=b$, or put another way, when $a/b=1$.

When $a/b<1$, the graph of the limaçon will include a small loop.

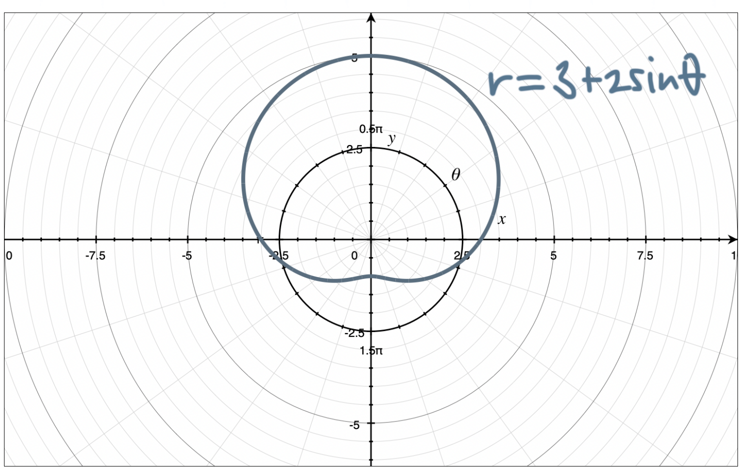


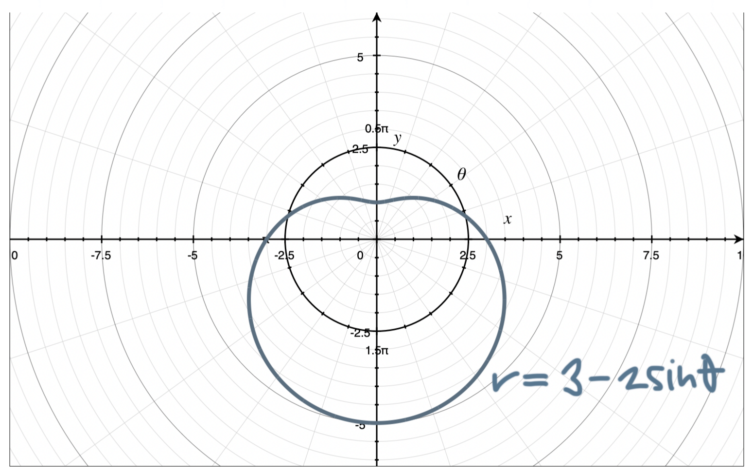


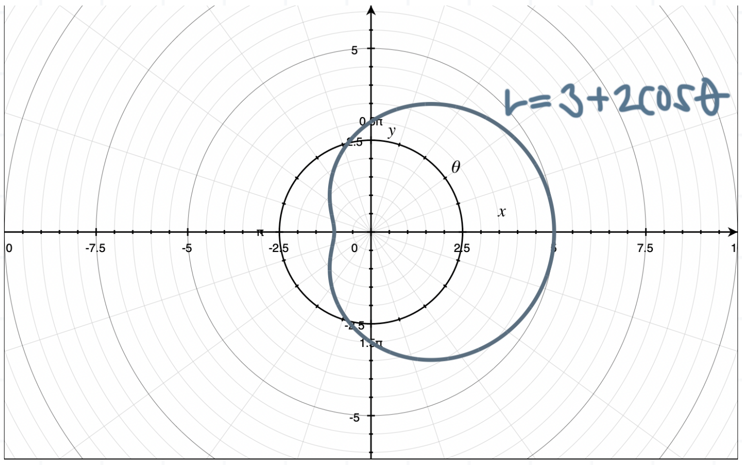


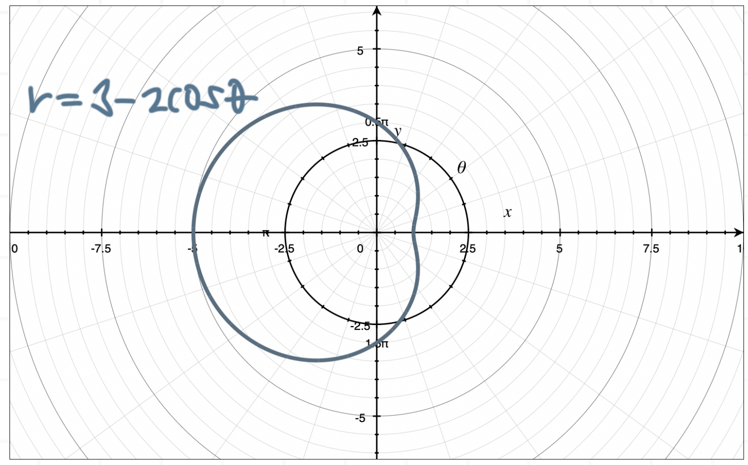


When $1 < a/b < 2$, the graph of the limaçon will include a small dip, like the cardioids we looked at earlier, but less severe.

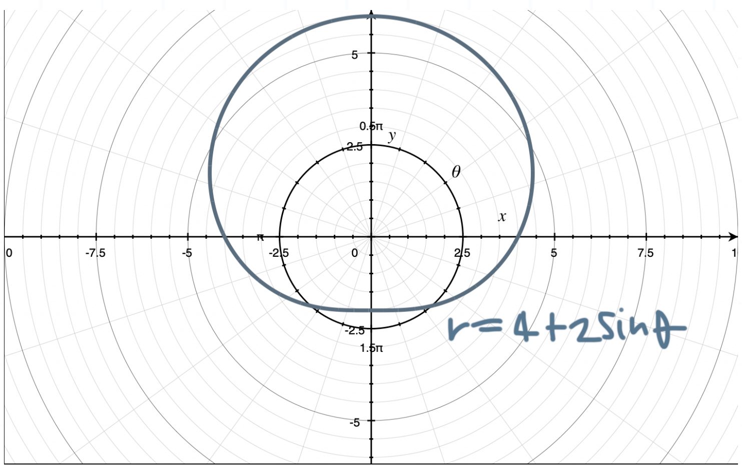


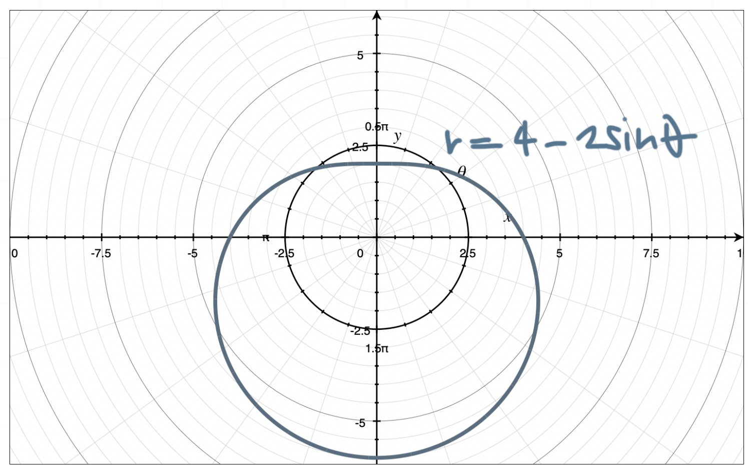


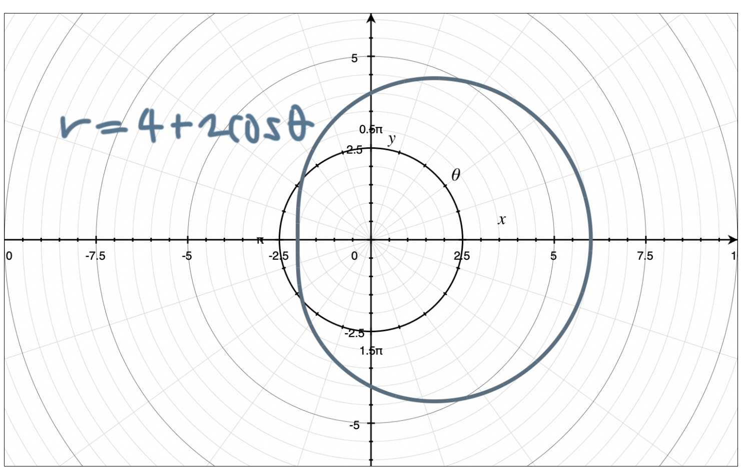


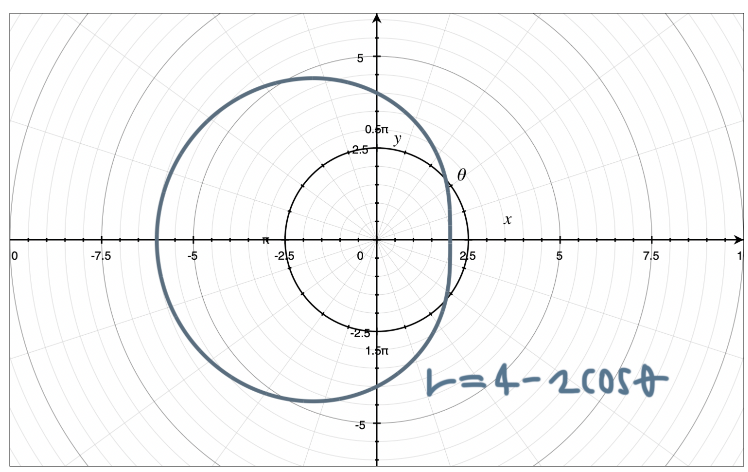


And when $a/b \geq 2$, the graph of the limaçon will include the smallest dip yet, even less severe than the $1 < a/b < 2$ case. The shape of these curves is close to, but not quite, a perfect circle.









How to sketch limaçons

We'll use the same approach to sketch limaçons that we've used previously to sketch circles, roses, and cardioids:

1. Set the argument of the trigonometric function equal to $\pi/2$, and then solve the equation for θ.

2. Evaluate the polar curve at multiples of the θ-value we solved for in Step 1, starting with $\theta = 0$, and plot the resulting points on the polar graph.

3. Connect the points on the polar graph with a smooth curve.

Let's do an example where we sketch the graph of a cosine limaçon.

Example

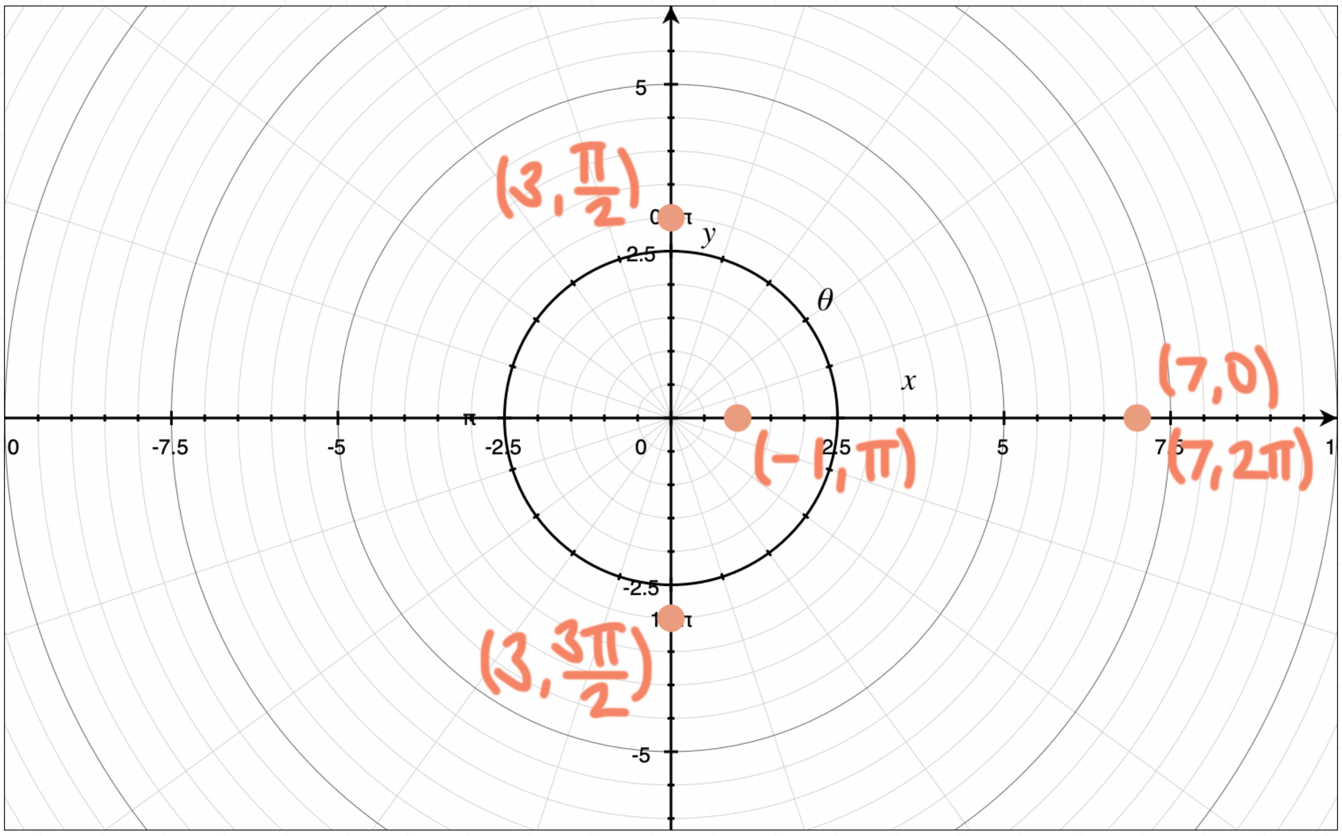
Sketch the graph of $r = 3 + 4\cos\theta$.

With $a = 3$ and $b = 4$, we get $a/b = 3/4 < 1$, which means the graph of this limaçon will include a small loop. And because this is a cosine limaçon with a positive sign separating the terms, it'll be symmetric about the horizontal axis, and sit mostly to the right of the vertical axis.

Now we'll make a table with for $\theta = 0$, $\pi/2$, π, $3\pi/2$, and 2π, and include the values of r that correspond to each of these θ-values.

theta	0	π/2	π	3π/2	2π
r	7	3	-1	3	7

Plotting these points on the polar graph gives



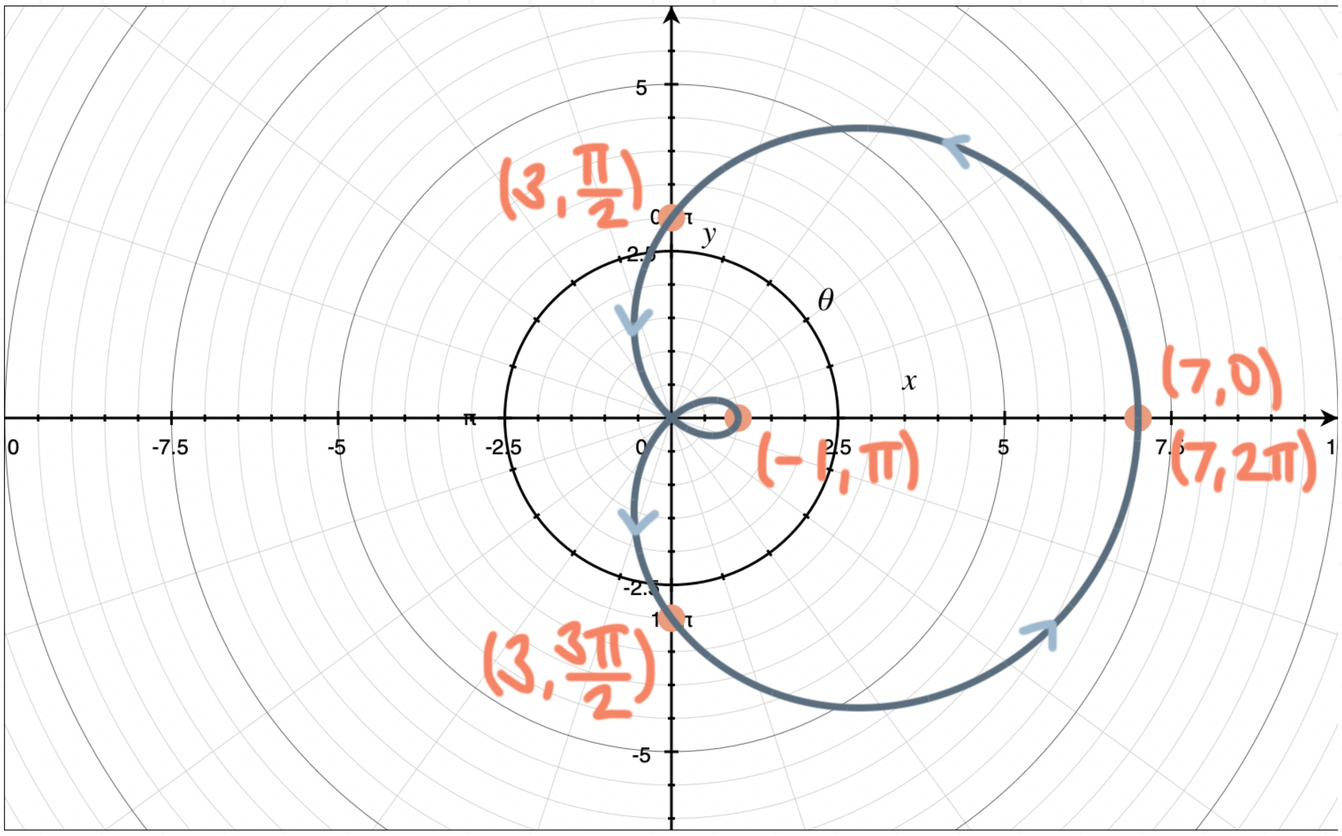
And if we connect these points with a smooth curve, in order, we see the graph of the limaçon. We start at $(7,0)$, and then

loop to $(3,\pi/2)$,

loop to $(-1,\pi)$,

loop to $(3,3\pi/2)$,

then finally loop back to $(7,2\pi)$, which is actually the same point as $(7,0)$. From then on, we're retracing the same pieces of the limaçon over and over.



Let's do another example, this time with a sine limaçon and a different a/b ratio.

Example

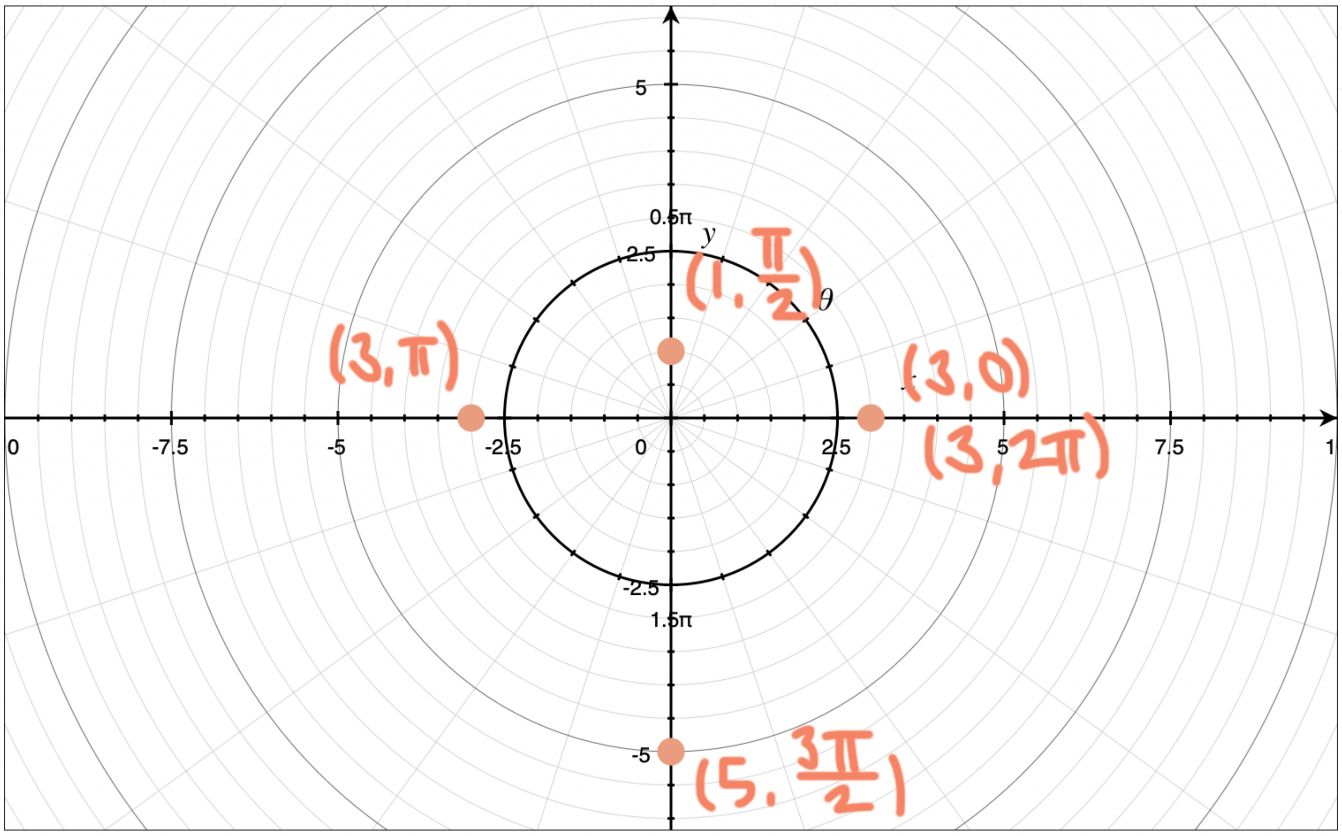
Sketch the graph of $r = 3 - 2\sin\theta$.

With $a = 3$ and $b = 2$, we get $a/b = 3/2$, and $1 < 3/2 < 2$, which means the graph of this limaçon will include a small dip. And because this is a sine limaçon with a negative sign separating the terms, it'll be symmetric about the vertical axis, and sit mostly below the vertical axis.

Now we'll make a table with for $\theta = 0$, $\pi/2$, π, $3\pi/2$, and 2π, and include the values of r that correspond to each of these θ-values.

theta	0	π/2	π	3π/2	2π
r	3	1	3	5	3

Plotting these points on the polar graph gives



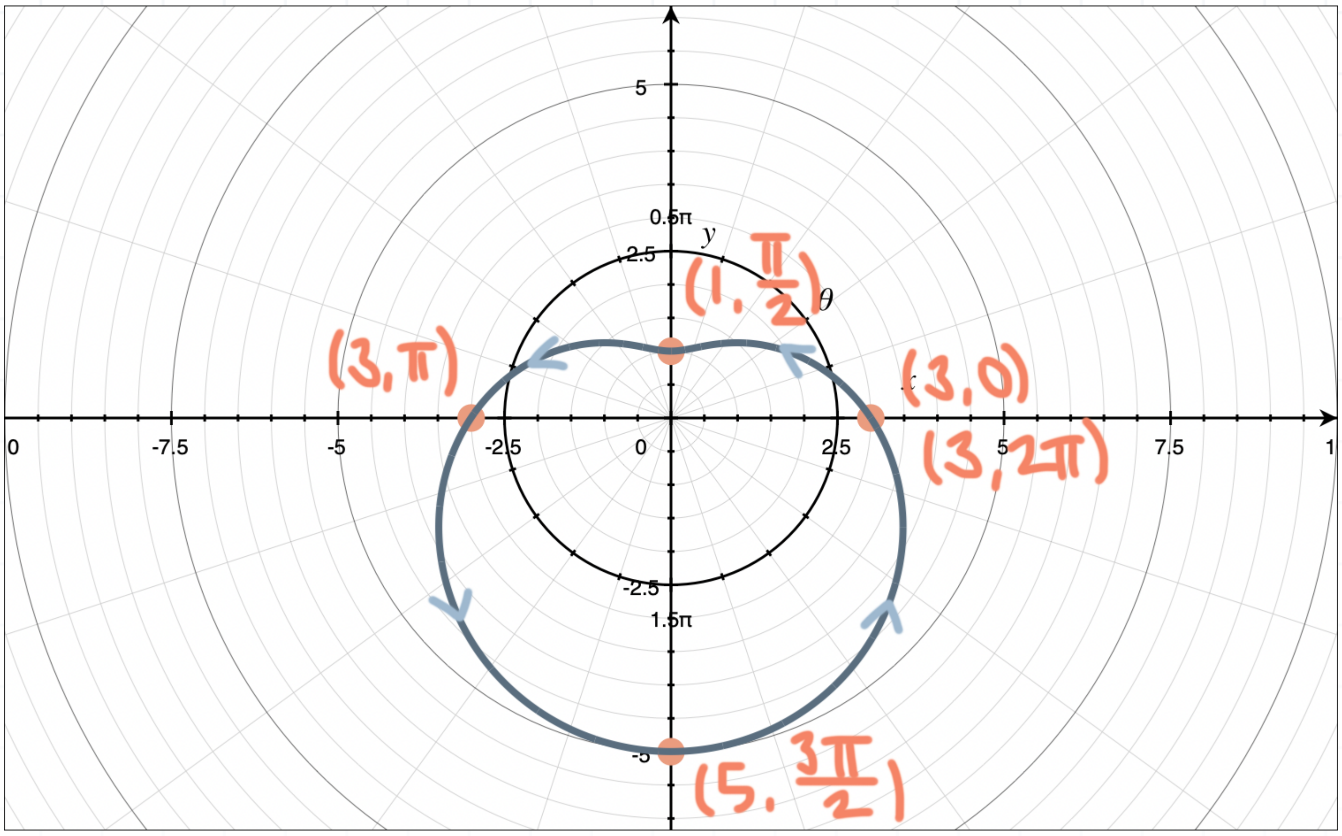
And if we connect these points with a smooth curve, in order, we see the graph of the limaçon. We start at $(3,0)$, and then

loop to $(1,\pi/2)$,

loop to $(3,\pi)$,

loop to $(5,3\pi/2)$,

then finally loop back to $(3,2\pi)$, which is actually the same point as $(3,0)$. From then on, we're retracing the same pieces of the limaçon over and over.


5
0.5π
2.5
y
θ
(1, $\frac{\pi}{2}$)
(3, π)
(3, 0)
(3, 2π)
-7.5
-5
-2.5
0
2.5
5
7.5
-2.5
1.5π
-5
(5, $\frac{3\pi}{2}$)

Graphing lemniscates

The last shape we'll look at is a figure-eight shape called a "**lemniscate**." The equations of lemniscates take the form

$$r^2 = c^2 \sin(2\theta) \qquad r^2 = -c^2 \sin(2\theta)$$

$$r^2 = c^2 \cos(2\theta) \qquad r^2 = -c^2 \cos(2\theta)$$

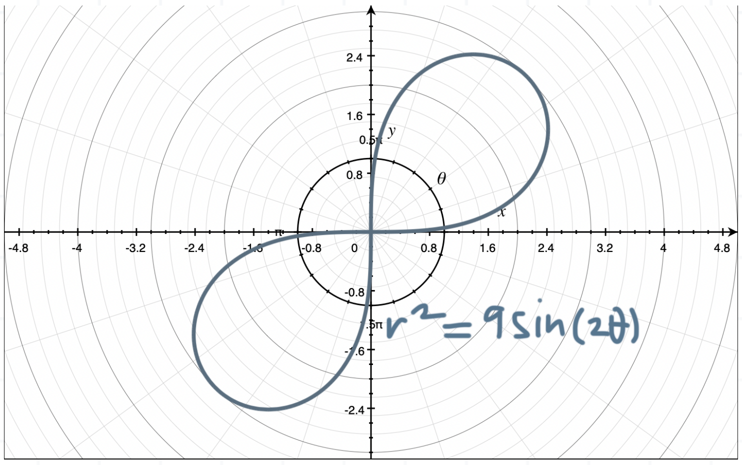
where c is positive constant. Lemniscates are always symmetric around the pole.

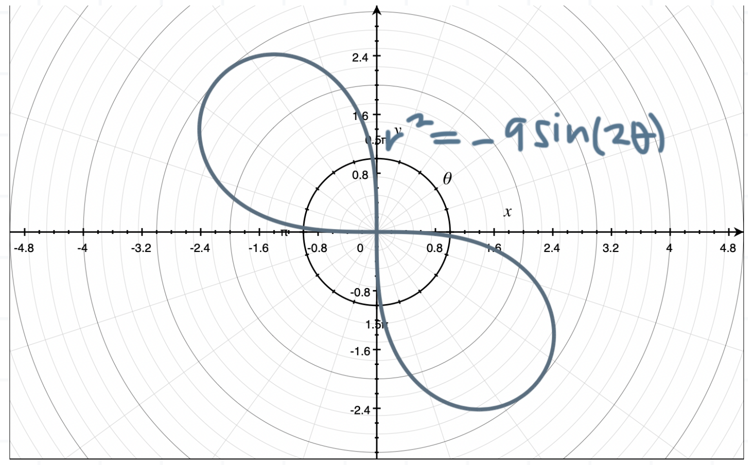
Properties of lemniscates

The argument of the trigonometric function in the equation of a lemniscate is always (2θ), which is a reminder that a lemniscate always has two loops. These are almost like the petals in a rose, except that there are always exactly two loops, instead of the varying number of petals that we find in roses.

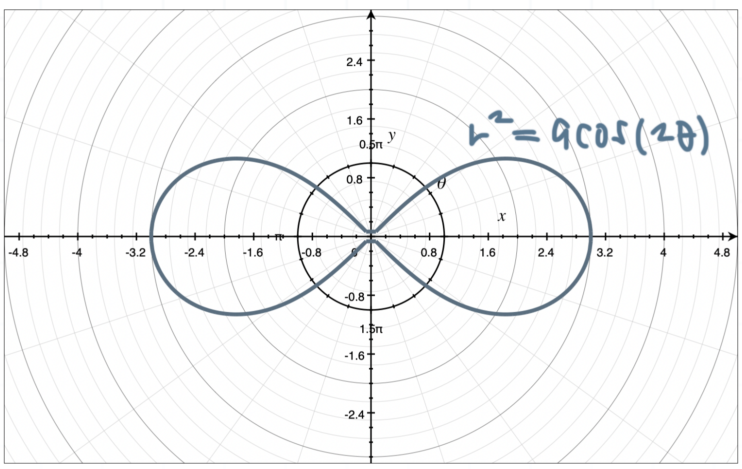
The tip of the two loops of the lemniscate will always lie a distance of c away from the pole. Where the loops themselves lie depends on whether we have a sine lemniscate or cosine lemniscate, and whether the equation includes c^2 or $-c^2$.

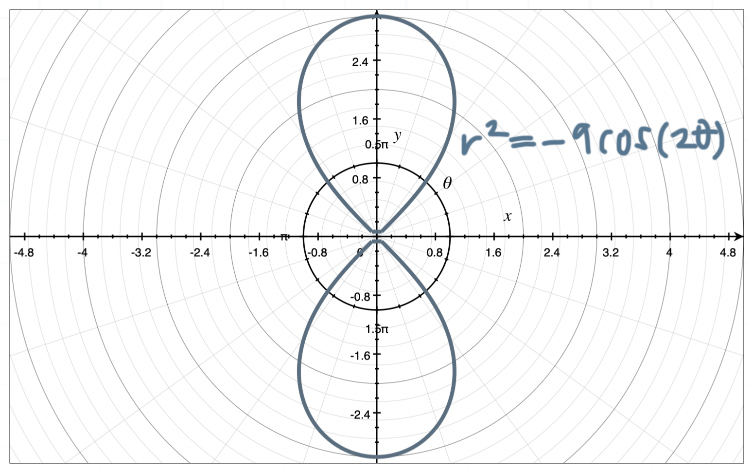
The loops of a sine lemniscate with c^2 lie in the first and third quadrants, while the loops of a sine lemniscate with $-c^2$ lie in the second and fourth quadrants.





The loops of a cosine lemniscate with c^2 lie along the horizontal axis, while the loops of a cosine lemniscate with $-c^2$ lie along the vertical axis.





The loops of the lemniscate will extend out from the pole to a distance of $r = c$.

How to sketch lemniscates

Because of the r^2 value that we see in lemniscate equations, we can sometimes find angles θ at which the lemniscate is undefined. For instance, take the lemniscate $r^2 = 9\cos(2\theta)$ that we graphed above. If we try to evaluate this polar equation at $\theta = \pi/2$, we get

$$r^2 = 9\cos\left(2 \cdot \frac{\pi}{2}\right)$$

$$r^2 = 9\cos\pi$$

$$r^2 = 9(-1)$$

$$r^2 = -9$$

$$r = \pm\sqrt{-9}$$

We can't use real numbers to take the square root of a negative value, so we run into a problem when we try to sketch this polar equation at $\theta = \pi/2$.

Because of this issue, we'll simplify our plan for sketching lemniscates and just use the value of c^2, the positive or negative sign in front of the c^2, and whether the lemniscate is a sine or cosine lemniscate, in order to sketch the graph. Our plan will be to

1. Identify the lemniscate as a sine or cosine lemniscate.

2. Identify the equation's leading sign as positive or negative.

3. Determine the value of c (not c^2).

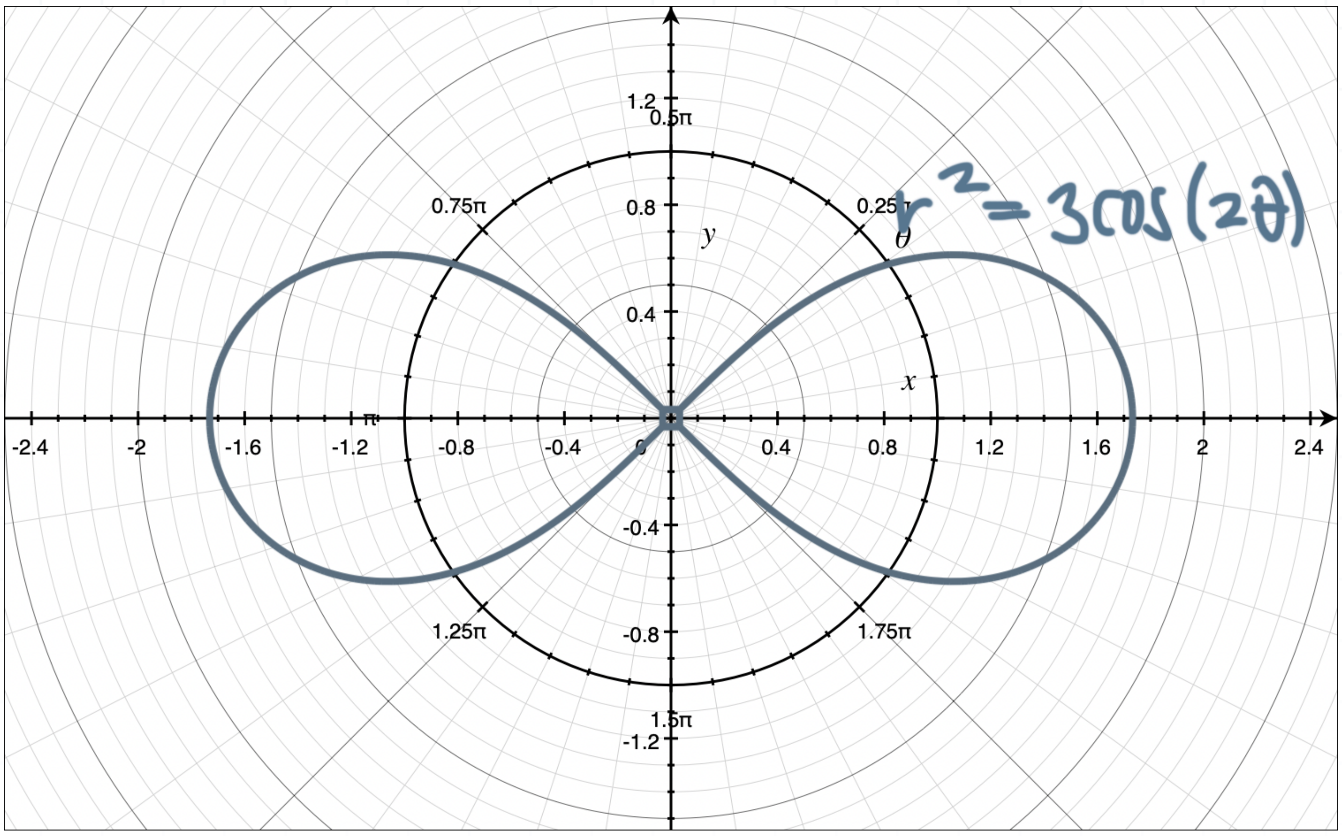
4. Use these three facts to sketch the lemniscate.

Let's do an example where we sketch the graph of a cosine lemniscate.

Example

Sketch the graph of $r^2 = 3\cos(2\theta)$.

This is a positive cosine lemniscate, which means the two loops will sit along the horizontal axis. And with $c = \sqrt{3} \approx 1.73$, the two loops will extend out a distance of $r \approx 1.73$. Therefore, the graph of the lemniscate will be

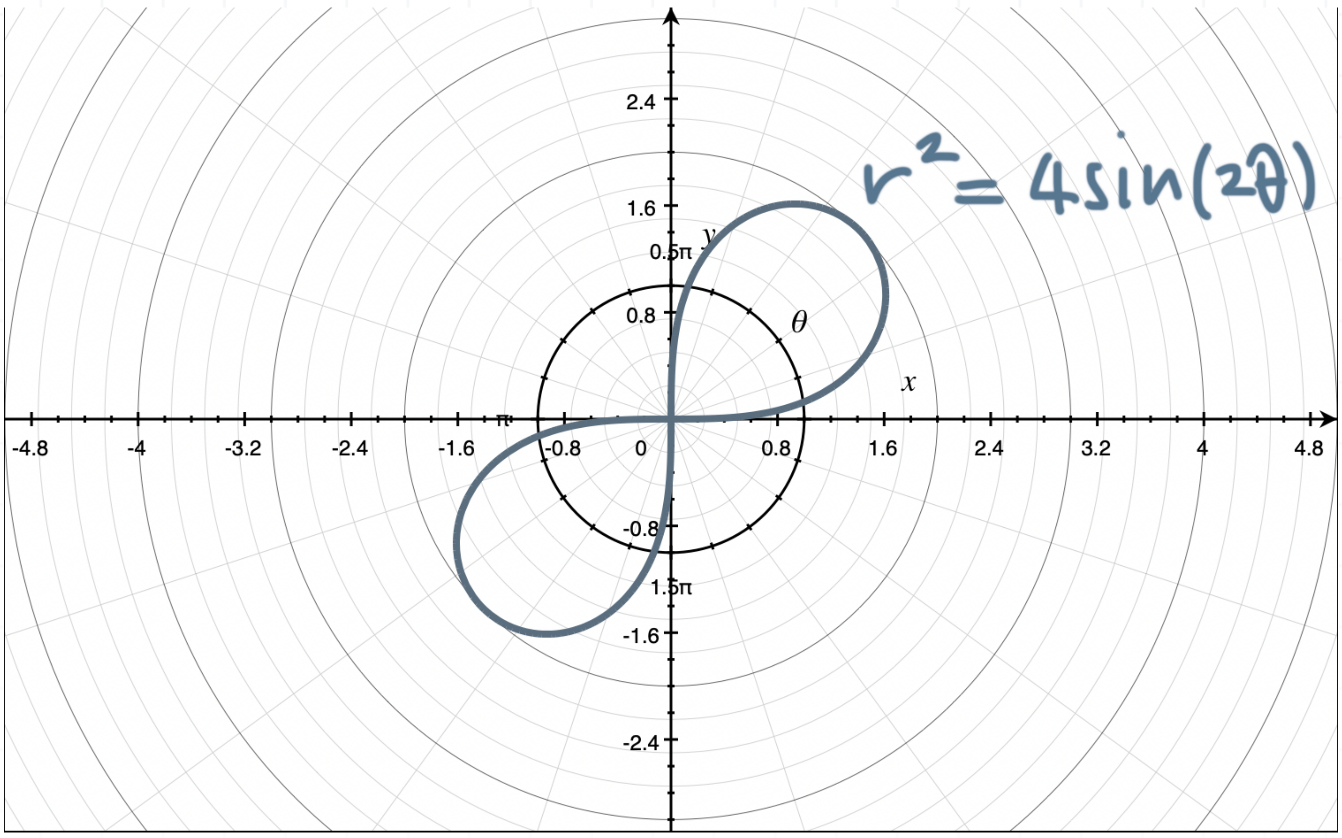


Now let's do an example with a sine lemniscate.

Example

Graph the lemniscate $r^2 = 4\sin(2\theta)$.

This is a positive sine lemniscate, which means the two loops will sit in the first and third quadrants. And with $c = \sqrt{4} = 2$, the loops extend out a distance of $r = 2$. Therefore, the graph of the lemniscate will be



$r^2 = 4\sin(2\theta)$
2.4
1.6
0.5π
0.8
θ
x
y
π
-4.8
-4
-3.2
-2.4
-1.6
-0.8
0
0.8
1.6
2.4
3.2
4
4.8
-0.8
1.5π
-1.6
-2.4

Intersection of polar curves

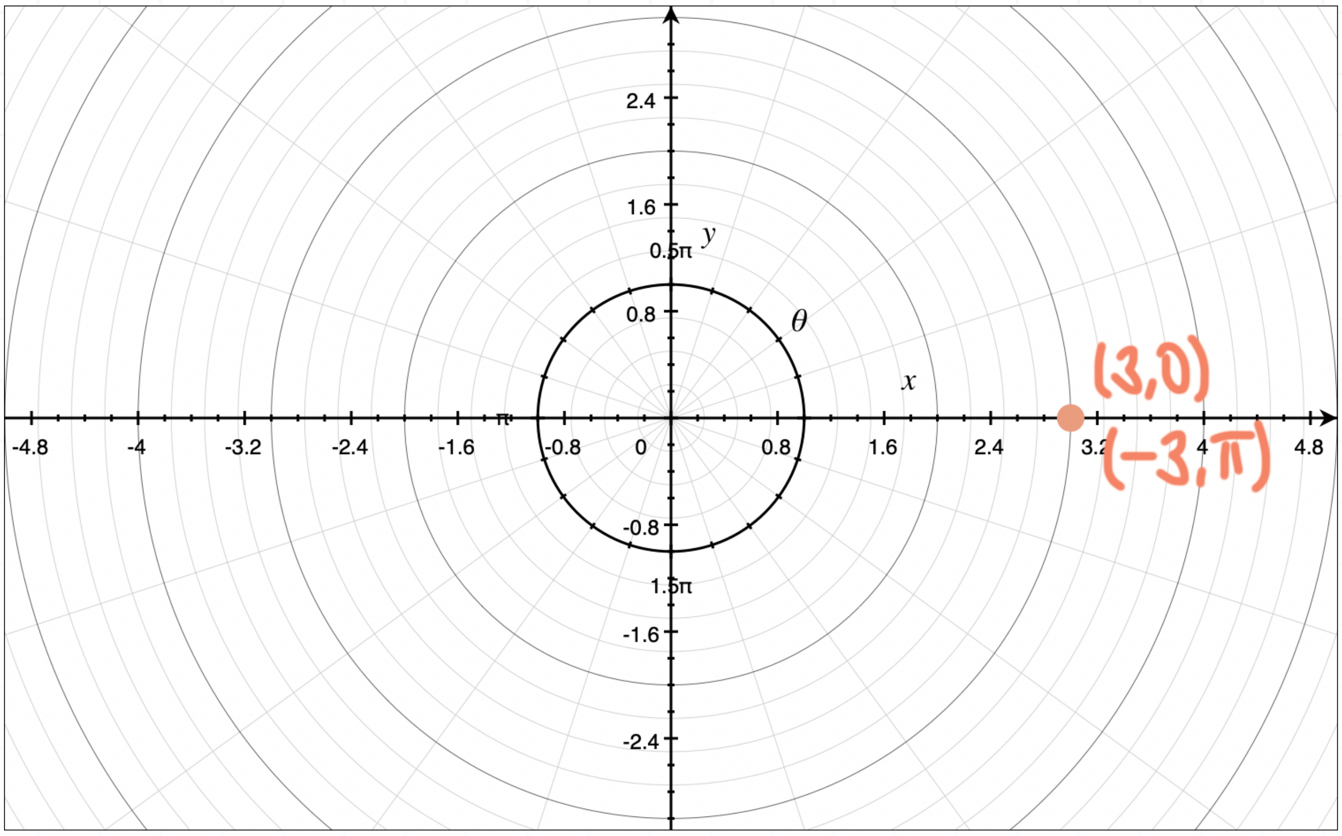
We also want to be able to find the points at which two polar curves intersect one another. We did the same thing with linear equations back in Algebra.

In both cases, with linear equations in Algebra, and now here with polar equations in Precalculus, we'll set the curves equal to each other in order to solve for points of intersection.

"Hidden" points of intersection

But we have to be careful. The points that satisfy both polar equations aren't necessarily the only points of intersection. Because we can represent the same polar point multiple ways (as we learned earlier), it's possible for two polar curves to intersect one another at the same location in space, but at different values of r and θ.

For example, we know that $(3,0)$ and $(-3,\pi)$ represent the same point in polar space.



Let's say that we have one polar curve that passes through $(3,0)$, and another polar curve that passes through $(-3,\pi)$. When we set those two polar curves equal to one another, we won't find this point of intersection. That's because one curve satisfies $(3,0)$ but not $(-3,\pi)$, while the other curve satisfies $(-3,\pi)$ but not $(3,0)$. So even though both $(3,0)$ and $(-3,\pi)$ represent the same point in space, because they're represented differently, the intersection point won't show up when we set the curves equal to each other.

Therefore, our strategy for finding points of intersection for two polar curves will be to start by setting the curves equal to each other and solving for coordinate points that satisfy both equations.

But then we'll follow that up by sketching both curves to look for other "hidden" points of intersection that we might have missed. We'll pay special attention to the pole $(0,0)$, which is very common hidden point of intersection.

If we find any hidden points of intersection, then we can use any representation of that point that we choose. In other words, in our example from earlier, we could choose either $(3,0)$ or $(-3,\pi)$ as the representation of the intersection. If possible, we try to choose a representation with an angle θ in the principal interval $[0,2\pi)$, and a positive value of r. So we'd prefer $(3,0)$ over $(-3,\pi)$, since r is positive in $(3,0)$.

Let's do an example where we find the intersection points of two circles.

Example

Find the points of intersection of $r = 2$ and $r = -4\sin\theta$.

Because both equations are equal to r, we can set their right sides equal to one another.

$$2 = -4\sin\theta$$

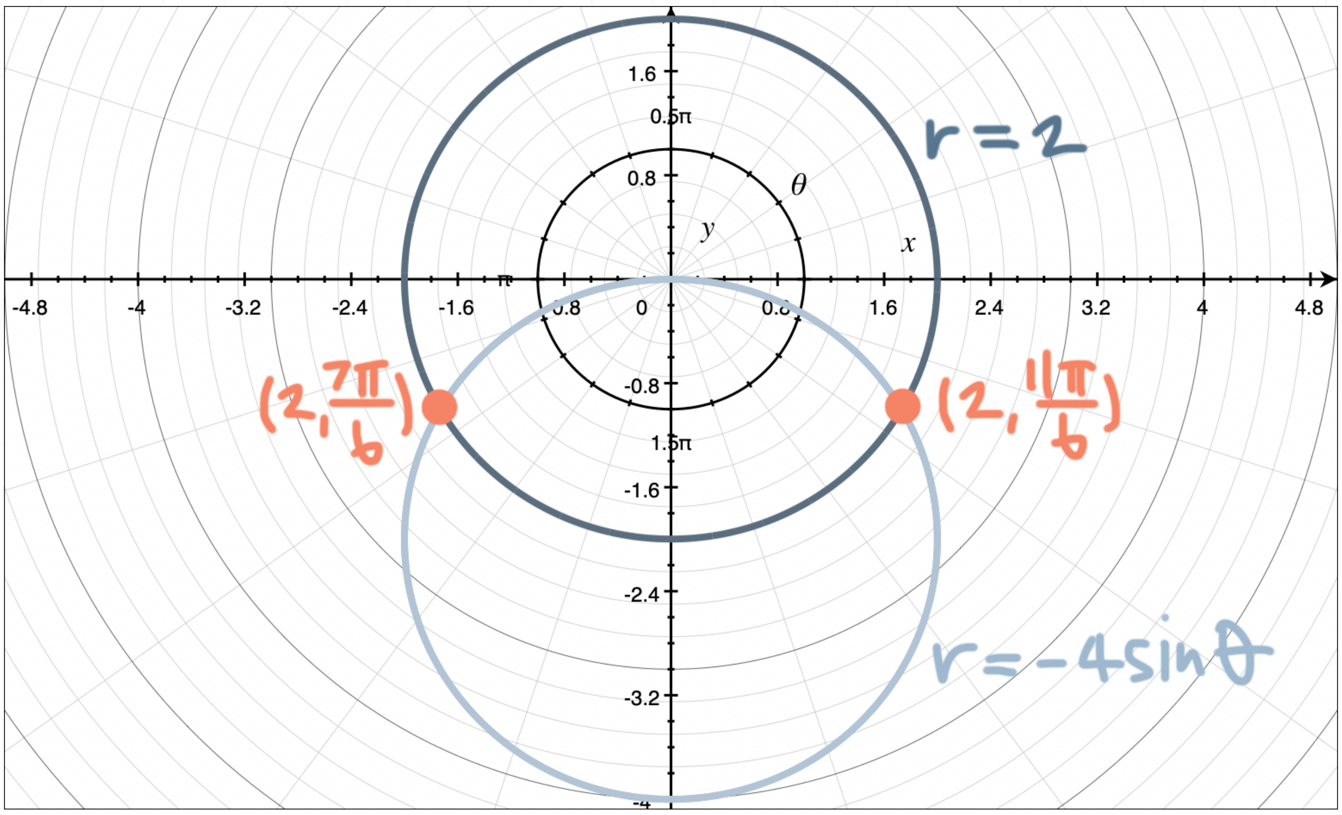
$$\sin\theta = -\frac{1}{2}$$

From the unit circle from Trigonometry, we know that sine has a value of $-1/2$ at $\theta = 7\pi/6$ and $\theta = 11\pi/6$. If we plug these angles back into either of the polar equations, we get $r = 2$ at both angles. Which means the points of intersection for these curves are

$$(r,\theta) = \left(2,\frac{7\pi}{6}\right)$$

$$(r, \theta) = \left(2, \frac{11\pi}{6}\right)$$

If we sketch the graph of both curves on the same set of axes, we can see these points of intersection, and confirm that there aren't any other points of intersection of than these two.



Let's do another example, this time with a circle and a limaçon.

Example

Find the points of intersection of $r = 2$ and $r = 2 - 4\cos\theta$.

In the limaçon equation, we have $a=2$ and $b=4$, which means $a/b<1$, which tells us that the limaçon has a small loop. We'll start by setting the curves equal to one another.

$$2=2-4\cos\theta$$

$$0=-4\cos\theta$$

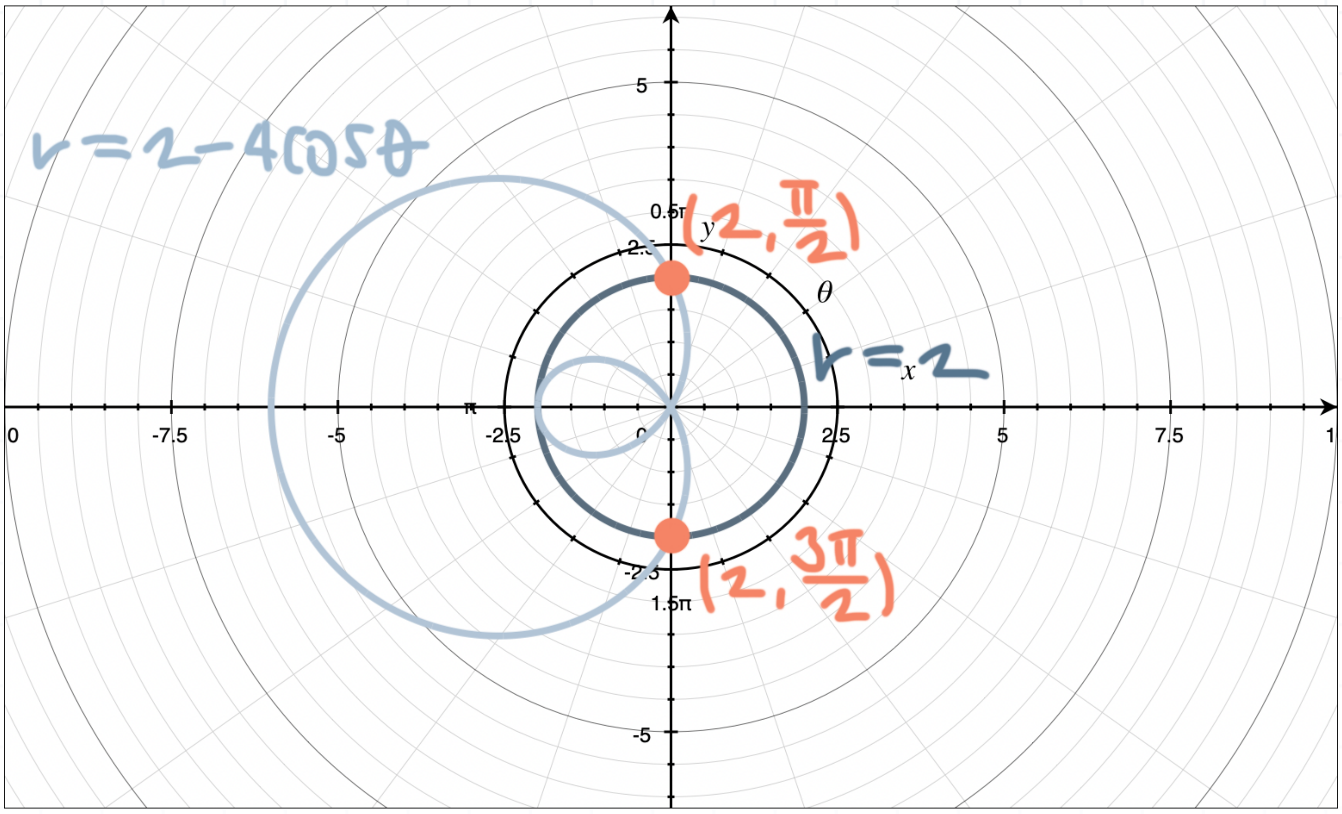
$$\cos\theta=0$$

From the unit circle from Trigonometry, we know that cosine has a value of 0 at $\theta=\pi/2$ and $\theta=3\pi/2$. If we plug these angles back into either of the polar equations, we get $r=2$ at both angles. Which means the points of intersection for these curves are

$$(r,\theta)=\left(2,\frac{\pi}{2}\right)$$

$$(r,\theta)=\left(2,\frac{3\pi}{2}\right)$$

If we sketch the graph of both curves on the same set of axes, we can see these points of intersection.



What we notice when we sketch the curves is that we actually have a third "hidden" point of intersection at $(2,\pi)$ that we didn't uncover when we set the curves equal to each other and solved for θ.

That's because the circle reaches that third point at $(r,\theta)=(2,\pi)$, while the limaçon reaches that point at $(r,\theta)=(-2,0)$. We can pick either representation of this third point; we'll choose $(2,\pi)$ because it includes a positive value of r, so the points of intersection of the curves are

$$(r,\theta)=\left(2,\frac{\pi}{2}\right)$$

$$(r,\theta)=\left(2,\frac{3\pi}{2}\right)$$

$$(r,\theta)=(2,\pi)$$

Let's look at an example with a circle and a cardioid, where we again have "hidden" intersections.

Example

Find the points of intersection of $r = 3\sin\theta$ and $r = 1 + \sin\theta$.

The cardioid is a sine cardioid with a positive sign between the terms, which means its graph is symmetric about the vertical axis and will sit mostly above the horizontal axis. We'll start by setting the curves equal to one another.

$$3\sin\theta = 1 + \sin\theta$$

$$2\sin\theta = 1$$

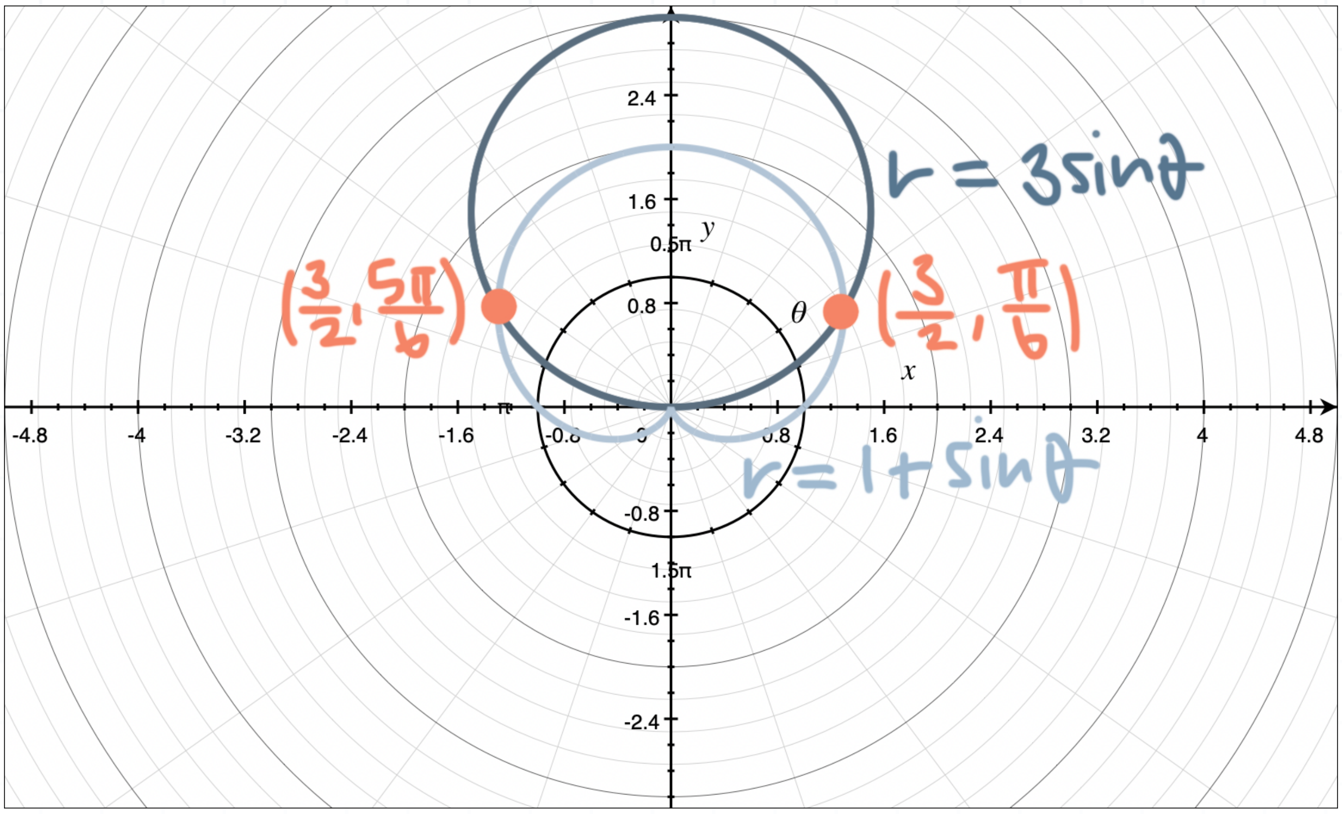
$$\sin\theta = \frac{1}{2}$$

From the unit circle from Trigonometry, we know that sine has a value of $1/2$ at $\theta = \pi/6$ and $\theta = 5\pi/6$. If we plug these angles back into either of the polar equations, we get $r = 3/2$ at both angles. Which means the points of intersection for these curves are

$$(r, \theta) = \left(\frac{3}{2}, \frac{\pi}{6}\right)$$

$$(r, \theta) = \left(\frac{3}{2}, \frac{5\pi}{6}\right)$$

If we sketch the graph of both curves on the same set of axes, we can see these points of intersection.



What we notice when we sketch the curves is that we actually have a third "hidden" point of intersection at $(0,0)$ that we didn't uncover when we set the curves equal to each other and solved for θ.

That's because the circle reaches that third point at $(r, \theta) = (0,0)$, while the cardioid reaches that point at $(r, \theta) = (0, 3\pi/2)$. We can pick either representation of this third point; we'll choose $(0,0)$, so the points of intersection of the curves are

$$(r, \theta) = \left(\frac{3}{2}, \frac{\pi}{6}\right)$$

$$(r, \theta) = \left(\frac{3}{2}, \frac{5\pi}{6}\right)$$

$$(r, \theta) = (0,0)$$

Finally, let's look at the intersection of a rose and a lemniscate.

Example

Find the points of intersection of $r = 3\cos(2\theta)$ and $r^2 = 9\cos(2\theta)$.

The rose has $|2n| = |2(2)| = 4$ petals that extend out to a distance of $c = 3$ from the origin. Because the equation of the rose is given for r, while the equation of the lemniscate is given for r^2, we'll square the rose equation,

$$r = 3\cos(2\theta)$$

$$r^2 = (3\cos(2\theta))^2$$

$$r^2 = 9\cos^2(2\theta)$$

then set the curves equal to one another.

$$9\cos^2(2\theta) = 9\cos(2\theta)$$

$$\cos^2(2\theta) = \cos(2\theta)$$

$$\cos^2(2\theta) - \cos(2\theta) = 0$$

$$\cos(2\theta)(\cos(2\theta) - 1) = 0$$

Apply the Zero Theorem from Algebra to create two equations that we can solve individually. We get

$$\cos(2\theta) = 0$$

$$2\theta = \frac{\pi}{2}, \frac{3\pi}{2}, \frac{5\pi}{2}, \frac{7\pi}{2}, \ldots$$

$$\theta = \frac{\pi}{4}, \frac{3\pi}{4}, \frac{5\pi}{4}, \frac{7\pi}{4}, \ldots$$

and

$$\cos(2\theta) - 1 = 0$$

$$\cos(2\theta) = 1$$

$$2\theta = 0, 2\pi, 4\pi, 6\pi, \ldots$$

$$\theta = 0, \pi, 2\pi, 3\pi, \ldots$$

Combining these angle sets into one gives the complete set of angles that satisfy both equations.

$$\theta = 0, \frac{\pi}{4}, \frac{3\pi}{4}, \pi, \frac{5\pi}{4}, \frac{7\pi}{4}, 2\pi, \ldots$$

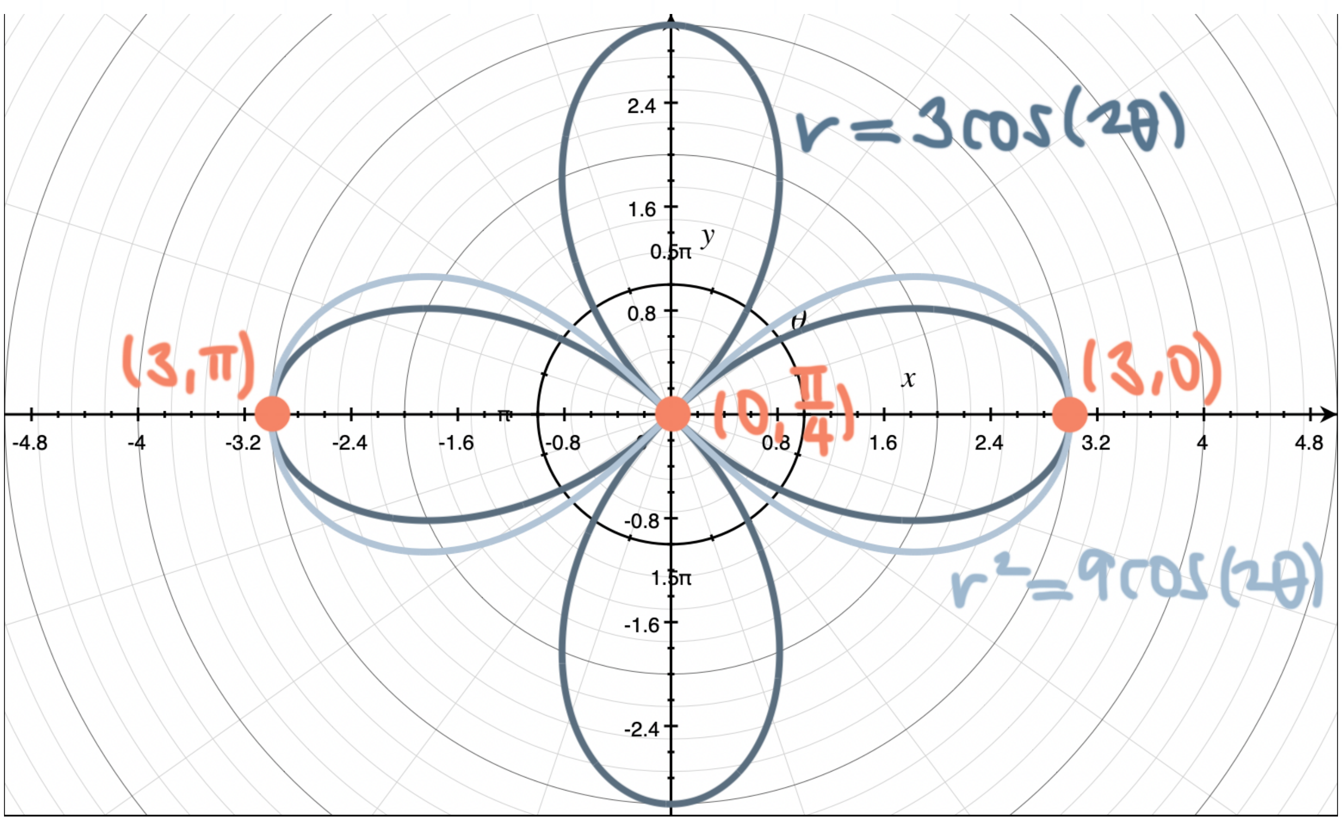
If we plug these angles back into both polar equations, we get

	$r = 3\cos(2\theta)$	$r^2 = 9\cos(2\theta)$
$\theta = 0$	$r = 3$	$r = \pm 3$
$\theta = \pi/4$	$r = 0$	$r = 0$
$\theta = 3\pi/4$	$r = 0$	$r = 0$
$\theta = \pi$	$r = 3$	$r = \pm 3$

$\theta = 5\pi/4$	$r = 0$	$r = 0$
$\theta = 7\pi/4$	$r = 0$	$r = 0$
$\theta = 2\pi$	$r = 3$	$r = \pm 3$

We have points of intersection only where we have matching values of r. We get $r = 3$ from both curves at $\theta = 0$, so $(r, \theta) = (3,0)$ is a point of intersection. We get $r = 0$ from both curves at $\theta = \pi/4$, so $(r, \theta) = (0,\pi/4)$ is a second point of intersection. We get $r = 3$ from both curves at $\theta = \pi$, so $(r, \theta) = (3,\pi)$ is a third point of intersection.

The table appears to include many more points of intersection, but any other matching values we found will just give us duplicates of the same three points, $(3,0)$, $(0,\pi/4)$, and $(3,\pi)$. If we sketch the graph of both curves on the same set of axes, we can see these points of intersection.



Complex numbers

We know from Algebra about the Cartesian (or rectangular) coordinate system, and we've learned now also about the polar coordinate system.

The reason we have different systems like these are because they allow us to solve problems that we wouldn't otherwise be able to if we were using just one system alone. Different systems can also make it easier to solve a particular kind of problem.

Imaginary numbers

Now we want to introduce the **complex number system**, which is based on the **imaginary unit** i. This imaginary number is defined as

$$i^2 = -1 \text{ or } i = \sqrt{-1}$$

Without imaginary numbers, we have no way to find the value of x in an equation like $x^2 = -16$, because we'd take the square root of both sides to get

$$x = \pm\sqrt{-16}$$

and then we'd be stuck, since the square root of a negative number isn't defined in the real number system. But if we use imaginary numbers, we're able to simplify this value of x as

$$x = \pm\sqrt{-16}$$

$$x = \pm\sqrt{16(-1)}$$

$$x = \pm \sqrt{16}\sqrt{-1}$$

$$x = \pm 4i$$

Simplifying powers of imaginary numbers

A natural extension of the definition of the imaginary number i is that the powers of i follow a predictable, cyclical pattern. We already know that any non-zero value raised to the power of 0 is 1, so $i^0 = 1$. Putting that together with the definition of i, we get

$$i^0 = 1$$

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

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

$$i^3 = i \cdot i^2 = -\sqrt{-1} = -i$$

$$i^4 = i^2 \cdot i^2 = (-1)(-1) = 1$$

$$i^5 = i \cdot i^4 = \sqrt{-1}(1) = \sqrt{-1}$$

$$i^6 = i^2 \cdot i^4 = (-1)(1) = -1$$

$$i^7 = i^3 \cdot i^4 = -i \cdot 1 = -i$$

This $\sqrt{-1}$, -1, $-i$, 1 pattern repeats over an over again, no matter how large we make the exponent on the imaginary number. Because the pattern is predictable, we can simplify any power of i just by pulling out the largest power that's divisible by 4. For instance,

$$i^{202}$$

$$i^{200} \cdot i^2$$

$$(i^4)^{50} \cdot i^2$$

$$(1)^{50} \cdot i^2$$

$1 \cdot i^2$

i^2

-1

A similar pattern also holds for imaginary numbers with negative exponents. As long as we rationalize the fraction whenever we're left with an imaginary number in the denominator, we get

$$i^{-1} = \frac{1}{i^1} = \frac{1}{i} = \frac{1}{i}\left(\frac{i}{i}\right) = \frac{i}{i^2} = \frac{i}{-1} = -i$$

$$i^{-2} = \frac{1}{i^2} = \frac{1}{-1} = -1$$

$$i^{-3} = \frac{1}{i^3} = \frac{1}{-i} = -\frac{1}{i} = -\frac{1}{i}\left(\frac{i}{i}\right) = -\frac{i}{i^2} = -\frac{i}{-1} = i$$

$$i^{-4} = \frac{1}{i^4} = \frac{1}{1} = 1$$

Complex numbers

We've defined imaginary numbers, and we already know about real numbers from previous math classes. The interesting thing is that complex numbers actually include all of the real numbers *and* all of the imaginary numbers. In other words, all real numbers are complex numbers, and all imaginary numbers are complex numbers. Let's talk about why.

The standard form of a **complex number** is $z = a + bi$, where a and b are real numbers. If a and b are real numbers, then we can see that a complex number is always the sum of the real number a, and the imaginary number bi.

The **real part** of the complex number is a, and we describe that as Re(z). The **imaginary part** of the complex number is b, and we describe that as Im(z). These are all examples of complex numbers:

$$z = 4 + i$$

$$z = -2 + 3i$$

$$z = \pi - 6i$$

$$z = \sqrt{2} - ei$$

In each of these complex numbers, both the real part and the imaginary part are non-zero. If the real-number part of a complex number is 0, then we end up with

$$z = a + bi$$

$$z = 0 + bi$$

$$z = bi$$

Since we're left with only the imaginary part, we call $z = bi$ a **pure imaginary** number. But if instead the imaginary part of a complex number is 0, then we end up with

$$z = a + bi$$

$$z = a + 0i$$

$$z = a$$

Since we're left with only the real part, we call $z = a$ just a **real number**. But notice that in both cases we started with $z = a + bi$. Therefore,

1. all real numbers and all imaginary numbers are complex numbers,

2. in the special case where $a = 0$, the complex number is a pure imaginary number, and

3. in the special case where $b = 0$, the complex number is a real number.

Complex number operations

Just like with real numbers, we want to be able to add, subtract, multiply, and divide complex numbers. So in this lesson, we'll talk about how handle arithmetic with complex numbers.

Adding and subtracting complex numbers

To add two complex numbers, we add their real parts, and then separately add their imaginary parts. There's no logical way to combine the real and imaginary parts themselves, so the result is still a complex number.

Example

Find the sum of the complex numbers $-2+17i$ and $6-8i$.

Add the real parts together, and then separately add the imaginary parts.

$$(-2+17i)+(6-8i)$$

$$(-2+6)+(17i-8i)$$

$$(-2+6)+(17-8)i$$

$$4+9i$$

Just like complex number addition, to subtract two complex numbers, subtract their real parts and their imaginary parts separately.

Example

Find the difference of the complex numbers $13 + 4i$ and $5 + 9i$.

Taking the difference of their real parts and imaginary parts separately, we get

$$(13 + 4i) - (5 + 9i)$$

$$(13 - 5) + (4i - 9i)$$

$$(13 - 5) + (4 - 9)i$$

$$8 - 5i$$

Let's do another example with both addition and subtraction.

Example

Find the sum and difference of $21 + 16i$ and $-9 + 4i$.

The sum of the complex numbers is

$$(21 + 16i) + (-9 + 4i)$$

$$(21 - 9) + (16i + 4i)$$

$$(21 - 9) + (16 + 4)i$$

$$12 + 20i$$

Their difference is

$$(21 + 16i) - (-9 + 4i)$$

We have to remember to be careful here with the subtraction in the middle, and distribute it across both parts of the second complex number.

$$21 + 16i + 9 - 4i$$

$$(21 + 9) + (16i - 4i)$$

$$30 + 12i$$

Multiplying and dividing complex numbers

In Algebra, we learned to use FOIL to multiply two binomials by multiplying their First terms, Outer terms, Inner terms, and Last terms, and then adding those products. Complex numbers are binomials by definition, $z = a + bi$, so we'll use FOIL to multiply them.

Here's what happens when we multiply the real and imaginary parts of complex numbers:

Real $\times$ real $=$ real $\qquad\qquad 3 \times 4 = 12$

Real $\times$ imaginary = imaginary $\quad 3 \times 4i = 12i$

Imaginary $\times$ imaginary = real $\quad 3i \times 4i = 12i^2 = 12(-1) = -12$

For instance, the product of 25 and $8i$ is

$$(25)(8i)$$

$$(25 \cdot 8)i$$

$$200i$$

And the product of i and $-i$ is

$$(i)(-i)$$

$$-(i^2)$$

$$-(-1)$$

$$1$$

We can also multiply complex numbers by 0, and just like real numbers, the result will be 0. The product of $-36 + 11i$ and 0 is

$$(-36 + 11i)(0)$$

$$(-36)(0) + (11i)(0)$$

$$0 + (11)(0)i$$

$$0 + 0i$$

$$0$$

Let's do an example of multiplying two complex numbers as binomials $z = a + bi$ to see how to use FOIL to find the product.

Example

Find the product of $-6 + 2i$ and $4 - 5i$.

Using FOIL, we'll multiply the first, outer, inner, and last terms.

$$(-6 + 2i)(4 - 5i)$$

$$(-6)(4) + (-6)(-5i) + (2i)(4) + (2i)(-5i)$$

$$-24 + 30i + 8i - 10i^2$$

$$-24 + 38i - 10i^2$$

Because $i^2 = -1$, we get

$$-24 + 38i - 10(-1)$$

$$-24 + 38i + 10$$

$$-14 + 38i$$

When we divide a complex number by a real number, we'll get

$$\frac{a + bi}{c} \quad \to \quad \frac{a}{c} + \frac{bi}{c} \quad \to \quad \frac{a}{c} + \frac{b}{c}i$$

After the division, a/c is now the real part of the complex number, and b/c is the imaginary part.

When we divide a complex number by a pure imaginary number

$$\frac{a+bi}{ci} \quad \to \quad \frac{a}{ci}+\frac{bi}{ci} \quad \to \quad \frac{a}{c}i^{-1}+\frac{b}{c} \quad \to$$

$$\frac{a}{c}(-i)+\frac{b}{c} \quad \to \quad -\frac{a}{c}i+\frac{b}{c} \quad \to \quad \frac{b}{c}-\frac{a}{c}i$$

After the division, b/c is now the real part of the complex number, and $-a/c$ the imaginary part.

But what about when we divide the full complex binomial $z=a+bi$ by another complex binomial $z=c+di$?

$$\frac{a+bi}{c+di}$$

This is where conjugates come in. The **conjugate** of a binomial is the same two terms in the binomial, but with the opposite sign between them. The **complex conjugate** is the same thing, but when the binomial is a complex number. For example, the complex conjugate of $3+4i$ is $3-4i$, the complex conjugate of $11-6i$ is $11+6i$, and the complex conjugate of $a+bi$ is $a-bi$.

Example

Find the complex conjugate of each of the complex numbers.

$$13+5i$$

$$7-4i$$

$-6+i$

For each of these, we keep the real part (13, 7, or -6) and change the sign of the imaginary part (from 5 to -5, from -4 to 4, or from 1 to -1):

The complex conjugate of $13+5i$ is $13-5i$.

The complex conjugate of $7-4i$ is $7+4i$.

The complex conjugate of $-6+i$ is $-6-i$.

To simplify the quotient of two complex numbers, we need to multiply both the numerator and denominator of the fraction by the complex conjugate of the denominator.

$$\frac{a+bi}{c+di}\left(\frac{c-di}{c-di}\right)$$

$$\frac{(a+bi)(c-di)}{(c+di)(c-di)}$$

Then we FOIL the numerator, and separately FOIL the denominator.

$$\frac{ac-adi+bci-bdi^2}{c^2-cdi+cdi-d^2i^2}$$

When we multiply by the complex conjugate like this, the two terms in the middle of the denominator will always cancel with one another.

$$\frac{ac - adi + bci - bdi^2}{c^2 - d^2i^2}$$

$$\frac{ac - adi + bci - bd(-1)}{c^2 - d^2(-1)}$$

$$\frac{ac - adi + bci + bd}{c^2 + d^2}$$

$$\frac{(ac + bd) + (-ad + bc)i}{c^2 + d^2}$$

$$\frac{(ac + bd) + (bc - ad)i}{c^2 + d^2}$$

$$\frac{ac + bd}{c^2 + d^2} + \frac{bc - ad}{c^2 + d^2}i$$

After reworking this expression, we just find another complex number in standard form. The first fraction is the real number part, and the second fraction is the imaginary part. So when we divide two complex number binomials, the result will be a complex number binomial.

Example

Find the quotient of $7 + 3i$ and $9 - 4i$.

First, we'll write the quotient as

$$\frac{7 + 3i}{9 - 4i}$$

and then multiply both the numerator and denominator by the complex conjugate of $9-4i$, which is $9+4i$.

$$\frac{7+3i}{9-4i}\left(\frac{9+4i}{9+4i}\right)$$

$$\frac{(7+3i)(9+4i)}{(9-4i)(9+4i)}$$

FOIL across the numerator and denominator, then simplify.

$$\frac{63+28i+27i+12i^2}{81+36i-36i-16i^2}$$

$$\frac{63+55i+12i^2}{81-16i^2}$$

$$\frac{63+55i+12(-1)}{81-16(-1)}$$

$$\frac{63+55i-12}{81+16}$$

$$\frac{51+55i}{97}$$

Split the fraction to write the quotient in standard form.

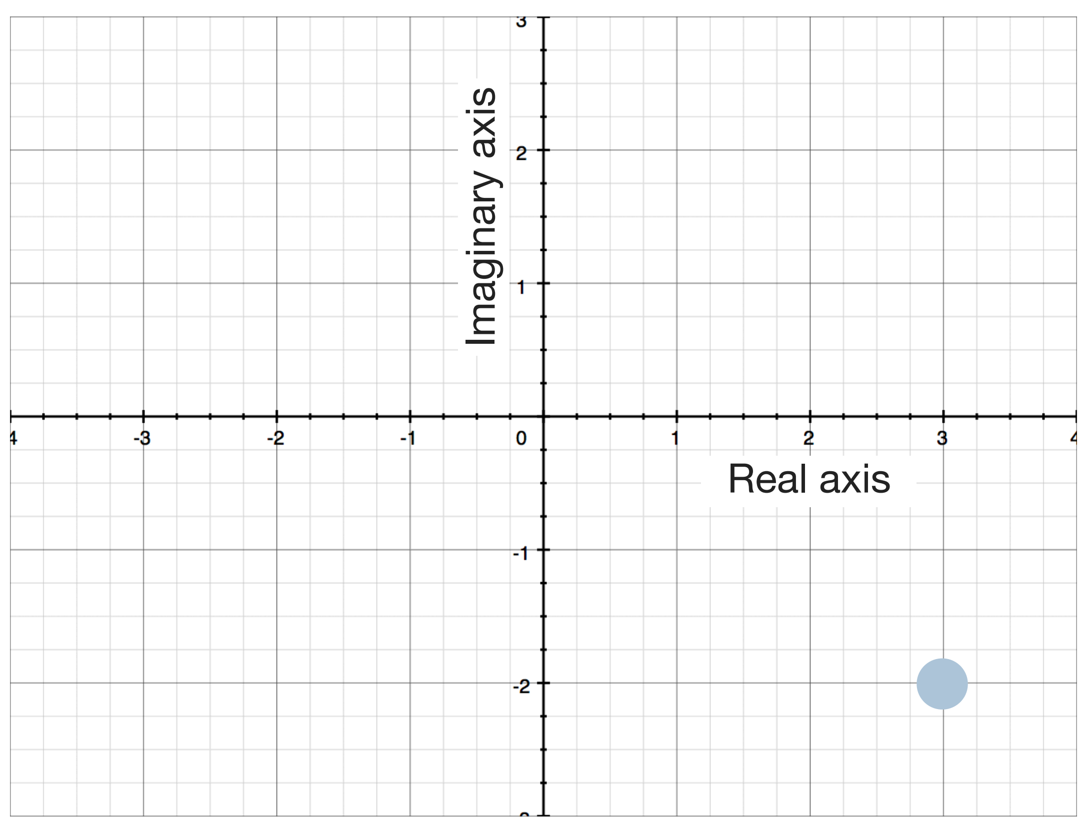
$$\frac{51}{97}+\frac{55}{97}i$$

Graphing complex numbers

Let's start with a little review. When we graph a rectangular coordinate point (x, y), we graph it in the xy-plane, also called the rectangular coordinate plane, or Cartesian coordinate plane. The x-value is represented along the horizontal axis, and the y-value is represented along the vertical axis.

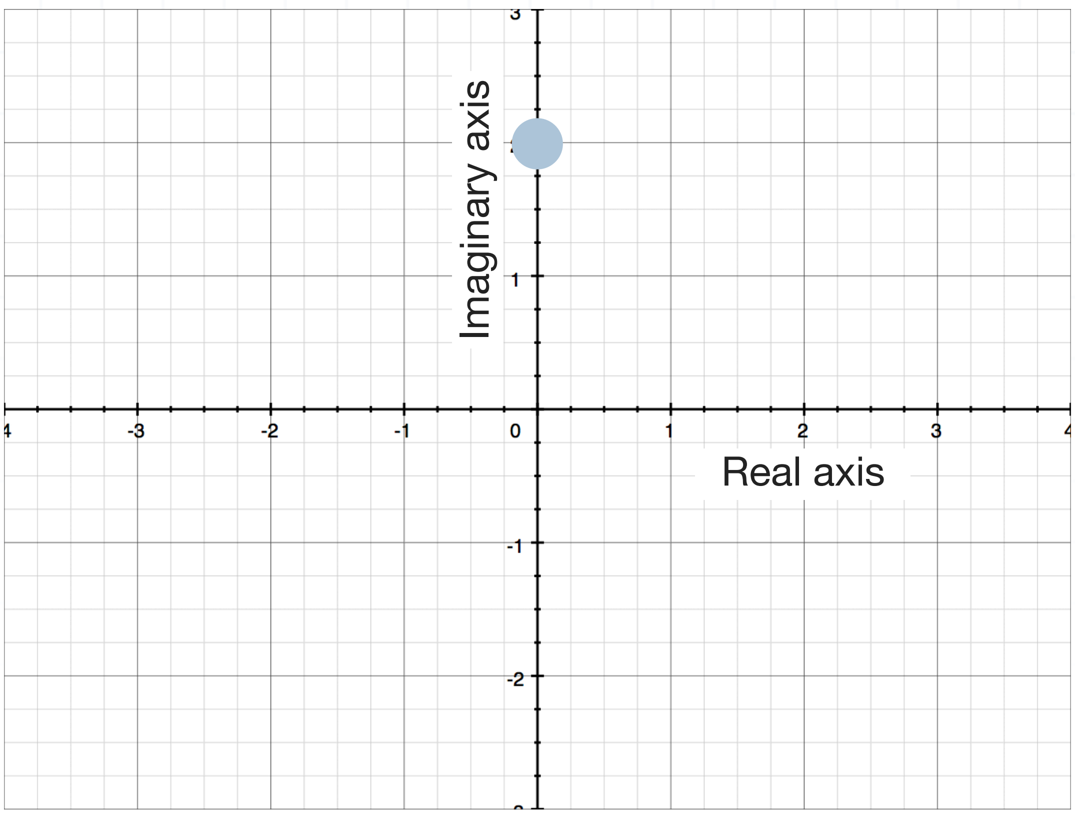
Complex numbers can't be graphed in the same xy-plane. Instead, we graph them in the complex plane. The **complex plane** has horizontal and vertical axes like the xy-plane, but the horizontal axis represents the real part of the complex number, and the vertical axis represents the imaginary part of the complex number.

So the complex number $z = 3 - 2i$ with real part 3 and imaginary part -2 is plotted as

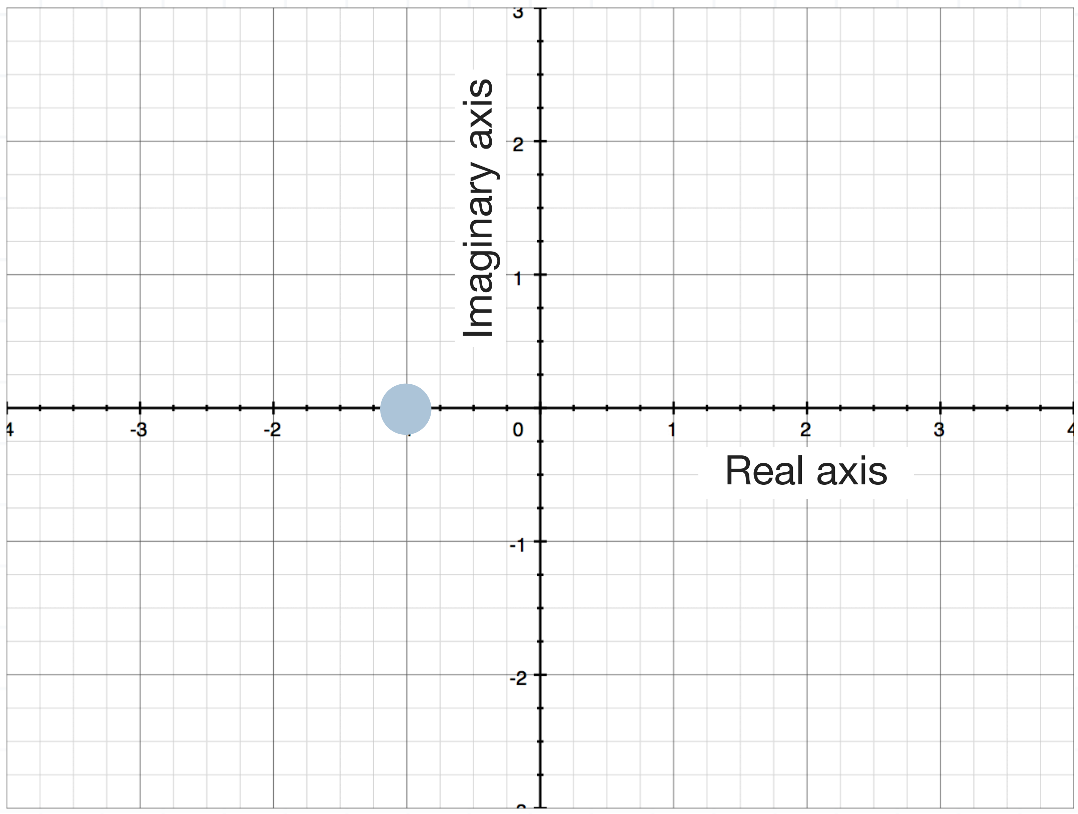


because a real part of $+3$ means we move to the right along the horizontal axis, and because an imaginary part of -2 means we move down along the vertical axis.

If the real part of the complex number is 0, such that we have the pure imaginary number $z = 0 + bi = bi$, then the complex number will lie on the vertical axis in the complex plane. For instance, $z = 2i$ is plotted as



Or if the imaginary part of the complex number is 0, such that we have the real number $z = a + 0i = a$, then the complex number will lie on the real axis in the complex plane. For instance, $z = -1$ is plotted as



Let's do a few more examples.

Example

Plot $3 + 2i$, $-1 - 3i$, and $1.5i$ in the complex plane.

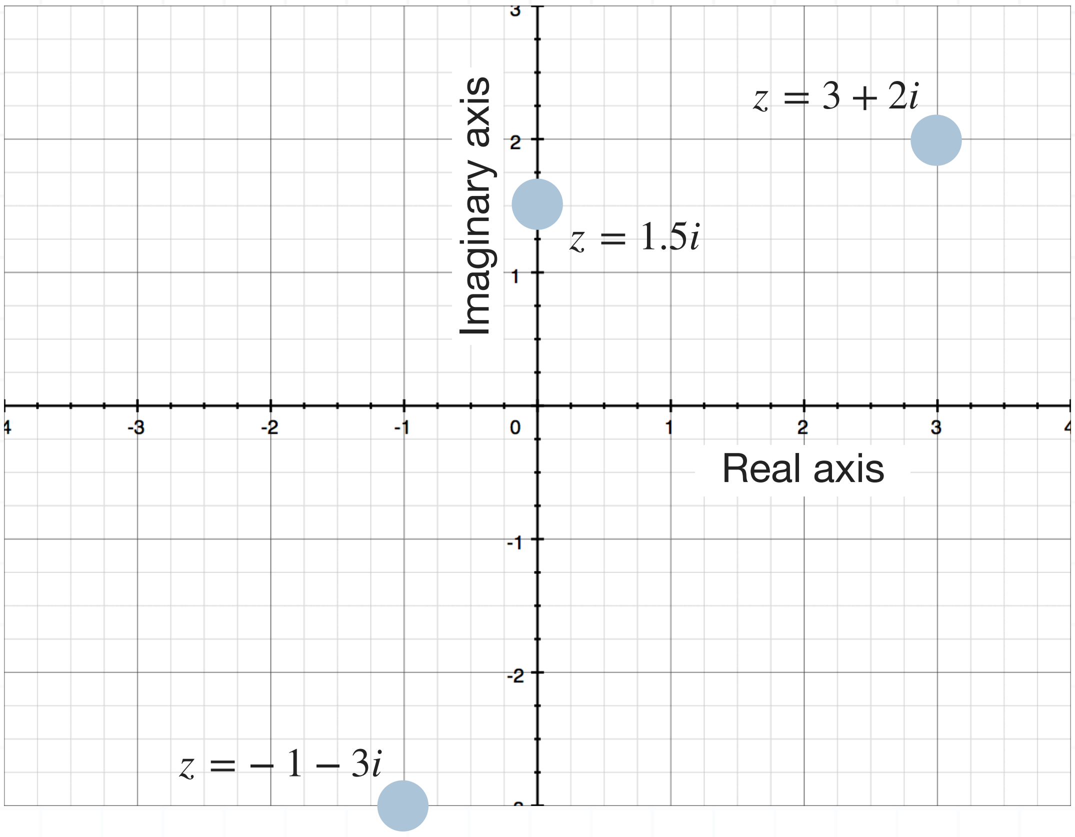
Let's break down each complex number.

For $3 + 2i$, the real part is $a = 3$ and the imaginary part is $b = 2$.

For $-1 - 3i$, the real part is $a = -1$ and the imaginary part is $b = -3$.

For $1.5i$, the real part is $a = 0$ and the imaginary part is $b = 1.5$.

Now we can plot all three of them in the same complex plane.



Plotting the sum or difference of complex numbers

Remember that when we add two complex numbers, we just add the real parts, and then separately add the imaginary parts. So the result is still a complex number.

Therefore, if we're asked to graph the sum of a complex number, we'll find the sum of the complex numbers first, and then plot the resulting sum.

Example

Graph the sum of the complex numbers $-4 + 3i$ and $2 - 6i$.

First, find the sum.

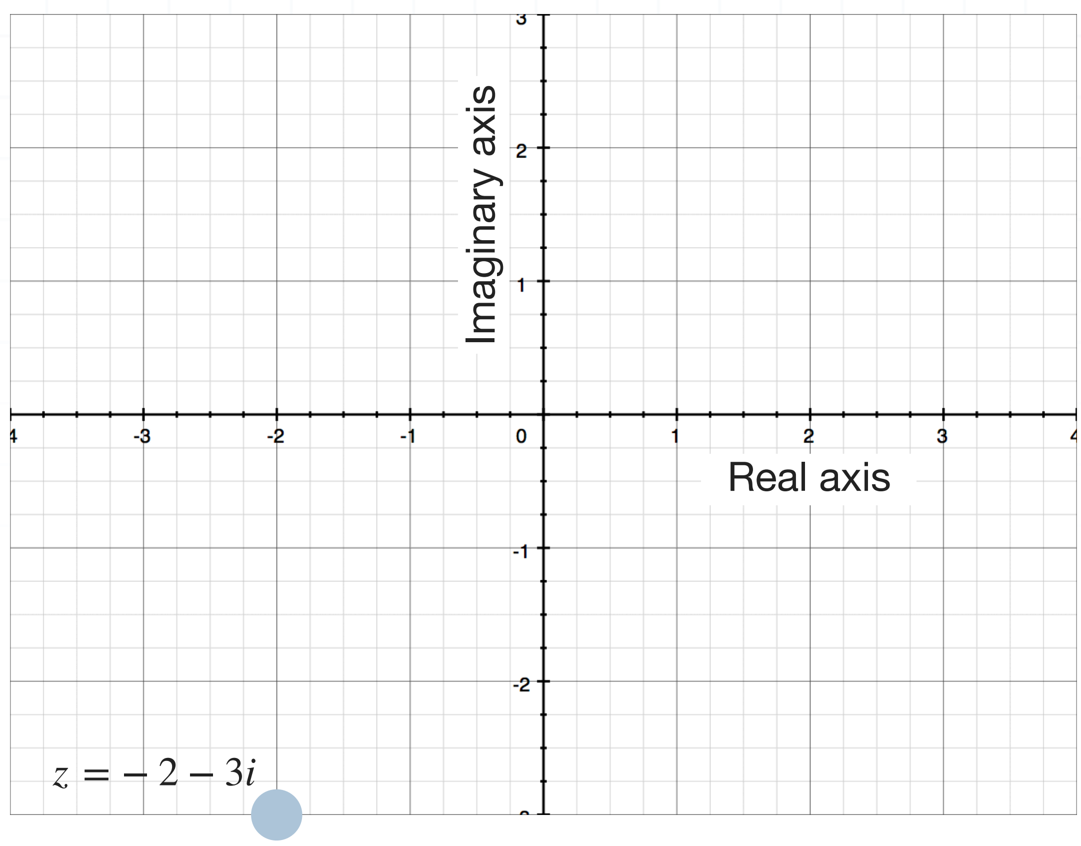
$$(-4+3i)+(2-6i)$$

$$(-4+2)+(3+(-6))i$$

$$(-4+2)+(3-6)i$$

$$-2-3i$$

Now plot the complex number $-2-3i$, which has real part -2 and imaginary part -3.



In the same way, to graph the difference of two complex numbers, find the difference first, and then plot the difference.

Example

Graph the difference of the complex numbers 1 and $5-3i$.

Write 1 as $1 + 0i$. Then the difference of the complex numbers is

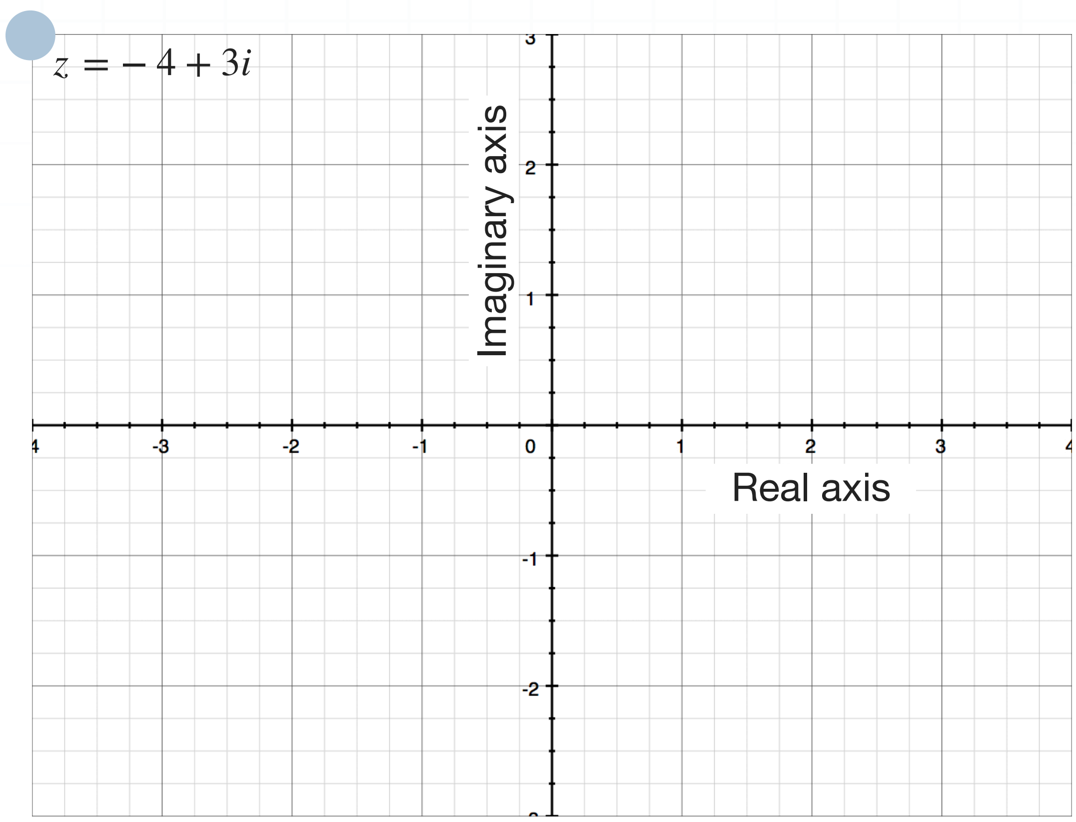
$$(1 + 0i) - (5 - 3i)$$

$$(1 - 5) + (0 - (-3))i$$

$$(-4) + (0 + 3)i$$

$$-4 + 3i$$

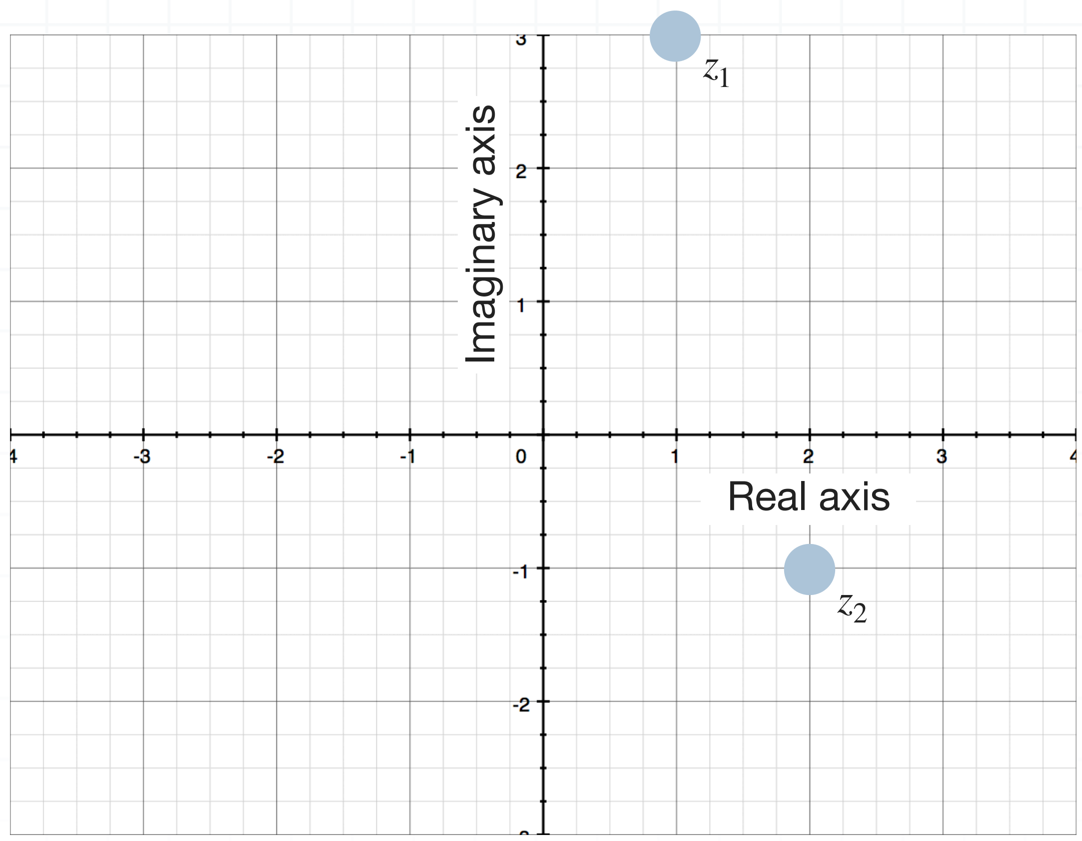
Now plot the complex number $-4 + 3i$, which has real part -4 and imaginary part of 3.



We can also work backwards. If we have imaginary numbers graphed in the complex plane, we can break them down into their real and imaginary parts, then find the sum or difference, and plot those.

Example

Graph the sum of the complex numbers z_1 and z_2.



The point z_1 is 1 unit to the right of the vertical axis and 3 units above the horizontal axis, which means that complex number is $z_1 = 1 + 3i$.

The point z_2 is 2 units to the right of the vertical axis and 1 unit below the horizontal axis, which means that complex number is $z_2 = 2 - 1i = 2 - i$.

The sum of z_1 and z_2 is

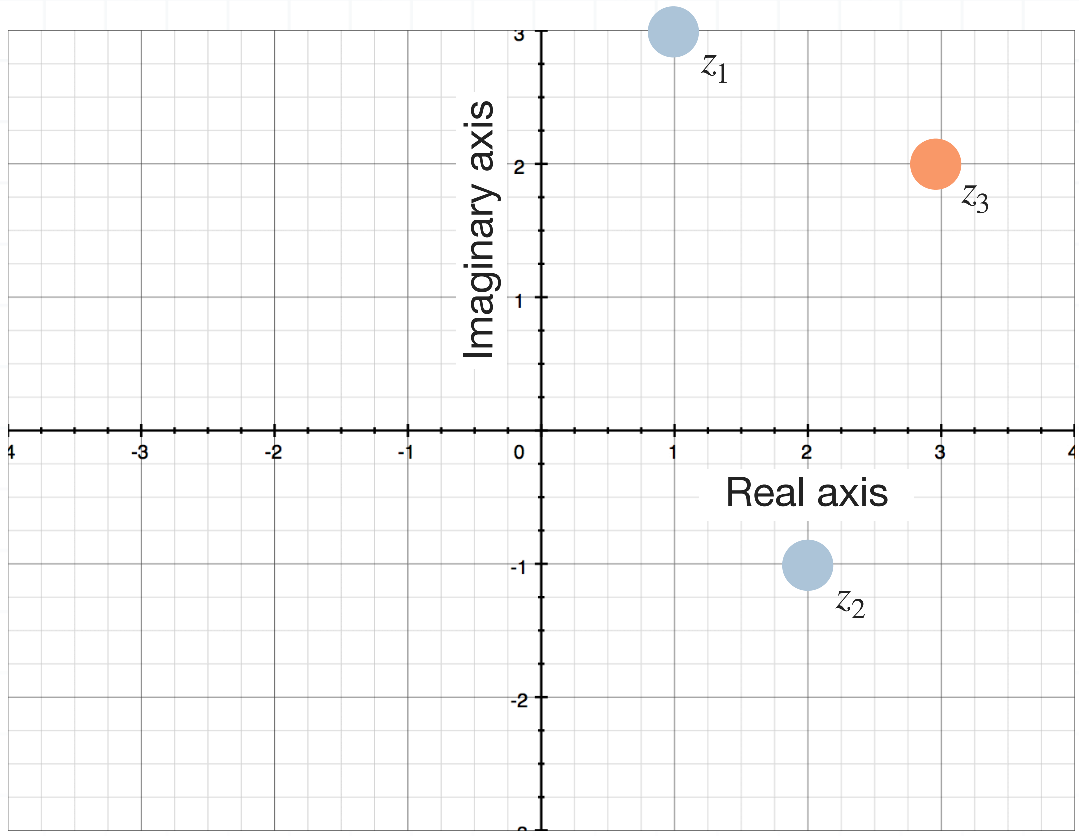
$$z_1 + z_2 = (1 + 3i) + (2 - i)$$

$$z_1 + z_2 = (1 + 2) + (3 + (-1))i$$

$$z_1 + z_2 = (1 + 2) + (3 - 1)i$$

$$z_1 + z_2 = 3 + 2i$$

So if we plot the sum on the same set of axes, we get



Distances and midpoints

We've learned in the past how to find the distance between two points in rectangular coordinate space. We can do it algebraically or graphically.

Algebraically, given the points $(3,2)$ and $(-2,-2)$, the distance between them can be found using the **distance formula**.

$$d=\sqrt{(x_1-x_2)^2+(y_1-y_2)^2}$$

$$d=\sqrt{(3-(-2))^2+(2-(-2))^2}$$

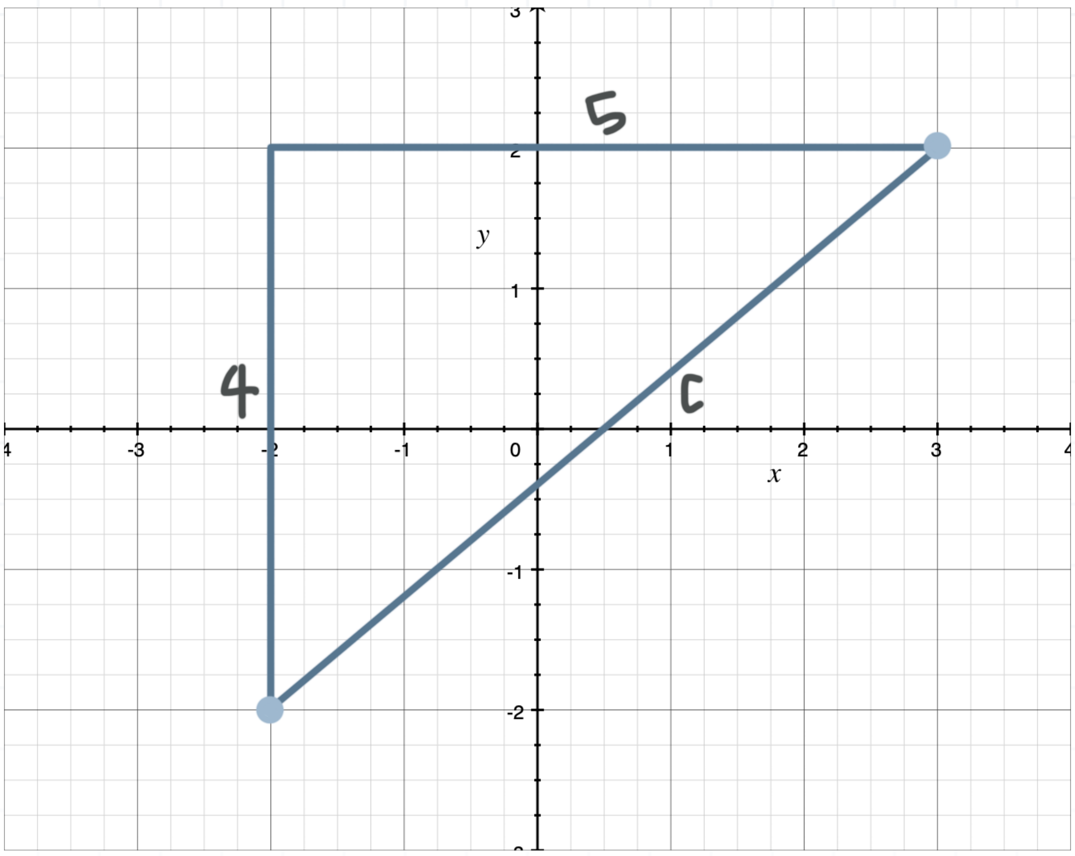
$$d=\sqrt{(3+2)^2+(2+2)^2}$$

$$d=\sqrt{5^2+4^2}$$

$$d=\sqrt{25+16}$$

$$d=\sqrt{41}$$

Or we can do this graphically using the Pythagorean Theorem. If we graph $(3,2)$ and $(-2,-2)$, and measure the horizontal and vertical distances between them, we get



Then the Pythagorean Theorem gives the distance between the points as the hypotenuse of the triangle, c.

$$4^2 + 5^2 = c^2$$

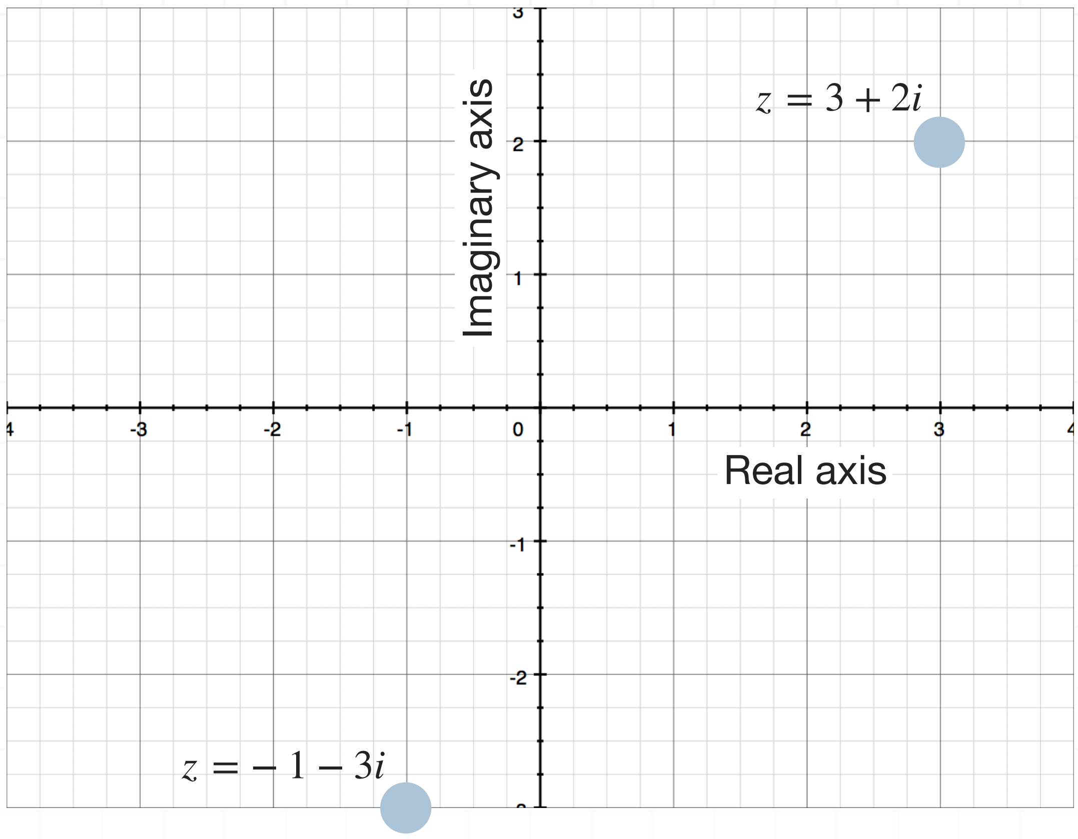
$$16 + 25 = c^2$$

$$41 = c^2$$

$$c = \sqrt{41}$$

Distance between complex numbers

Finding the distance between complex numbers follows exactly the same process. For instance, given the complex numbers $z = -1 - 3i$ and $z = 3 + 2i$,



we can find the distance between them by finding the difference between the real parts and the imaginary parts. The distance between the real parts is $3-(-1)=3+1=4$, and the distance between the imaginary parts is $2-(-3)=2+3=5$. Then by the Pythagorean Theorem, the distance between $z=-1-3i$ and $z=3+2i$ is

$$4^2+5^2=c^2$$

$$16+25=c^2$$

$$41=c^2$$

$$c=\sqrt{41}$$

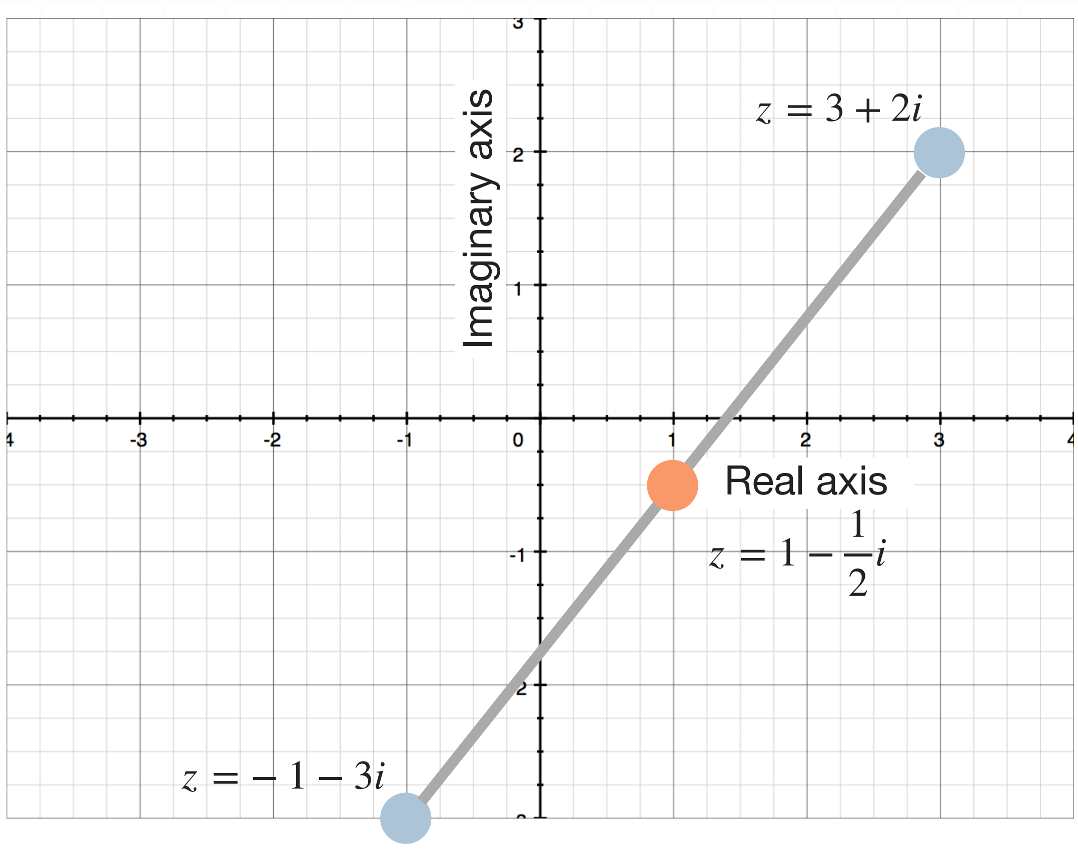
The midpoint between complex numbers

To find the midpoint between complex numbers, we just find the midpoint of the real parts, and separately the midpoint of the imaginary parts.

The distance between the real parts of $z=-1-3i$ and $z=3+2i$ is $3-(-1)=3+1=4$. Half of that distance is $4/2=2$, so we look for the value that's 2 units from -1 and 2 units from 3, so the midpoint between those real parts must be 1.

The distance between the imaginary parts of $z=-1-3i$ and $z=3+2i$ is $2-(-3)=2+3=5$. Half of that distance is $5/2=2.5$, so we look for the value that's 2.5 units from -3 and 2.5 units from 2, so the midpoint between those imaginary parts must be -0.5, or $-1/2$.

Therefore, the midpoint between $z=-1-3i$ and $z=3+2i$ is $z=1-(1/2)i$. If we graph all three of these in the complex plane, we get



Complex numbers in polar form

We've learned how to graph complex numbers in the complex plane, also called the **Argand plane**. But we'd also like to be able to graph complex numbers in polar coordinates. We can't do it without first learning how to find the absolute value of a complex number.

Absolute value

Remember that **absolute value** (also called the **magnitude**) really just means "distance from the origin." The distance from the origin of every complex number can be found using the Pythagorean Theorem with the real part and the imaginary part of the complex number.

For instance, given $z = -2 - i$, the distance of the real part from the origin is 2 units, and the distance of the imaginary part from the origin is 1 unit. Therefore, the distance of $z = -2 - i$ from the origin is

$$|z| = \sqrt{2^2 + 1^2}$$

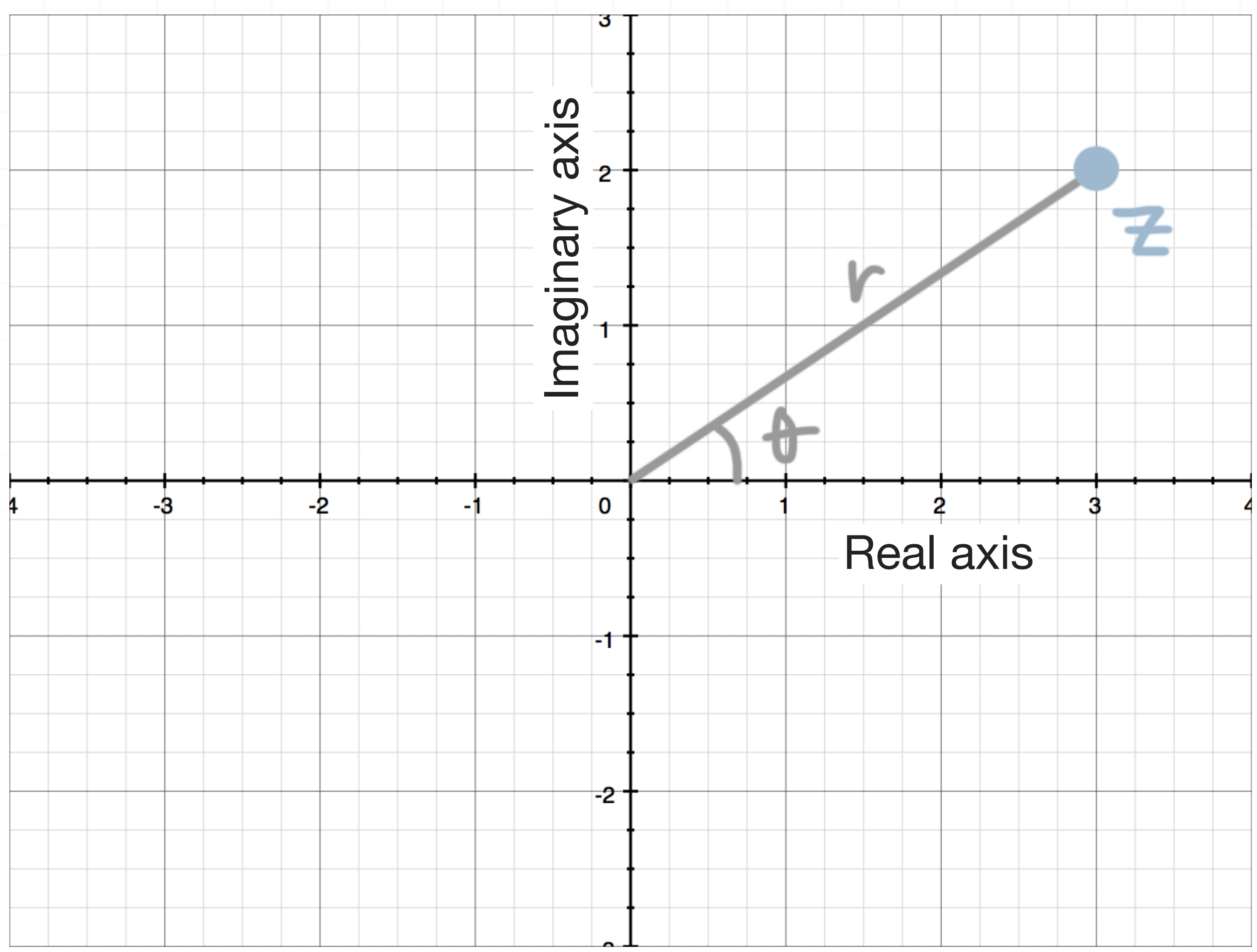
$$|z| = \sqrt{4 + 1}$$

$$|z| = \sqrt{5}$$

Polar form of complex numbers

We know that rectangular coordinates (x, y) always locate a position in coordinate space as the horizontal distance from the origin x, and the vertical distance from the origin y.

But polar coordinates locate points differently. They define points by distance from the origin r, and angle between the point and the positive side of the horizontal axis θ.



To convert a complex number in **rectangular form** $z = a + bi$ into polar form, we can use what we know geometrically about the complex number.

Remember that the real part of the complex number a is the horizontal distance from the origin to z, and the imaginary part of the complex number b is the vertical distance from the origin to z. So a, b, and r form a right triangle between the origin and z, where r is the hypotenuse of the triangle.

We already said that the magnitude, or the absolute value, of the complex number $|z|$ is the distance of z from the origin. So $r = |z|$ and we can say

$$r = |z| = \sqrt{a^2 + b^2}$$

To find a formula for θ, we need to remember the SOH-CAH-TOA rule from Trigonometry:

Sine = Opposite / Hypotenuse

Cosine = Adjacent / Hypotenuse

Tangent = Opposite / Adjacent

We know that the side opposite the angle θ is b, and that the side adjacent to the angle θ is a. So, given the opposite and adjacent side lengths as the imaginary and real parts of the complex number, respectively, we can use the tangent rule to write a formula for θ.

$$\tan\theta = \frac{b}{a}$$

$$\arctan(\tan\theta) = \arctan\left(\frac{b}{a}\right)$$

$$\theta = \arctan\left(\frac{b}{a}\right)$$

That's all assuming that we have values for a and b. But sometimes we'll have values for r and θ instead. In that case, we go back to SOH-CAH-TOA, and build these two equations:

$$\cos\theta = \frac{a}{r}, \text{ so } a = r\cos\theta$$

$$\sin\theta = \frac{b}{r}, \text{ so } b = r\sin\theta$$

Then, if we plug those values for a and b into the complex number equation, we get the **polar form** of a complex number.

$$z = a + bi$$

$$z = r\cos\theta + (r\sin\theta)i$$

$$z = r(\cos\theta + i\sin\theta)$$

This is a topic for a different time, but the value inside the parentheses $(\cos\theta + i\sin\theta)$ is equal to $e^{i\theta}$, where e is Euler's number $e \approx 2.718$ and i is the imaginary number. Which means the complex number can also be written in **exponential form** as

$$z = re^{i\theta}$$

Let's do an example where we convert a complex number into polar form.

Example

Write the complex number in polar form.

$$z = \frac{\sqrt{2}}{2} - \frac{\sqrt{2}}{2}i$$

The real part of this complex number is $a = \sqrt{2}/2$ and its imaginary part is $b = -\sqrt{2}/2$, so the value of r will be

$$r = \sqrt{a^2 + b^2}$$

$$r = \sqrt{\left(\frac{\sqrt{2}}{2}\right)^2 + \left(-\frac{\sqrt{2}}{2}\right)^2}$$

$$r = \sqrt{\frac{1}{2} + \frac{1}{2}}$$

$$r = \sqrt{1}$$

$$r = 1$$

The value of θ is

$$\tan\theta = \frac{-\frac{\sqrt{2}}{2}}{\frac{\sqrt{2}}{2}} = -1$$

$$\arctan(\tan\theta) = \arctan(-1)$$

$$\theta = \arctan(-1)$$

$$\theta = -\frac{\pi}{4}$$

Then the complex number written in polar form is

$$z = r(\cos\theta + i\sin\theta)$$

$$z = 1\left[\cos\left(-\frac{\pi}{4}\right) + i\sin\left(-\frac{\pi}{4}\right)\right]$$

Let's do another example where we convert the complex number into polar form.

Example

Write the complex number in polar form.

$$z = 6i$$

We can write the complex number $6i$ as $0 + 6i$, so its real part is $a = 0$ and its imaginary part is $b = 6$. The distance of $6i$ from the origin is

$$r = \sqrt{a^2 + b^2} = \sqrt{0^2 + 6^2} = \sqrt{0 + 36} = \sqrt{36} = 6$$

Since the imaginary part of $6i$ is 6, that means the complex number is located on the positive imaginary axis, so $\theta = \pi/2$. In polar form, the complex number is

$$r(\cos\theta + i\sin\theta)$$

$$6\left[\cos\left(\frac{\pi}{2}\right) + i\sin\left(\frac{\pi}{2}\right)\right]$$

The takeaway here is that $a + bi$ (the **rectangular form**) and $r\cos\theta + r\sin\theta i$ (the **polar form**) are two different ways to express the same thing. That's why we call both of them z.

$$z = a + bi$$

$$z = r\cos\theta + (r\sin\theta)i$$

Comparing these equations to each other gives us these two equations:

$$a = r\cos\theta$$

$$b = r\sin\theta$$

To convert back and forth between rectangular and polar forms, we use the conversion equations

$$r = |z| = \sqrt{a^2 + b^2}$$

$$\theta = \arctan\left(\frac{b}{a}\right)$$

Multiplying and dividing polar forms

Now that we know about complex, polar, and exponential forms, let's look at how to multiply and divide polar forms.

Multiplying polar forms

To multiply two complex numbers $z_1 = r_1(\cos\theta_1 + i\sin\theta_1)$ and $z_2 = r_2(\cos\theta_2 + i\sin\theta_2)$ that are given in polar form, we first pull out r_1 and r_2.

$$z_1z_2 = [r_1(\cos\theta_1 + i\sin\theta_1)][r_2(\cos\theta_2 + i\sin\theta_2)]$$

$$z_1z_2 = r_1r_2(\cos\theta_1 + i\sin\theta_1)(\cos\theta_2 + i\sin\theta_2)$$

Multiply the binomials.

$$z_1z_2 = r_1r_2(\cos\theta_1\cos\theta_2 + i\sin\theta_2\cos\theta_1 + i\sin\theta_1\cos\theta_2 + i^2\sin\theta_1\sin\theta_2)$$

$$z_1z_2 = r_1r_2(\cos\theta_1\cos\theta_2 + i\sin\theta_2\cos\theta_1 + i\sin\theta_1\cos\theta_2 + (-1)\sin\theta_1\sin\theta_2)$$

$$z_1z_2 = r_1r_2(\cos\theta_1\cos\theta_2 + i\sin\theta_2\cos\theta_1 + i\sin\theta_1\cos\theta_2 - \sin\theta_1\sin\theta_2)$$

Group together the real and imaginary parts. Then use the sum identities for sine and cosine to simplify.

$$z_1z_2 = r_1r_2[(\cos\theta_1\cos\theta_2 - \sin\theta_1\sin\theta_2) + (\sin\theta_2\cos\theta_1 + \sin\theta_1\cos\theta_2)i]$$

$$z_1z_2 = r_1r_2[\cos(\theta_1 + \theta_2) + i\sin(\theta_1 + \theta_2)]$$

When we look at this product, we can see that

1. its distance from the origin is the product of the original distances $r = r_1 r_2$, and

2. its angle is the sum of the original angles $\theta = \theta_1 + \theta_2$.

Let's work through an example where we multiply complex numbers given in polar form.

Example

Find the product $z_1 z_2$ of the complex numbers in polar form.

$$z_1 = 3\left(\cos\left(\frac{3\pi}{8}\right) + i\sin\left(\frac{3\pi}{8}\right)\right)$$

$$z_2 = \sqrt{5}\left(\cos\left(\frac{7\pi}{6}\right) + i\sin\left(\frac{7\pi}{6}\right)\right)$$

Plugging these complex numbers into the formula for the product of complex numbers gives

$$z_1 z_2 = r_1 r_2[\cos(\theta_1 + \theta_2) + i\sin(\theta_1 + \theta_2)]$$

$$z_1 z_2 = (3 \cdot \sqrt{5})\left[\cos\left(\frac{3\pi}{8} + \frac{7\pi}{6}\right) + i\sin\left(\frac{3\pi}{8} + \frac{7\pi}{6}\right)\right]$$

$$z_1 z_2 = 3\sqrt{5}\left[\cos\left(\frac{9\pi}{24} + \frac{28\pi}{24}\right) + i\sin\left(\frac{9\pi}{24} + \frac{28\pi}{24}\right)\right]$$

$$z_1 z_2 = 3\sqrt{5}\left(\cos\left(\frac{37\pi}{24}\right) + i\sin\left(\frac{37\pi}{24}\right)\right)$$

Dividing polar forms

To divide two complex numbers $z_1 = r_1(\cos\theta_1 + i\sin\theta_1)$ and $z_2 = r_2(\cos\theta_2 + i\sin\theta_2)$ that are given in polar form, we get

$$\frac{z_1}{z_2} = \frac{r_1(\cos\theta_1 + i\sin\theta_1)}{r_2(\cos\theta_2 + i\sin\theta_2)}$$

We could simplify this division using the complex conjugate of the denominator. But it's actually much easier to convert the complex numbers in the division to exponential form.

$$\frac{z_1}{z_2} = \frac{r_1 e^{i\theta_1}}{r_2 e^{i\theta_2}}$$

Then exponent rules let us simplify this to

$$\frac{z_1}{z_2} = \frac{r_1}{r_2} e^{i\theta_1 - i\theta_2}$$

$$\frac{z_1}{z_2} = \frac{r_1}{r_2} e^{i(\theta_1 - \theta_2)}$$

If, instead of converting to exponential form, we had gone forward with multiplying by the complex conjugate, we would have found

$$\frac{z_1}{z_2} = \frac{r_1}{r_2}[\cos(\theta_1 - \theta_2) + i\sin(\theta_1 - \theta_2)]$$

We can use either of these formulas (the polar form or the exponential form) to find the quotient of two complex numbers.

Example

Find the quotient z_1/z_2 of the complex numbers in polar form.

$$z_1 = 13\left(\cos\left(\frac{8\pi}{5}\right) + i\sin\left(\frac{8\pi}{5}\right)\right)$$

$$z_2 = 8\left(\cos\left(\frac{2\pi}{3}\right) + i\sin\left(\frac{2\pi}{3}\right)\right)$$

Plugging these complex numbers into the formula for the quotient of complex numbers gives

$$\frac{z_1}{z_2} = \frac{r_1}{r_2}[\cos(\theta_1 - \theta_2) + i\sin(\theta_1 - \theta_2)]$$

$$\frac{z_1}{z_2} = \frac{13}{8}\left[\cos\left(\frac{8\pi}{5} - \frac{2\pi}{3}\right) + i\sin\left(\frac{8\pi}{5} - \frac{2\pi}{3}\right)\right]$$

$$\frac{z_1}{z_2} = \frac{13}{8}\left[\cos\left(\frac{24\pi}{15} - \frac{10\pi}{15}\right) + i\sin\left(\frac{24\pi}{15} - \frac{10\pi}{15}\right)\right]$$

$$\frac{z_1}{z_2} = \frac{13}{8}\left(\cos\left(\frac{14\pi}{15}\right) + i\sin\left(\frac{14\pi}{15}\right)\right)$$

Powers of complex numbers and De Moivre's Theorem

In this lesson, we want to build on our understanding of multiplication of complex numbers in polar form, by looking at how to find powers of complex numbers, like z^2, z^3, z^4, etc.

Exponential form for large powers

We don't know yet how to raise the complex number like

$$z = 3\left(\cos\left(\frac{3\pi}{8}\right) + i\sin\left(\frac{3\pi}{8}\right)\right)$$

to a power. For instance, if we want to find z^{12}, it seems like it would be pretty tedious to do it this way:

$$z = \left[3\left(\cos\left(\frac{3\pi}{8}\right) + i\sin\left(\frac{3\pi}{8}\right)\right)\right]^{12}$$

$$z = 3^{12}\left(\cos\left(\frac{3\pi}{8}\right) + i\sin\left(\frac{3\pi}{8}\right)\right)^{12}$$

We'll actually learn how to do this in the next part of this lesson, but for now we want to realize that we can convert the complex number from polar form to exponential form. In exponential form, we get

$$z = re^{i\theta}$$

$$z^{12} = (re^{i\theta})^{12}$$

$$z^{12} = r^{12}e^{12i\theta}$$

$$z^{12} = 3^{12}e^{12i\cdot\frac{3\pi}{8}}$$

$$z^{12} = 531{,}441e^{\frac{9\pi}{2}i}$$

De Moivre's Theorem

We don't always have to convert to exponential form to find the power of a complex number. We can keep complex numbers in polar form. Let's talk about how.

We know that when we multiply two complex numbers z_1 and z_2, their product is

$$z_1z_2 = r_1r_2[\cos(\theta_1 + \theta_2) + i\sin(\theta_1 + \theta_2)]$$

In the previous lesson, we always multiplied different complex numbers. But what happens when we multiply two equivalent complex numbers? Let's say we want to find the product of z_1, multiplied by z_1. We know from the product formula that the result is

$$z_1z_2 = r_1r_2[\cos(\theta_1 + \theta_2) + i\sin(\theta_1 + \theta_2)]$$

$$z_1z_1 = r_1r_1[\cos(\theta_1 + \theta_1) + i\sin(\theta_1 + \theta_1)]$$

$$(z_1)^2 = (r_1)^2[\cos(2\theta_1) + i\sin(2\theta_1)]$$

Let's say we want to take this result and multiply by z_1 again. We'd get

$$(z_1)^2z_1 = (r_1)^2r_1[\cos(2\theta_1 + \theta_1) + i\sin(2\theta_1 + \theta_1)]$$

$$(z_1)^3 = (r_1)^3[\cos(3\theta_1) + i\sin(3\theta_1)]$$

If we multiply again by z_1, we get

$$(z_1)^3 z_1 = (r_1)^3 r_1[\cos(3\theta_1 + \theta_1) + i\sin(3\theta_1 + \theta_1)]$$

$$(z_1)^4 = (r_1)^4[\cos(4\theta_1) + i\sin(4\theta_1)]$$

When we put these results together,

$$z^2 = r^2[\cos(2\theta) + i\sin(2\theta)]$$

$$z^3 = r^3[\cos(3\theta) + i\sin(3\theta)]$$

$$z^4 = r^4[\cos(4\theta) + i\sin(4\theta)]$$

we see a pattern emerging. The pattern we're getting is

$$z^n = r^n[\cos(n\theta) + i\sin(n\theta)]$$

This is **De Moivre's Theorem** (sometimes called De Moivre's formula), and it comes directly from what we already knew about multiplying complex numbers. Looking at the formula, we can see that it tells us that, if we want to raise a complex number to the 5th power, we just raise r to the 5th power, and multiply θ by 5.

Example

Use De Moivre's Theorem to find z^3, the third power of z.

$$z = 7\left(\cos\left(\frac{\pi}{6}\right) + i\sin\left(\frac{\pi}{6}\right)\right)$$

To find z^3, we need to use $n = 3$ in De Moivre's Theorem. With $n = 3$, and $r = 7$ and $\theta = \pi/6$ from z, we plug into De Moivre's Theorem to get

$$z^n = r^n[\cos(n\theta) + i\sin(n\theta)]$$

$$z^3 = r^3[\cos(3\theta) + i\sin(3\theta)]$$

$$z^3 = 7^3\left(\cos\left(3 \cdot \frac{\pi}{6}\right) + i\sin\left(3 \cdot \frac{\pi}{6}\right)\right)$$

$$z^3 = 343\left(\cos\left(\frac{\pi}{2}\right) + i\sin\left(\frac{\pi}{2}\right)\right)$$

We could leave the complex number in its polar form, or we could rewrite it in rectangular form.

$$z^3 = 343(0 + i(1))$$

$$z^3 = 343i$$

Let's do another example with different n, r, and θ values.

Example

Find z^8.

$$z = \frac{1}{2}\left(\cos\left(\frac{\pi}{7}\right) + i\sin\left(\frac{\pi}{7}\right)\right)$$

Plugging $r = 1/2$, $\theta = \pi/7$, and $n = 8$ into De Moivre's Theorem, we get

$$z^n = r^n[\cos(n\theta) + i\sin(n\theta)]$$

$$z^8 = r^8[\cos(8\theta) + i\sin(8\theta)]$$

$$z^8 = \left(\frac{1}{2}\right)^8\left(\cos\left(8 \cdot \frac{\pi}{7}\right) + i\sin\left(8 \cdot \frac{\pi}{7}\right)\right)$$

$$z^8 = \frac{1}{256}\left(\cos\left(\frac{8\pi}{7}\right) + i\sin\left(\frac{8\pi}{7}\right)\right)$$

Power of a rectangular complex number

We know from the previous examples how to find the power of a complex number when the complex number is already in polar form. If we want to find the power of a complex number that's given in rectangular form, we need to first convert the rectangular complex number into a polar complex number.

For instance, given the rectangular complex number $z = 1 - \sqrt{3}i$, we'll first convert it to polar form by finding the modulus $|z|$ and the angle θ. The distance from the origin is

$$|z| = r = \sqrt{a^2 + b^2}$$

$$|z| = r = \sqrt{1^2 + (-\sqrt{3})^2}$$

$$|z| = r = \sqrt{1+3}$$

$$|z| = r = \sqrt{4}$$

$$|z| = r = 2$$

and the angle is

$$\theta = \arctan\left(\frac{b}{a}\right) = \arctan\left(\frac{-\sqrt{3}}{1}\right) = \arctan(-\sqrt{3}) = -\frac{\pi}{3}$$

Then $z = 1 - \sqrt{3}i$ in polar form is

$$z = r(\cos\theta + i\sin\theta)$$

$$z = 2\left(\cos\left(-\frac{\pi}{3}\right) + i\sin\left(-\frac{\pi}{3}\right)\right)$$

And now that the complex number is in polar form, we know that we can find any power of it using De Moivre's Theorem. If we want to find z^4, we'll plug into De Moivre's formula to get

$$z^n = r^n[\cos(n\theta) + i\sin(n\theta)]$$

$$z^4 = r^4[\cos(4\theta) + i\sin(4\theta)]$$

With $r = 2$ and $\theta = -\pi/3$, this becomes

$$z^4 = 2^4\left(\cos\left(4\left(-\frac{\pi}{3}\right)\right) + i\sin\left(4\left(-\frac{\pi}{3}\right)\right)\right)$$

$$z^4 = 16\left(\cos\left(-\frac{4\pi}{3}\right) + i\sin\left(-\frac{4\pi}{3}\right)\right)$$

We can leave this in polar form, or rewrite it in rectangular form as

$$z^4 = 16\left(-\frac{1}{2} + \frac{\sqrt{3}}{2}i\right)$$

$$z^4 = 8(-1 + \sqrt{3}i)$$

$$z^4 = -8 + 8\sqrt{3}i$$

Complex number equations

Let's talk about another application of powers of complex numbers. We can use what we've learned already about complex numbers to solve equations like $z^4 = 16$.

In an equation like this one, z represents a complex number, which means we're looking for the complex numbers that satisfy the equation. We'll do this using a system of equations.

First, we want to focus on the left side of $z^4 = 16$. Using De Moivre's Theorem, we know we can rewrite z^4 as

$$z^n = r^n[\cos(n\theta) + i\sin(n\theta)]$$

$$z^4 = r^4[\cos(4\theta) + i\sin(4\theta)]$$

Now we'll look just at the right side of $z^4 = 16$. We know we can rewrite 16 as the complex number $16 + 0i$ in rectangular form. If we find the modulus and angle of this complex number, we get

$$r = \sqrt{16^2 + 0^2}$$

$$r = \sqrt{16^2}$$

$$r = 16$$

and

$$\theta = \arctan\left(\frac{b}{a}\right) = \arctan\left(\frac{0}{16}\right) = \arctan 0 = 0$$

Keep in mind though that this arctan equation is true at an angle of 0, but it's also true at angles coterminal with 0, including 2π, 4π, 6π, 8π, etc. So if we put this in polar form, we get

$$z = 16[\cos(0,\ 2\pi,\ 4\pi,\ 8\pi,\ \dots) + i\sin(0,\ 2\pi,\ 4\pi,\ 8\pi,\ \dots)]$$

$$z = 16[\cos(2\pi k) + i\sin(2\pi k)]$$

We could also write this in degrees instead of radians as

$$z = 16[\cos(360^\circ k) + i\sin(360^\circ k)]$$

Starting again with $z^4 = 16$, we can now make substitutions.

$$z^4 = 16$$

$$r^4[\cos(4\theta) + i\sin(4\theta)] = 16$$

$$r^4[\cos(4\theta) + i\sin(4\theta)] = 16[\cos(360^\circ k) + i\sin(360^\circ k)]$$

Because we have the same form on both sides of the equation, we can equate corresponding values to produce a system of equations.

$$r^4 = 16$$

$$4\theta = 360^\circ k$$

From these equations, we get

$r^4 = 16$, so $r = 2$

$4\theta = 360^\circ k$, so $\theta = 90^\circ k$

If we plug $k = 0,\ 1,\ 2,\ 3,\ \dots$ into $\theta = 90^\circ k$, we get

For $k = 0$, $\theta = 90^\circ(0) = 0^\circ$

For $k = 1$, $\theta = 90^\circ(1) = 90^\circ$

For $k = 2$, $\theta = 90^\circ(2) = 180^\circ$

For $k = 3$, $\theta = 90^\circ(3) = 270^\circ$

...

We could keep going for $k = 4, 5, 6, 7, \dots$, but $k = 4$ gives 360°, which is coterminal with the 0° value we already found for $k = 0$, so we realize that we'll just be starting to repeat the same solutions. So the only solutions we need to consider are $\theta = 0^\circ, 90^\circ, 180^\circ$, and 270°.

Plugging these four angles and $r = 2$ into the formula for the polar form of a complex number, we'll get the solutions to $z^4 = 16$.

$$z_1 = 2[\cos(0^\circ) + i\sin(0^\circ)] = 2[1 + i(0)] = 2$$

$$z_2 = 2[\cos(90^\circ) + i\sin(90^\circ)] = 2[0 + i(1)] = 2i$$

$$z_3 = 2[\cos(180^\circ) + i\sin(180^\circ)] = 2[-1 + i(0)] = -2$$

$$z_4 = 2[\cos(270^\circ) + i\sin(270^\circ)] = 2[0 + i(-1)] = -2i$$

We can double-check that these are all roots of 16 by substituting into $z^4 = 16$ for z.

For $z_1 = 2$, we get

$$2^4 = 16$$

$$16 = 16$$

For $z_2 = 2i$, we get

$$(2i)^4 = 16$$

$$16i^4 = 16$$

$$16i^2i^2 = 16$$

$$16(-1)(-1) = 16$$

$$16 = 16$$

For $z_3 = -2$, we get

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

$$16 = 16$$

For $z_4 = -2i$, we get

$$(-2i)^4 = 16$$

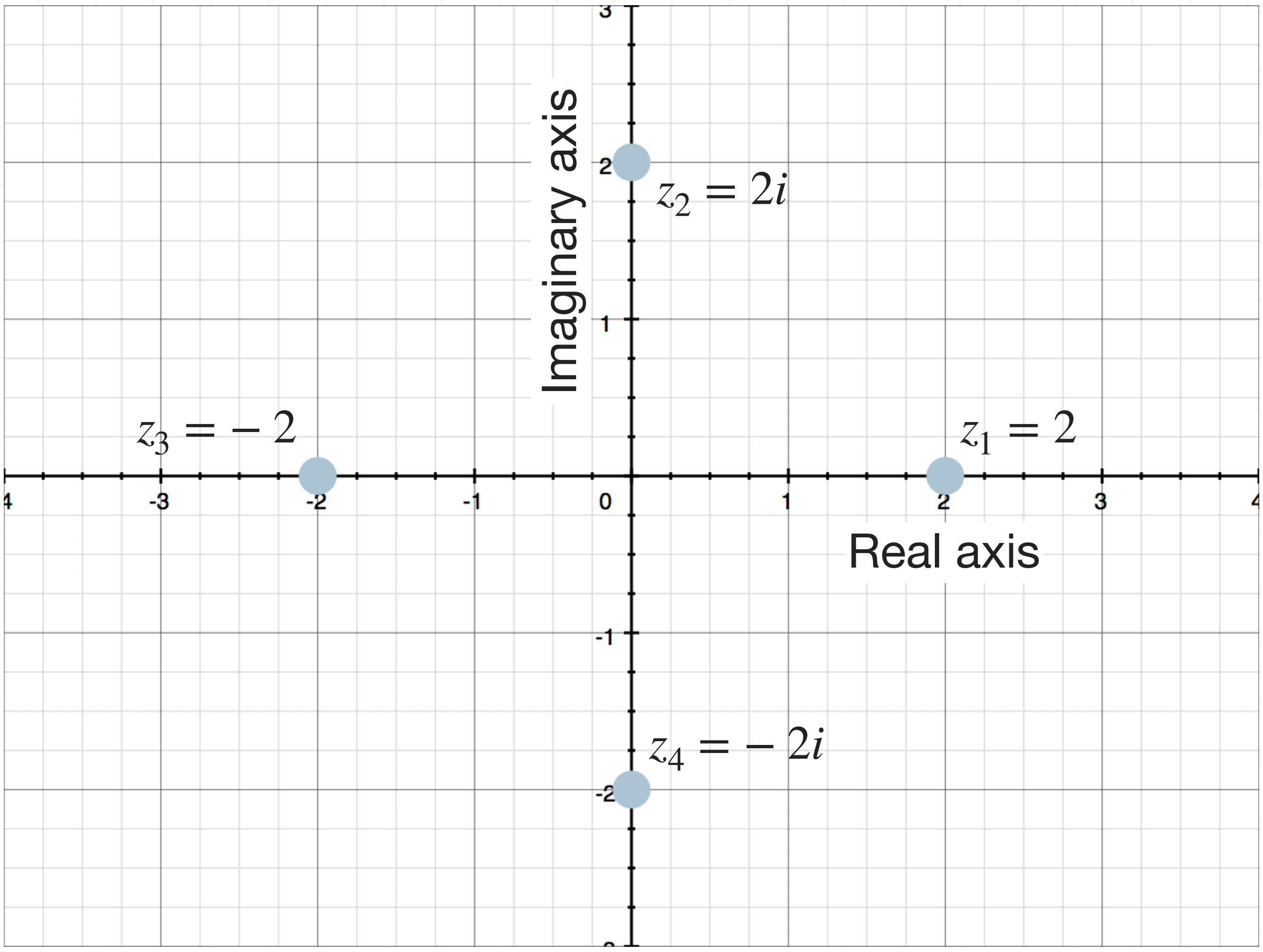
$$16i^4 = 16$$

$$16i^2i^2 = 16$$

$$16(-1)(-1) = 16$$

$$16 = 16$$

And if we graph these complex numbers, we see the four solutions to $z^4 = 16$ in the complex plane.



Imaginary axis
Real axis
$z_1 = 2$
$z_2 = 2i$
$z_3 = -2$
$z_4 = -2i$

Roots of complex numbers

We know how to raise a complex number to a power, but in this lesson we want to talk about the opposite operation: how to find the roots of a complex number.

To do this, we always want to start with a complex number in polar form, $z = r(\cos\theta + i\sin\theta)$. If the complex number is given in rectangular form $z = a + bi$, we'll convert it to polar form to get started.

Once the complex number is in polar form, then its nth roots are given in radians by

$$\sqrt[n]{z} = \sqrt[n]{r}\left[\cos\left(\frac{\theta + 2\pi k}{n}\right) + i\sin\left(\frac{\theta + 2\pi k}{n}\right)\right]$$

or in degrees by

$$\sqrt[n]{z} = \sqrt[n]{r}\left[\cos\left(\frac{\theta + 360^\circ k}{n}\right) + i\sin\left(\frac{\theta + 360^\circ k}{n}\right)\right]$$

What do we mean by the nth roots? Well, if we're finding the 5th roots of a complex number, it means we're finding the set of values that, when raised to the 5th power, are equal to the complex number.

Keep in mind also that there are always n nth roots. So if we're looking for 5th roots, we'll find five of them. If we're looking for 3rd roots, we'll find three of them.

And the roots are always defined by the values $k = 0, 1, 2, 3, \ldots n - 1$. So the 5th roots will be given by $k = 0, 1, 2, 3, 4$.

Example

Find the 4th roots of the complex number.

$$z = 81(\cos 180^\circ + i \sin 180^\circ)$$

Because we're looking for 4th roots, we know there will be four of them, and they'll be given by $k = 0, 1, 2, 3$.

And since this complex number is given in degrees, we'll plug $n = 4$ into the formula for nth roots in degrees.

$$\sqrt[n]{z} = \sqrt[n]{r}\left[\cos\left(\frac{\theta + 360^\circ k}{n}\right) + i \sin\left(\frac{\theta + 360^\circ k}{n}\right)\right]$$

$$\sqrt[4]{z} = \sqrt[4]{r}\left[\cos\left(\frac{\theta + 360^\circ k}{4}\right) + i \sin\left(\frac{\theta + 360^\circ k}{4}\right)\right]$$

For the given complex number, $r = 81$ and $\theta = 180^\circ$, so

$$\sqrt[4]{z} = \sqrt[4]{81}\left[\cos\left(\frac{180^\circ + 360^\circ k}{4}\right) + i \sin\left(\frac{180^\circ + 360^\circ k}{4}\right)\right]$$

Now we'll find values for $k = 0, 1, 2, 3$.

For $k = 0$:

$$\sqrt[4]{z}_{k=0} = \sqrt[4]{81}\left[\cos\left(\frac{180^\circ + 360^\circ(0)}{4}\right) + i\sin\left(\frac{180^\circ + 360^\circ(0)}{4}\right)\right]$$

$$\sqrt[4]{z}_{k=0} = \sqrt[4]{81}\left(\cos\frac{180^\circ}{4} + i\sin\frac{180^\circ}{4}\right)$$

$$\sqrt[4]{z}_{k=0} = \sqrt[4]{81}(\cos 45^\circ + i\sin 45^\circ)$$

$$\sqrt[4]{z}_{k=0} = 3\left(\frac{\sqrt{2}}{2} + \frac{\sqrt{2}}{2}i\right)$$

$$\sqrt[4]{z}_{k=0} = \frac{3\sqrt{2}}{2} + \frac{3\sqrt{2}}{2}i$$

For $k = 1$:

$$\sqrt[4]{z}_{k=1} = \sqrt[4]{81}\left[\cos\left(\frac{180^\circ + 360^\circ(1)}{4}\right) + i\sin\left(\frac{180^\circ + 360^\circ(1)}{4}\right)\right]$$

$$\sqrt[4]{z}_{k=1} = \sqrt[4]{81}\left(\cos\frac{540^\circ}{4} + i\sin\frac{540^\circ}{4}\right)$$

$$\sqrt[4]{z}_{k=1} = \sqrt[4]{81}(\cos 135^\circ + i\sin 135^\circ)$$

$$\sqrt[4]{z}_{k=1} = 3\left(-\frac{\sqrt{2}}{2} + \frac{\sqrt{2}}{2}i\right)$$

$$\sqrt[4]{z}_{k=1} = -\frac{3\sqrt{2}}{2} + \frac{3\sqrt{2}}{2}i$$

For $k = 2$:

$$\sqrt[4]{z}_{k=2} = \sqrt[4]{81}\left[\cos\left(\frac{180^\circ + 360^\circ(2)}{4}\right) + i\sin\left(\frac{180^\circ + 360^\circ(2)}{4}\right)\right]$$

$$\sqrt[4]{z}_{k=2} = \sqrt[4]{81}\left(\cos\frac{900^\circ}{4} + i\sin\frac{900^\circ}{4}\right)$$

$$\sqrt[4]{z}_{k=2} = \sqrt[4]{81}(\cos 225^\circ + i\sin 225^\circ)$$

$$\sqrt[4]{z}_{k=2} = 3\left(-\frac{\sqrt{2}}{2} - \frac{\sqrt{2}}{2}i\right)$$

$$\sqrt[4]{z}_{k=2} = -\frac{3\sqrt{2}}{2} - \frac{3\sqrt{2}}{2}i$$

For $k = 3$:

$$\sqrt[4]{z}_{k=3} = \sqrt[4]{81}\left[\cos\left(\frac{180^\circ + 360^\circ(3)}{4}\right) + i\sin\left(\frac{180^\circ + 360^\circ(3)}{4}\right)\right]$$

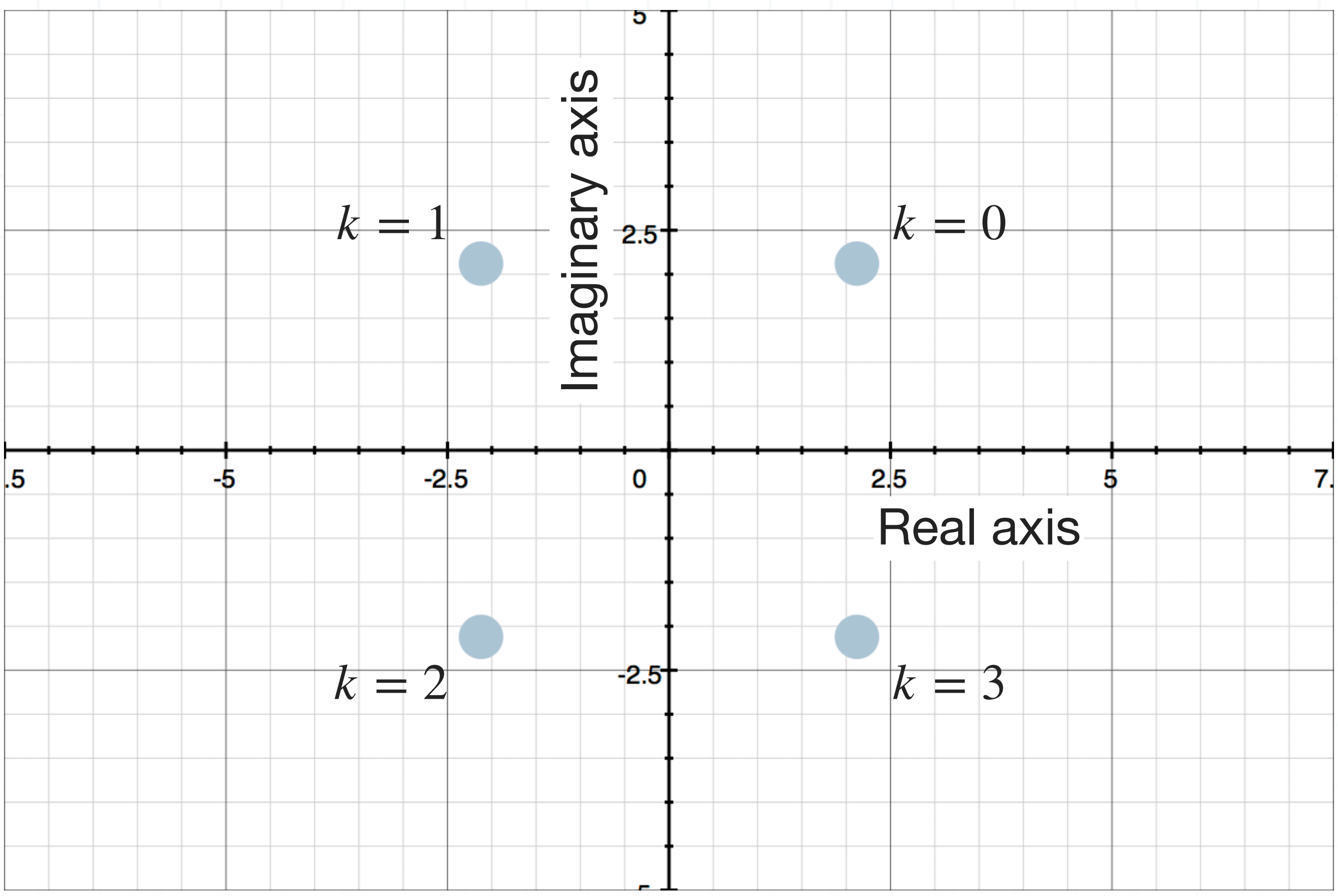
$$\sqrt[4]{z}_{k=3} = \sqrt[4]{81}\left(\cos\frac{1{,}260^\circ}{4} + i\sin\frac{1{,}260^\circ}{4}\right)$$

$$\sqrt[4]{z}_{k=3} = \sqrt[4]{81}(\cos 315^\circ + i\sin 315^\circ)$$

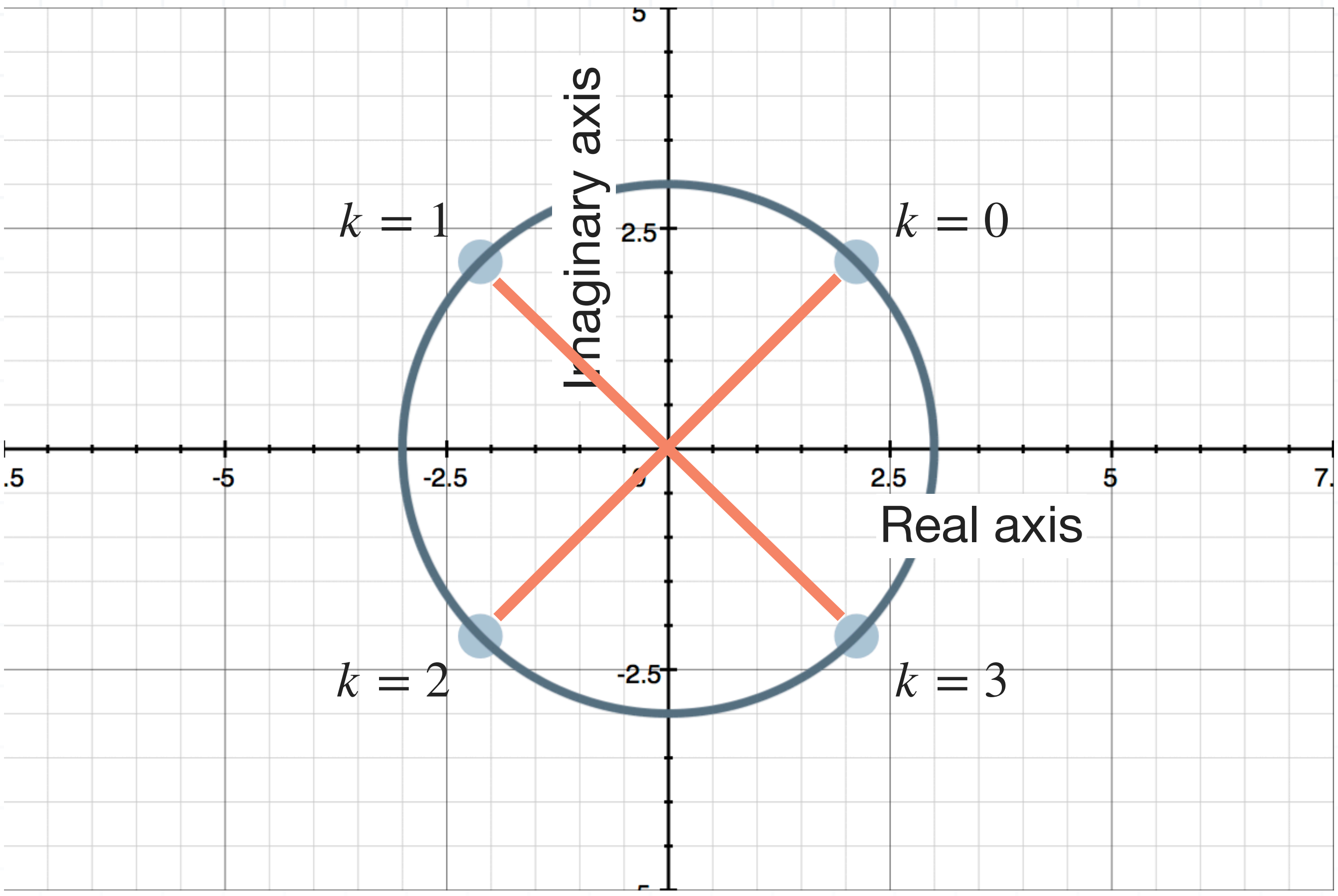
$$\sqrt[4]{z}_{k=3} = 3\left(\frac{\sqrt{2}}{2} - \frac{\sqrt{2}}{2}i\right)$$

$$\sqrt[4]{z}_{k=3} = \frac{3\sqrt{2}}{2} - \frac{3\sqrt{2}}{2}i$$

We found four 4th roots. If we graph them in the complex plane, we can see what they look like.



Notice how these fourth roots of the complex number (in which $r = 3$), divide the circle with radius 3 into four equal parts.



5
2.5
-2.5
-5
.5
-5
-2.5
0
2.5
5
7.
Imaginary axis
Real axis
$k = 1$
$k = 0$
$k = 2$
$k = 3$

Matrix dimensions and entries

A **matrix** is a rectangular array of values, where each value is an **entry** in both a row and a column.

Matrix dimensions

A matrix is often described by its number of rows and columns. For instance, a 3×4 matrix is a matrix with 3 rows and 4 columns. In the description "3×4," the number of rows always comes first, and the number of columns always comes second, so remember:

"rows $\times$ columns"

A matrix can be as small as 1×1, with one row and one column, in which case it looks like

$$[a]$$

Or it can have infinitely many rows and/or columns. It can have the same number of rows and columns, more rows than columns, or more columns than rows.

A 2×2 matrix: $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$

A 2×3 matrix: $\begin{bmatrix} a & b & c \\ d & e & f \end{bmatrix}$

A 3×2 matrix: $\begin{bmatrix} a & b \\ c & d \\ e & f \end{bmatrix}$

Matrices are just a way of arranging data, especially large amounts of data. And in this section we'll learn about different ways to work with matrices, like how to add them or multiply them.

And these kinds of matrix operations are useful in all kinds of fields, like statistics, economics, data analysis, and computer programming.

Example

Give the dimensions of each matrix.

$$A = \begin{bmatrix} 7 & 3 & 4 \\ 1 & 6 & 1 \\ 2 & 2 & 3 \end{bmatrix} \qquad B = \begin{bmatrix} 1 & 9 & 0 & 0 & 2 \end{bmatrix} \qquad C = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

We always give the dimensions of a matrix as rows $\times$ columns. Matrix A has 3 rows and 3 columns, so A is a 3×3 matrix. Matrix B has 1 row and 5 columns, so B is a 1×5 matrix. Matrix C has 3 rows and 1 column, so C is a 3×1 matrix.

Matrix entries

We call out a particular entry in a matrix using the name of the matrix and the row and column where the entry is sitting. So if the matrix is called upper-case K (upper-case letters are often used to name matrices), and

$$K = \begin{bmatrix} a & b & c \\ d & e & f \end{bmatrix}$$

then we'll use lower-case k (or possibly still upper-case K) to name a particular entry. If we want the entry in the first row, third column, we write that as $k_{1,3} = c$, since c is the entry in the first row, third column, of matrix K.

Example

Identify $A_{2,3}$, $B_{1,4}$, and $C_{3,1}$.

$$A = \begin{bmatrix} 7 & 3 & 4 \\ 1 & 6 & 1 \\ 2 & 2 & 3 \end{bmatrix} \qquad B = \begin{bmatrix} 1 & 9 & 0 & 0 & 2 \end{bmatrix} \qquad C = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

The value of $A_{2,3}$ is the entry in the second row, third column of matrix A, which is 1, so $A_{2,3} = 1$. The value of $B_{1,4}$ is the entry in the first row, fourth column of matrix B, which is 0, so $B_{1,4} = 0$. The value of $C_{3,1}$ is the entry in the third row, first column of matrix C, so $C_{3,1} = 1$.

Representing systems with matrices

In Algebra, we learned how to solve systems of linear equations using three methods:

1. Substitution

2. Elimination

3. Graphing

But we can also use matrices to solve systems of equations.

Representing systems

When we use a matrix to represent a system of linear equations, we call it an **augmented matrix**. Each row in the augmented matrix represents one equation in the system, and each column represents a different variable, or the constants. For instance, given the linear system

$$3x + 2y = 7$$

$$x - 6y = 0$$

the augmented matrix for the system would be

$$\begin{bmatrix} 3 & 2 & 7 \\ 1 & -6 & 0 \end{bmatrix}$$

What we want to notice is that each row corresponds to one of the equations. In the first row, the values 3, 2, and 7 come from the first equation, $3x + 2y = 7$. And in the second row, the values 1, -6, and 0 come from the second equation, $x - 6y = 0$.

Looking at the columns, we can see that the values from the first column are the coefficients on x, the values from the second column are the coefficients on y, and the values from the third column are the constants that come from the right side of each equation.

We can use augmented matrices to represent systems of any size. If we added two more equations to the system, we'd simply add two more rows to the matrix. Or if we added another variable to the system, like z, we'd simply add one more column to the matrix.

Let's do an example with a few more variables.

Example

Represent the system with an augmented matrix called M.

$$-2x + y - t = 7$$

$$x - y + z + 4t = 0$$

We always want to look at all the variables that are included in the system, not just the first equation, since the first equation may not include all the variables.

This particular system includes x, y, z, and t. Which means the augmented matrix will have four columns, one for each variable, plus a column for the constants, so five columns in total. Because there are two equations in the system, the matrix will have two rows. We could set up the matrix as

$$M = \begin{bmatrix} x_1 & y_1 & z_1 & t_1 & C_1 \\ x_2 & y_2 & z_2 & t_2 & C_2 \end{bmatrix}$$

Because there's no z-term in the first equation, the value of z_1 will be 0. If we fill in the matrix with that value and all the other coefficients and constants, we get

$$M = \begin{bmatrix} -2 & 1 & 0 & -1 & 7 \\ 1 & -1 & 1 & 4 & 0 \end{bmatrix}$$

Checking the order

As we're building out an augmented matrix, we want to be sure that we have all the variables in the same order, and all our constants grouped together on the same side of the equation. That way, with everything lined up, it'll be easy to make sure that each entry in a column represents the same variable or constant, and that each row in the matrix captures the entire equation.

Let's do an example where the terms aren't already in order.

Example

Express the system of linear equations as a matrix called B.

$$2x + 3y - z = 11$$

$$7y = 6 - x - 4z$$

$$-8z + 3 = y$$

Before we do anything, we want to put each equation in order, with x, then y, then z on the left side, and the constant on the right side.

$$2x + 3y - z = 11$$

$$x + 7y + 4z = 6$$

$$-y - 8z = -3$$

We could also recognize that there's no x-term in the third equation, and add in a 0 "filler" term in its place.

$$2x + 3y - z = 11$$

$$x + 7y + 4z = 6$$

$$0x - y - 8z = -3$$

Plugging the coefficients and constants into an augmented matrix gives

$$B = \begin{bmatrix} x_1 & y_1 & z_1 & C_1 \\ x_2 & y_2 & z_2 & C_2 \\ x_3 & y_3 & z_3 & C_3 \end{bmatrix}$$

$$B = \begin{bmatrix} 2 & 3 & -1 & 11 \\ 1 & 7 & 4 & 6 \\ 0 & -1 & -8 & -3 \end{bmatrix}$$

Simple row operations

In this lesson we'll look at some different ways to manipulate rows in matrices.

Switching two rows

We can switch any two rows in a matrix without changing the value of the matrix. In this matrix, we'll switch rows 1 and 2, which we write as $R_1 \leftrightarrow R_2$.

$$\begin{bmatrix} 3 & 2 & 7 \\ 1 & -6 & 0 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & -6 & 0 \\ 3 & 2 & 7 \end{bmatrix}$$

Keep in mind that we can also make multiple row switches. For instance, in this 3×3 matrix, we could first switch the second row with the third row,

$$\begin{bmatrix} 7 & 3 & 4 \\ 1 & 6 & 1 \\ 2 & 2 & 3 \end{bmatrix} \longrightarrow \begin{bmatrix} 7 & 3 & 4 \\ 2 & 2 & 3 \\ 1 & 6 & 1 \end{bmatrix}$$

and then switch the first row with the second row.

$$\begin{bmatrix} 7 & 3 & 4 \\ 1 & 6 & 1 \\ 2 & 2 & 3 \end{bmatrix} \longrightarrow \begin{bmatrix} 2 & 2 & 3 \\ 7 & 3 & 4 \\ 1 & 6 & 1 \end{bmatrix}$$

Let's do an example of a row switch in a larger matrix.

Example

Write the new matrix after $R_3 \leftrightarrow R_2$.

$$\begin{bmatrix} 2 & 3 & -1 & 11 \\ 1 & 7 & 4 & 6 \\ 0 & -1 & -8 & -3 \end{bmatrix}$$

The operation described by $R_3 \leftrightarrow R_2$ is switching row 2 with row 3. Nothing will happen to row 1. The matrix after $R_3 \leftrightarrow R_2$ is

$$\begin{bmatrix} 2 & 3 & -1 & 11 \\ 0 & -1 & -8 & -3 \\ 1 & 7 & 4 & 6 \end{bmatrix}$$

Multiplying a row by a constant

We can multiply any row by any non-zero constant without changing the value of the matrix. When we multiply by a constant, we call that constant a **scalar**, because it has the effect of "scaling" that row.

As an example, if we multiply through the first row of this matrix by 2, we don't actually change the value of the matrix.

$$\begin{bmatrix} 3 & 2 & 7 \\ 1 & -6 & 0 \end{bmatrix} \longrightarrow \begin{bmatrix} 2 \cdot 3 & 2 \cdot 2 & 2 \cdot 7 \\ 1 & -6 & 0 \end{bmatrix} \longrightarrow \begin{bmatrix} 6 & 4 & 14 \\ 1 & -6 & 0 \end{bmatrix}$$

How can it be true that multiplying a row by a scalar doesn't change the value of the matrix? Aren't the entires in the matrix now different? Well, to get an intuitive understanding of this, remember that a row in a matrix can represent a linear equation. For instance, the matrix

$$\begin{bmatrix} 6 & 4 & 14 \\ 1 & -6 & 0 \end{bmatrix}$$

could represent the linear system

$$6x + 4y = 14$$

$$x - 6y = 0$$

But given $6x + 4y = 14$, we know we can divide through the equation by 2, and it doesn't change the value of the equation, it just gives us $3x + 2y = 7$, which will produce the same solution as $6x + 4y = 14$.

So in the same way, we can pull the 2 back out of the matrix, undoing the operation from earlier,

$$\begin{bmatrix} 6 & 4 & 14 \\ 1 & -6 & 0 \end{bmatrix} \longrightarrow \begin{bmatrix} 2 \cdot 3 & 2 \cdot 2 & 2 \cdot 7 \\ 1 & -6 & 0 \end{bmatrix} \longrightarrow \begin{bmatrix} 3 & 2 & 7 \\ 1 & -6 & 0 \end{bmatrix}$$

and the matrix still has the same value.

Keep in mind that we're not limited to applying a scalar to only one row of the matrix. We can apply scalars to as many rows as we like, and the scalars don't even have to be the same (they just have to be non-zero).

For example, we can multiply the first row of the matrix by 2 (which we write as $2R_1 \to R_1$), and multiply the second row of the matrix by 3 (which we write as $3R_2 \to R_2$),

$$\begin{bmatrix} 3 & 2 & 7 \\ 1 & -6 & 0 \end{bmatrix} \longrightarrow \begin{bmatrix} 2 \cdot 3 & 2 \cdot 2 & 2 \cdot 7 \\ 3 \cdot 1 & 3 \cdot -6 & 3 \cdot 0 \end{bmatrix} \longrightarrow \begin{bmatrix} 6 & 4 & 14 \\ 3 & -18 & 0 \end{bmatrix}$$

and we still won't have changed the value of the matrix, since those constants could be factored right back out again.

Let's do an example that makes us apply a scalar and do a row switch at the same time.

Example

Write the new matrix after $3R_1 \leftrightarrow 2R_3$.

$$\begin{bmatrix} 2 & 3 & -1 & 11 \\ 1 & 7 & 4 & 6 \\ 0 & -1 & -8 & -3 \end{bmatrix}$$

The operation described by $3R_1 \leftrightarrow 2R_3$ is multiplying row 1 by a constant of 3, multiplying row 3 by a constant of 2, and then switching those two rows. Nothing will happen to row 2. The matrix after $3R_1$ is

$$\begin{bmatrix} 6 & 9 & -3 & 33 \\ 1 & 7 & 4 & 6 \\ 0 & -1 & -8 & -3 \end{bmatrix}$$

Then the matrix after $2R_3$ is

$$\begin{bmatrix} 6 & 9 & -3 & 33 \\ 1 & 7 & 4 & 6 \\ 0 & -2 & -16 & -6 \end{bmatrix}$$

Then the matrix after $3R_1 \leftrightarrow 2R_3$ is

$$\begin{bmatrix} 0 & -2 & -16 & -6 \\ 1 & 7 & 4 & 6 \\ 6 & 9 & -3 & 33 \end{bmatrix}$$

Adding a row to another row

It's also acceptable to add one row to another. Keep in mind though that this doesn't consolidate two rows into one. Instead, we replace a row with the sum of itself and another row. For instance, in this matrix,

$$\begin{bmatrix} 3 & 2 & 7 \\ 1 & -6 & 0 \end{bmatrix}$$

we could replace the first row with the sum of the first and second rows, which we write as $R_1 + R_2 \to R_1$. When we perform that operation, we're replacing the entries in row 1, but row 2 stays the same.

$$\begin{bmatrix} 3+1 & 2-6 & 7+0 \\ 1 & -6 & 0 \end{bmatrix} \longrightarrow \begin{bmatrix} 4 & -4 & 7 \\ 1 & -6 & 0 \end{bmatrix}$$

Let's do an example with row addition.

Example

Write the new matrix after $R_1 + 4R_3 \to R_1$.

$$\begin{bmatrix} 2 & 3 & -1 & 11 \\ 1 & 7 & 4 & 6 \\ 0 & -1 & -8 & -3 \end{bmatrix}$$

The operation described by $R_1 + 4R_3 \to R_1$ is multiplying row 3 by a constant of 4, adding that resulting row to row 1, and using that result to replace row 1. The row R_3 is

$$\begin{bmatrix} 0 & -1 & -8 & -3 \end{bmatrix}$$

So the row $4R_3$ would be

$$\begin{bmatrix} 4(0) & 4(-1) & 4(-8) & 4(-3) \end{bmatrix}$$

$$\begin{bmatrix} 0 & -4 & -32 & -12 \end{bmatrix}$$

Then the row $R_1 + 4R_3$ is

$$\begin{bmatrix} 2+0 & 3+(-4) & -1+(-32) & 11+(-12) \end{bmatrix}$$

$$\begin{bmatrix} 2 & -1 & -33 & -1 \end{bmatrix}$$

The matrix after $R_1 + 4R_3 \to R_1$, which is replacing row 1 with this row we just built, is

$$\begin{bmatrix} 2 & -1 & -33 & -1 \\ 1 & 7 & 4 & 6 \\ 0 & -1 & -8 & -3 \end{bmatrix}$$

Gauss-Jordan elimination and reduced row-echelon form

We know from Algebra that solving a system of linear equations is about finding the value of each unknown variable in the system. For instance, in this system,

$$-x-5y+z=17$$

$$-5x-5y+5z=5$$

$$2x+5y-3z=-10$$

we would solve the system by finding a set of values (x, y, z) that satisfy all three equations at the same time.

Whether by substitution, elimination, or graphing, which are the three methods we've used in the past to solve systems, we need to figure out that the values of x, y, and z that satisfy this system are $x=-1$, $y=-4$ and $z=-4$, or $(x, y, z)=(-1,-4,-4)$.

Row-echelon and reduced row-echelon forms

But now that we know how to use row operations to manipulate matrices, we have a new tool for solving systems of linear equations. In this lesson we want to look at how putting an augmented matrix into row-echelon form or reduced row-echelon form is another way to solve a system.

A matrix is in **row-echelon form (ref)** if

1. all rows consisting of only 0s are at the bottom of the matrix, and

2. the first non-zero entry in each row sits in a column to the right of the first non-zero entries in all the rows above it. In other words, the non-zero entries sit in a staircase pattern.

For a 3×3 augmented matrix, row-echelon form might look like

$$\begin{bmatrix} \mathbf{4} & 1 & 0 & = & 17 \\ 0 & \mathbf{2} & 5 & = & 10 \\ 0 & 0 & -\mathbf{3} & = & 2 \end{bmatrix}$$

The bolded first non-zero entry in each row is called a **pivot** or **pivot entry**, and each pivot is in a column to the right of the column that houses the pivot from each row above it. Any column that houses a pivot entry is called a **pivot column**.

If a matrix is

1. in row-echelon form, and

2. if all the pivot entries are equal to 1, and

3. if all of the non-pivot entries in the matrix are equal to 0 (other than the constants in the far-right column),

then the matrix is specifically in **reduced row-echelon form (rref)**. Reduced row-echelon form for a 3×3 matrix might look like

$$\begin{bmatrix} 1 & 0 & 0 & = & C_1 \\ 0 & 1 & 0 & = & C_2 \\ 0 & 0 & 1 & = & C_3 \end{bmatrix}$$

Here are examples of rref matrices that are 2×2, 3×3, and 4×4:

For 2×2:

$$\begin{bmatrix} 1 & 0 & = & C_1 \\ 0 & 1 & = & C_2 \end{bmatrix}$$

For 3×3:

$$\begin{bmatrix} 1 & 0 & 0 & = & C_1 \\ 0 & 1 & 0 & = & C_2 \\ 0 & 0 & 1 & = & C_3 \end{bmatrix}$$

For 4×4:

$$\begin{bmatrix} 1 & 0 & 0 & 0 & = & C_1 \\ 0 & 1 & 0 & 0 & = & C_2 \\ 0 & 0 & 1 & 0 & = & C_3 \\ 0 & 0 & 0 & 1 & = & C_4 \end{bmatrix}$$

Solving systems

If we try to solve the system

$$-x - 5y + z = 17$$

$$-5x - 5y + 5z = 5$$

$$2x + 5y - 3z = -10$$

using the substitution and elimination methods that we learned in Algebra, it'll be really tedious.

But if we can put this system into an augmented matrix, then we can use all of the row operations that we learned in the previous lesson in order to transform the original matrix into ref, or better yet, rref. Reduced row-echelon form will transform the system into

$$\begin{bmatrix} -1 & -5 & 1 & = & 17 \\ -5 & -5 & 5 & = & 5 \\ 2 & 5 & -3 & = & -10 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 0 & 0 & = & C_1 \\ 0 & 1 & 0 & = & C_2 \\ 0 & 0 & 1 & = & C_3 \end{bmatrix}$$

Let's talk for a second about why we would want to put the system of linear equations into an augmented matrix and then put the matrix into reduced row-echelon form.

Remember that the matrix

$$\begin{bmatrix} 1 & 0 & 0 & = & C_1 \\ 0 & 1 & 0 & = & C_2 \\ 0 & 0 & 1 & = & C_3 \end{bmatrix}$$

is still representing a system of linear equations. So if we've put the matrix into rref and then we pull back out the linear equations represented by the rref matrix, we get

$$1x + 0y + 0z = C_1$$

$$0x + 1y + 0z = C_2$$

$$0x + 0y + 1z = C_3$$

or just

$$x = C_1$$

$$y = C_2$$

$$z = C_3$$

In other words, from reduced row-echelon form, we automatically have the values of each variable, and we've solved the system. So what we're saying is that, if we put the matrix into its reduced row-echelon form, then we can pull out the value of each variable directly from the matrix.

Gaussian-Jordan elimination

So we know that it's helpful to put a matrix into reduced row-echelon form, and we've said that we can use matrix row operations to do this, but is there any systematic, orderly way that we go about these row operations?

Yes! **Gauss-Jordan elimination** is an **algorithm** (a specific set of steps that can be repeated over and over again) to get the matrix all the way to reduced row-echelon form. These are the steps:

1. Optional: Pull out any scalars from each row in the matrix.

2. If the first entry in the first row is 0, swap that first row with another row that has a non-zero entry in its first column. Otherwise, move to step 3.

3. Multiply through the first row by a scalar to make the leading entry equal to 1.

4. Add scaled multiples of the first row to every other row in the matrix until every entry in the first column, other than the leading 1 in the first row, is a 0.

5. Go back step 2 and repeat the process until the matrix is in reduced row-echelon form.

Let's walk through an example of how to use Gauss-Jordan elimination to rewrite an augmented matrix in rref so that we can pull out the value of each variable in the system represented by the matrix.

Example

Use Gauss-Jordan elimination to solve for the value of each variable.

$$\begin{bmatrix} -1 & -5 & 1 & = & 17 \\ -5 & -5 & 5 & = & 5 \\ 2 & 5 & -3 & = & -10 \end{bmatrix}$$

Remember first that this augmented matrix represents the linear system

$$-x - 5y + z = 17$$

$$-5x - 5y + 5z = 5$$

$$2x + 5y - 3z = -10$$

where the entries in the first column are the coefficients of x, the entries in the second column are the coefficients of y, and the entries in the third column are the coefficients of z. The entries in the fourth column are the constants.

Step 1:

Starting with the optional first step from Gauss-Jordan elimination, we could divide through the second row by 5, and that would reduce those values. After $(1/5)R_2 \to R_2$, the matrix is

$$\begin{bmatrix} -1 & -5 & 1 & = & 17 \\ -1 & -1 & 1 & = & 1 \\ 2 & 5 & -3 & = & -10 \end{bmatrix}$$

Step 2 (with the first row):

The first entry in the first row is non-zero, so there's no need to swap it with another row.

Step 3 (with the first row):

Multiply row 1 by -1 to get a leading 1 in the first row. After $-R_1 \to R_1$, the matrix is

$$\begin{bmatrix} -(-1) & -(-5) & -(1) & = & -(17) \\ -1 & -1 & 1 & = & 1 \\ 2 & 5 & -3 & = & -10 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 5 & -1 & = & -17 \\ -1 & -1 & 1 & = & 1 \\ 2 & 5 & -3 & = & -10 \end{bmatrix}$$

Step 4 (with the first row):

Replace row 2 with the sum of rows 1 and 2. After $R_1 + R_2 \to R_2$, the matrix is

$$\begin{bmatrix} 1 & 5 & -1 & = & -17 \\ 1-1 & 5-1 & -1+1 & = & -17+1 \\ 2 & 5 & -3 & = & -10 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 5 & -1 & = & -17 \\ 0 & 4 & 0 & = & -16 \\ 2 & 5 & -3 & = & -10 \end{bmatrix}$$

Replace row 3 with row 3 minus (2 times row 1). After $R_3 - 2R_1 \to R_3$, the matrix is

$$\begin{bmatrix} 1 & 5 & -1 & = & -17 \\ 0 & 4 & 0 & = & -16 \\ 2-2(1) & 5-2(5) & -3-2(-1) & = & -10-2(-17) \end{bmatrix}$$

$$\longrightarrow \begin{bmatrix} 1 & 5 & -1 & = & -17 \\ 0 & 4 & 0 & = & -16 \\ 0 & -5 & -1 & = & 24 \end{bmatrix}$$

We now have $1, 0, 0$ in the first column, which is exactly what we want. It's time to go back to step 2, but this time with the second row.

Step 2 (with the second row):

The second entry in the second row is non-zero, so there's no need to swap it with another row.

Step 3 (with the second row):

Multiply row 2 by $1/4$ to get a leading 1 in the second row. After $(1/4)R_2 \to R_2$, the matrix is

$$\begin{bmatrix} 1 & 5 & -1 & = & -17 \\ \frac{1}{4}(0) & \frac{1}{4}(4) & \frac{1}{4}(0) & = & \frac{1}{4}(-16) \\ 0 & -5 & -1 & = & 24 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 5 & -1 & = & -17 \\ 0 & 1 & 0 & = & -4 \\ 0 & -5 & -1 & = & 24 \end{bmatrix}$$

Step 4 (with the second row):

Replace row 1 with the sum of (-5 times row 2) and row 1. After $-5R_2 + R_1 \to R_1$, the matrix is

$$\begin{bmatrix} -5(0)+1 & -5(1)+5 & -5(0)-1 & = & -5(-4)-17 \\ 0 & 1 & 0 & = & -4 \\ 0 & -5 & -1 & = & 24 \end{bmatrix}$$

$$\longrightarrow \begin{bmatrix} 1 & 0 & -1 & = & 3 \\ 0 & 1 & 0 & = & -4 \\ 0 & -5 & -1 & = & 24 \end{bmatrix}$$

Replace row 3 with the sum of (5 times row 2) and row 3. After $5R_2 + R_3 \to R_3$, the matrix is

$$\begin{bmatrix} 1 & 5 & -1 & = & -17 \\ 0 & 1 & 0 & = & -4 \\ 5(0)+0 & 5(1)-5 & 5(0)-1 & = & 5(-4)+24 \end{bmatrix}$$

$$\longrightarrow \begin{bmatrix} 1 & 0 & -1 & = & 3 \\ 0 & 1 & 0 & = & -4 \\ 0 & 0 & -1 & = & 4 \end{bmatrix}$$

We now have 0, 1, 0 in the second column, which is exactly what we want. It's time to go back to step 2, but this time with the third row.

Step 2 (with the third row):

The third entry in the third row is non-zero, so there's no need to swap it with another row.

Step 3 (with the third row):

Multiply row 3 by -1 to get a leading 1 in the third row. After $-R_3 \to R_3$, the matrix is

$$\begin{bmatrix} 1 & 0 & -1 & = & 3 \\ 0 & 1 & 0 & = & -4 \\ -(0) & -(0) & -(-1) & = & -(4) \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 0 & -1 & = & 3 \\ 0 & 1 & 0 & = & -4 \\ 0 & 0 & 1 & = & -4 \end{bmatrix}$$

Step 4 (with the third row):

Replace row 1 with the sums of rows 1 and 3. After $R_1 + R_3 \to R_1$, the matrix is

$$\begin{bmatrix} 1+0 & 0+0 & -1+1 & = & 3-4 \\ 0 & 1 & 0 & = & -4 \\ 0 & 0 & 1 & = & -4 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 0 & 0 & = & -1 \\ 0 & 1 & 0 & = & -4 \\ 0 & 0 & 1 & = & -4 \end{bmatrix}$$

We now have 0, 0, 1 in the third column, which is exactly what we want. The matrix is now in reduced row-echelon form.

From this resulting matrix, we get the solution set

$$1x + 0y + 0z = -1$$

$$0x + 1y + 0z = -4$$

$$0x + 0y + 1z = -4$$

and then

$$x = -1$$

$$y = -4$$

$$z = -4$$

That's all it took to find that the solution to the system is $x = -1$, $y = -4$, and $z = -4$.

Matrix addition and subtraction

In the same way that we can add and subtract real numbers, we can also add and subtract matrices. But matrices must have the same dimensions in order for us to be able to add or subtract them.

For instance, given a 2×3 matrix, we can only add it to another 2×3 matrix or subtract it from another 2×3 matrix. We couldn't add a 2×3 matrix to a 2×2 matrix, and we couldn't subtract a 3×3 matrix from a 2×4 matrix.

To add matrices, we simply add entries from matching positions in each matrix. For instance, to add 2×2 matrices, we follow this pattern:

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} + \begin{bmatrix} 5 & 6 \\ 7 & 8 \end{bmatrix} = \begin{bmatrix} 1+5 & 2+6 \\ 3+7 & 4+8 \end{bmatrix} = \begin{bmatrix} 6 & 8 \\ 10 & 12 \end{bmatrix}$$

Or to add 2×4 matrices, we follow this pattern:

$$\begin{bmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \end{bmatrix} + \begin{bmatrix} 9 & 10 & 11 & 12 \\ 13 & 14 & 15 & 16 \end{bmatrix}$$

$$= \begin{bmatrix} 1+9 & 2+10 & 3+11 & 4+12 \\ 5+13 & 6+14 & 7+15 & 8+16 \end{bmatrix} = \begin{bmatrix} 10 & 12 & 14 & 16 \\ 18 & 20 & 22 & 24 \end{bmatrix}$$

Subtracting matrices follows the same pattern. We simply subtract entries from matching positions in the matrix. For instance, to add 2×2 matrices, we follow this pattern:

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} - \begin{bmatrix} 5 & 6 \\ 7 & 8 \end{bmatrix} = \begin{bmatrix} 1-5 & 2-6 \\ 3-7 & 4-8 \end{bmatrix} = \begin{bmatrix} -4 & -4 \\ -4 & -4 \end{bmatrix}$$

Matrices can also be part of an equation. For instance, in the same way that $x + 3 = 2$ gets solved as

$$x + 3 = 2$$

$$x = 2 - 3$$

$$x = -1$$

we can also solve an equation that contains matrices.

$$x + \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} = \begin{bmatrix} 6 & 8 \\ 10 & 12 \end{bmatrix}$$

$$x = \begin{bmatrix} 6 & 8 \\ 10 & 12 \end{bmatrix} - \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$$

$$x = \begin{bmatrix} 6 - 1 & 8 - 2 \\ 10 - 3 & 12 - 4 \end{bmatrix}$$

$$x = \begin{bmatrix} 5 & 6 \\ 7 & 8 \end{bmatrix}$$

In the simple linear equation $x + 3 = 2$, we subtracted 3 from both sides to get x by itself, and then simplified $2 - 3$ on the right to find $x = -1$. And we really did the same thing with the matrix equation. We subtracted the $1, 2, 3, 4$ matrix from both sides to get x by itself, and then simplified the difference of the matrices on the right to find a matrix value for x.

Let's do another example with matrix addition and subtraction.

Example

Solve for b.

$$\begin{bmatrix} 5 & -7 \\ -1 & 0 \end{bmatrix} + \begin{bmatrix} 3 & -4 \\ -1 & -6 \end{bmatrix} = b + \begin{bmatrix} 1 & 0 \\ 17 & 0 \end{bmatrix} - \begin{bmatrix} 2 & 4 \\ 11 & -9 \end{bmatrix}$$

Let's start with the matrix addition on the left side of the equation.

$$\begin{bmatrix} 5+3 & -7+(-4) \\ -1+(-1) & 0+(-6) \end{bmatrix} = b + \begin{bmatrix} 1 & 0 \\ 17 & 0 \end{bmatrix} - \begin{bmatrix} 2 & 4 \\ 11 & -9 \end{bmatrix}$$

$$\begin{bmatrix} 8 & -11 \\ -2 & -6 \end{bmatrix} = b + \begin{bmatrix} 1 & 0 \\ 17 & 0 \end{bmatrix} - \begin{bmatrix} 2 & 4 \\ 11 & -9 \end{bmatrix}$$

Subtract matrices on the right.

$$\begin{bmatrix} 8 & -11 \\ -2 & -6 \end{bmatrix} = b + \begin{bmatrix} 1-2 & 0-4 \\ 17-11 & 0-(-9) \end{bmatrix}$$

$$\begin{bmatrix} 8 & -11 \\ -2 & -6 \end{bmatrix} = b + \begin{bmatrix} -1 & -4 \\ 6 & 9 \end{bmatrix}$$

To isolate b, we'll subtract the matrix on the right from both sides in order to move it to the left.

$$\begin{bmatrix} 8 & -11 \\ -2 & -6 \end{bmatrix} - \begin{bmatrix} -1 & -4 \\ 6 & 9 \end{bmatrix} = b$$

$$\begin{bmatrix} 8-(-1) & -11-(-4) \\ -2-6 & -6-9 \end{bmatrix} = b$$

$$\begin{bmatrix} 9 & -7 \\ -8 & -15 \end{bmatrix} = b$$

So the value of b that makes the equation true is the matrix

$$b = \begin{bmatrix} 9 & -7 \\ -8 & -15 \end{bmatrix}$$

Properties of matrix addition and subtraction

When it comes to addition and subtraction, matrices follow the same rules as real numbers.

Addition

Matrix addition is commutative and associative. The fact that it's commutative means that we can add two matrices together in either order, and still get the same answer.

$$A + B = B + A$$

The fact that matrix addition is associative means that we can group the addition in different ways, and still get the same answer.

$$(A + B) + C = A + (B + C)$$

Subtraction

On the other hand, matrix subtraction is not commutative, and it's not associative. The fact that it's not commutative means that we won't get the same result if we subtract matrices in different orders.

$$A - B \neq B - A$$

The fact that matrix subtraction is not associative means that we can't group the subtraction in different ways and still get the same answer.

$$(A - B) - C \neq A - (B - C)$$

Scalar multiplication and zero matrices

We've learned about matrix addition and subtraction, and in this lesson we want to start looking at matrix multiplication. But before we multiply two matrices together, we'll tackle something a little simpler: scalar multiplication.

Scalar multiplication

We can multiply any matrix by a scalar. If we want to multiply a matrix by 3, then we distribute the scalar across every entry in the matrix, and the result of the scalar multiplication looks like

$$3\begin{bmatrix}6 & 2\\ -1 & -4\end{bmatrix}=\begin{bmatrix}3(6) & 3(2)\\ 3(-1) & 3(-4)\end{bmatrix}=\begin{bmatrix}18 & 6\\ -3 & -12\end{bmatrix}$$

Notice that this also translates to dividing through a matrix by a scalar. Because if we want to divide through a matrix by 6, that's the same as multiplying through the matrix by $1/6$, and we're right back to the same kind of scalar multiplication.

$$\frac{1}{6}\begin{bmatrix}6 & 2\\ -1 & -4\end{bmatrix}=\begin{bmatrix}\frac{1}{6}(6) & \frac{1}{6}(2)\\ \frac{1}{6}(-1) & \frac{1}{6}(-4)\end{bmatrix}=\begin{bmatrix}1 & \frac{1}{3}\\ -\frac{1}{6} & -\frac{2}{3}\end{bmatrix}$$

This translates as well to matrix equations, so let's do an example where we have to apply scalar multiplication before solving the matrix equation.

Example

Solve the equation for k.

$$2\begin{bmatrix} 5 & -7 \\ -1 & 0 \end{bmatrix} + k = -3\begin{bmatrix} 2 & 4 \\ 11 & -9 \end{bmatrix}$$

Apply the scalars to the matrices.

$$\begin{bmatrix} 2(5) & 2(-7) \\ 2(-1) & 2(0) \end{bmatrix} + k = \begin{bmatrix} -3(2) & -3(4) \\ -3(11) & -3(-9) \end{bmatrix}$$

$$\begin{bmatrix} 10 & -14 \\ -2 & 0 \end{bmatrix} + k = \begin{bmatrix} -6 & -12 \\ -33 & 27 \end{bmatrix}$$

Subtract the matrix on the left from both sides of the equation in order to isolate k.

$$k = \begin{bmatrix} -6 & -12 \\ -33 & 27 \end{bmatrix} - \begin{bmatrix} 10 & -14 \\ -2 & 0 \end{bmatrix}$$

$$k = \begin{bmatrix} -6 - 10 & -12 - (-14) \\ -33 - (-2) & 27 - 0 \end{bmatrix}$$

$$k = \begin{bmatrix} -16 & 2 \\ -31 & 27 \end{bmatrix}$$

Zero matrices

A **zero matrix** is a matrix with all zero values. These are all zero matrices with various dimensions:

$$\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \qquad \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \qquad \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

We always name the zero matrix with a capital O. And optionally, we can add a subscript with the dimensions of the zero matrix. Since the values in a zero matrix are all zeros, just having the dimensions of the zero matrix tells us what the entire matrix looks like. So $O_{2\times3}$ is

$$O_{2\times3} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

Adding and subtracting the zero matrix

Adding the zero matrix to any other matrix doesn't change that matrix's value, in the same way that subtracting the zero matrix from any other matrix doesn't change that matrix's value.

Just like with non-zero matrices, matrix dimensions still have to match before we can use addition or subtraction.

Adding the zero matrix:

$$\begin{bmatrix} 9 & -6 & 2 \\ 1 & 0 & -7 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 9 & -6 & 2 \\ 1 & 0 & -7 \end{bmatrix}$$

Subtracting the zero matrix:

$$\begin{bmatrix} 9 & -6 & 2 \\ 1 & 0 & -7 \end{bmatrix} - \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 9 & -6 & 2 \\ 1 & 0 & -7 \end{bmatrix}$$

Adding opposite matrices

Matrices K and $-K$ are **opposite matrices**. In other words, to get the opposite of a matrix, multiply it by a scalar of -1. So if

$$K = \begin{bmatrix} 9 & -6 & 2 \\ 1 & 0 & -7 \end{bmatrix}$$

then the opposite of K is

$$-K = (-1)\begin{bmatrix} 9 & -6 & 2 \\ 1 & 0 & -7 \end{bmatrix} = \begin{bmatrix} (-1)9 & (-1)(-6) & (-1)2 \\ (-1)1 & (-1)0 & (-1)(-7) \end{bmatrix} = \begin{bmatrix} -9 & 6 & -2 \\ -1 & 0 & 7 \end{bmatrix}$$

Adding opposite matrices results in the zero matrix.

$$K + (-K) = \begin{bmatrix} 9 & -6 & 2 \\ 1 & 0 & -7 \end{bmatrix} + \begin{bmatrix} -9 & 6 & -2 \\ -1 & 0 & 7 \end{bmatrix}$$

$$= \begin{bmatrix} 9 + (-9) & -6 + 6 & 2 + (-2) \\ 1 + (-1) & 0 + 0 & -7 + 7 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} = O_{2\times3}$$

Multiplying by a zero scalar

Multiplying any matrix by a scalar of 0 results in a zero matrix.

$$(0)\begin{bmatrix} 9 & -6 & 2 \\ 1 & 0 & -7 \end{bmatrix} = \begin{bmatrix} 9(0) & -6(0) & 2(0) \\ 1(0) & 0(0) & -7(0) \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} = O_{2\times3}$$

Matrix multiplication

We talked before about scalar multiplication, which is when we multiply a matrix by a real-number value. But **matrix multiplication** is what we do when we multiply two matrices together.

Dimensions matter

Based on what we've already learned about matrix addition and subtraction, we might think that multiplying matrices is just a matter of multiplying corresponding entries, since matrix addition is just a matter of adding corresponding entries, and matrix subtraction is just a matter of subtracting corresponding entries.

But in fact, we follow an entirely different process to multiply matrices, and we'll walk through exactly what that is in this lesson.

First, when we multiply two matrices A and B, the order matters. So $A \cdot B$ won't give us the same result as $B \cdot A$. Which means that matrices don't follow the commutative property of multiplication.

The reason the order matters is because of the way we multiply the matrices, which really depends on the dimensions. Here's the thing to remember about dimensions:

> We need the same number of columns in the first matrix as rows in the second matrix.

So for example, we can multiply a 3×2 matrix by any of these:

2×1

2×2

2×3

2×4

...

That's because, when we multiply one matrix by another, we multiply the rows in the first matrix by the columns in the second matrix. Let's say we want to multiply a 2×2 matrix called A by a 2×2 matrix called B.

$$A = \begin{bmatrix} 2 & 6 \\ 3 & -1 \end{bmatrix}$$

$$B = \begin{bmatrix} -4 & -2 \\ 1 & 0 \end{bmatrix}$$

If we call the first and second rows in A rows R_1 and R_2, and call the first and second columns in B columns C_1 and C_2,

$$A = \begin{array}{c} R_1 \rightarrow \\ R_2 \rightarrow \end{array} \begin{bmatrix} 2 & 6 \\ 3 & -1 \end{bmatrix}$$

$$B = \begin{array}{c} \begin{array}{cc} C_1 & C_2 \\ \downarrow & \downarrow \end{array} \\ \begin{bmatrix} -4 & -2 \\ 1 & 0 \end{bmatrix} \end{array}$$

then the product of A and B is

$$AB = \begin{bmatrix} R_1C_1 & R_1C_2 \\ R_2C_1 & R_2C_2 \end{bmatrix}$$

Let's look at each entry in the product:

$(AB)_{1,1}$ is the product of the **first row** and **first column**

$(AB)_{2,1}$ is the product of the **second row** and **first column**

$(AB)_{1,2}$ is the product of the **first row** and **second column**

$(AB)_{2,2}$ is the product of the **second row** and **second column**

The easy way to tell whether or not we can multiply matrices is to line up their dimensions. For instance, given that matrix A is a 2×3 and matrix B is a 3×4 matrix, we line up the product AB as

$$AB: \ 2 \times \boxed{3 \quad 3} \times 4$$

If the middle numbers match like they do here (they're both 3), then we can multiply the matrices to get a valid result, because we have the same number of columns in the first matrix as rows in the second matrix, which is what we said we needed. If we wanted to multiply B by A, we'd line up the product as

$$BA: \ 3 \times \boxed{4 \quad 2} \times 3$$

Because those middle numbers don't match (one is 4, the other is 2), we can't multiply the matrices. We don't have the same number of columns in the first matrix as rows in the second matrix, so the product isn't even defined.

Dimensions of the product

Now that we know how to determine whether or not the product of two matrices will be defined, let's talk about the dimensions of the product.

We said before that because we have the same number of columns in the first matrix as rows in the second matrix, AB will be defined in this case:

$$AB: \ 2\times\boxed{3 \quad 3}\times 4$$

Once we know that the product AB is defined, we can also quickly know the dimensions of the resulting product. To get those dimensions, just take the number of rows from the first matrix and the number of columns from the second matrix.

$$AB: \ \boxed{2}\times 3 \quad 3\times\boxed{4}$$

So the dimensions of the product AB will be 2×4. In other words, a 2×3 matrix multiplied by a 3×4 matrix will always result in a 2×4 matrix.

Example

If matrix A is 2×2 and matrix B is 4×2, say whether AB or BA is defined, and give the dimensions of the product if it is defined.

Line up the dimensions for the products AB and BA.

$$AB: \ 2\times\boxed{2 \quad 4}\times 2$$

$$BA: \ 4 \times \boxed{2 \quad 2} \times 2$$

For AB, the middle numbers don't match, so that product isn't defined. For BA, the middle numbers match, so that product is defined.

The dimensions of BA are given by the outside numbers,

$$BA: \boxed{4} \times 2 \quad 2 \times \boxed{2}$$

so the dimensions of BA will be 4×2.

Using the dot product to multiply matrices

The **dot product** is the tool we'll use to multiply an entire row of one matrix by an entire column of another matrix. When we're calculating a dot product, we want to think about ordered pairs.

For instance, we said that when we take the product of A and B,

$$A = \begin{bmatrix} 2 & 6 \\ 3 & -1 \end{bmatrix}$$

$$B = \begin{bmatrix} -4 & -2 \\ 1 & 0 \end{bmatrix}$$

the first entry we'll need to find is the product of the first row in A and the first column in B. The first row in A is the ordered pair $(2,6)$, and the first column in B is the ordered pair $(-4,1)$.

To take the dot product of these ordered pairs, we take the product of the first values, and then add that result to the product of the second values. In other words, the dot product of $(2,6)$ and $(-4,1)$ is

$$2(-4)+6(1)$$

$$-8+6$$

$$-2$$

Now that we understand how to use the dot product to multiply matrices, we can say that the product of matrices A and B is

$$AB=\begin{bmatrix}2 & 6\\ 3 & -1\end{bmatrix}\cdot\begin{bmatrix}-4 & -2\\ 1 & 0\end{bmatrix}$$

$$AB=\begin{bmatrix}R_1C_1 & R_1C_2\\ R_2C_1 & R_2C_2\end{bmatrix}$$

$$AB=\begin{bmatrix}2(-4)+6(1) & 2(-2)+6(0)\\ 3(-4)+(-1)(1) & 3(-2)+(-1)(0)\end{bmatrix}$$

$$AB=\begin{bmatrix}-8+6 & -4+0\\ -12+(-1) & -6+0\end{bmatrix}$$

$$AB=\begin{bmatrix}-2 & -4\\ -13 & -6\end{bmatrix}$$

Properties of matrix multiplication

When it comes to multiplication, matrices don't follow the same rules as real numbers.

As we already mentioned, matrix multiplication is not commutative. The fact that it's not commutative means that we can't multiply matrices in a different order and still get the same answer.

$$AB \neq BA$$

Matrix multiplication is associative. The fact that it's associative means that we can group the multiplication in different ways, and still get the same answer, as long as we don't change the order.

$$(AB)C = A(BC)$$

Matrix multiplication is distributive. The fact that it's distributive means that we can distribute multiplication across another value.

$$A(B + C) = AB + AC$$

$$(B + C)A = BA + CA$$

$$A(B - C) = AB - AC$$

$$(B - C)A = BA - CA$$

When it comes to the zero matrix, it doesn't matter whether we multiply a matrix by the zero matrix, or multiply the zero matrix by a matrix; we'll get the O matrix either way. But the dimensions of the zero matrix may change, depending on whether it's the first or second matrix in the multiplication.

- When $OA = O$, the zero matrix O must have the same number of columns as A has rows.

- When $AO = O$, the zero matrix O must have the same number of rows as A has columns.

Identity matrices

We already know that multiplying a matrix by a scalar of 0 will turn the matrix into a matrix with all 0 entries,

$$0\begin{bmatrix}6 & 2\\ -1 & -4\end{bmatrix}=\begin{bmatrix}0(6) & 0(2)\\ 0(-1) & 0(-4)\end{bmatrix}=\begin{bmatrix}0 & 0\\ 0 & 0\end{bmatrix}$$

and that multiplying a matrix by a scalar of 1 won't change the original matrix.

$$1\begin{bmatrix}6 & 2\\ -1 & -4\end{bmatrix}=\begin{bmatrix}1(6) & 1(2)\\ 1(-1) & 1(-4)\end{bmatrix}=\begin{bmatrix}6 & 2\\ -1 & -4\end{bmatrix}$$

But is there any matrix that we can multiply by another, without changing the value of that second matrix? The answer is yes, and we call it the **identity matrix**.

We always call the identity matrix I, and it's always a square matrix, like 2×2, 3×3, 4×4, etc. For that reason, it's common to abbreviate $I_{2\times 2}$ as just I_2, $I_{3\times 3}$ as just I_3, etc.

We'll talk more later about why the identity matrix is always square. But for now, here's what an identity matrix looks like:

$$I_2=\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}\qquad I_3=\begin{bmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\end{bmatrix}\qquad I_4=\begin{bmatrix}1 & 0 & 0 & 0\\ 0 & 1 & 0 & 0\\ 0 & 0 & 1 & 0\\ 0 & 0 & 0 & 1\end{bmatrix}$$

When we multiply the identity matrix by another matrix, we don't change the value of the other matrix. Let's see what happens when we multiply the identity matrix by another matrix.

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 7 & 3 & 4 \\ 1 & 6 & 1 \\ 2 & 2 & 3 \end{bmatrix}$$

$$= \begin{bmatrix} 7(1)+1(0)+2(0) & 3(1)+6(0)+2(0) & 4(1)+1(0)+3(0) \\ 7(0)+1(1)+2(0) & 3(0)+6(1)+2(0) & 4(0)+1(1)+3(0) \\ 7(0)+1(0)+2(1) & 3(0)+6(0)+2(1) & 4(0)+1(0)+3(1) \end{bmatrix}$$

$$= \begin{bmatrix} 7+0+0 & 3+0+0 & 4+0+0 \\ 0+1+0 & 0+6+0 & 0+1+0 \\ 0+0+2 & 0+0+2 & 0+0+3 \end{bmatrix}$$

$$= \begin{bmatrix} 7 & 3 & 4 \\ 1 & 6 & 1 \\ 2 & 2 & 3 \end{bmatrix}$$

Notice how multiplying by the identity matrix I_3 didn't change the value of the second matrix.

Dimensions of the identity matrix

Let's prove to ourselves that the identity matrix will always be square. We'll start with some other matrix, like this 3×2:

$$A = \begin{bmatrix} 4 & -6 \\ 1 & 1 \\ -2 & 9 \end{bmatrix}$$

Because we know that the identity matrix won't change the value of A, we can set up this equation:

$$I \cdot \begin{bmatrix} 4 & -6 \\ 1 & 1 \\ -2 & 9 \end{bmatrix} = \begin{bmatrix} 4 & -6 \\ 1 & 1 \\ -2 & 9 \end{bmatrix}$$

If we think about the dimensions of A in the context of this equation, we'll see why the identity matrix must be square. The dimensions of A are 3×2, so let's substitute those into the equation to get a visual picture of the dimensions.

$$I \cdot 3 \times 2 = 3 \times 2$$

Then let's break down the dimensions of the identity matrix as rows $\times$ columns, or $R \times C$.

$$R \times C \cdot 3 \times 2 = 3 \times 2$$

First, we know that in order to be able to multiply matrices at all, we need the same number of columns in the first matrix as rows in the second matrix. So we know the identity matrix must have 3 columns.

$$R \times \boxed{3 \cdot 3} \times 2 = 3 \times 2$$

We also know that the dimensions of the resulting matrix on the right come from the rows of the first matrix and the columns of the second matrix.

$$\boxed{R} \times 3 \cdot 3 \times \boxed{2} = \boxed{3} \times \boxed{2}$$

So we know the identity matrix must have 3 rows.

$$3 \times 3 \cdot 3 \times 2 = 3 \times 2$$

Therefore, the identity matrix in this case turns out to be a square 3×3 matrix. And this works for a matrix with any dimensions. Here are some examples:

For a 2×4 matrix, the identity matrix has to be I_2:

$$I \cdot 2 \times 4 = 2 \times 4$$

$$R \times C \cdot 2 \times 4 = 2 \times 4$$

$$\boxed{2} \times \boxed{2 \cdot 2} \times \boxed{4} = \boxed{2} \times \boxed{4}$$

$$\boxed{2 \times 2} \cdot 2 \times 4 = 2 \times 4$$

For a 3×1 matrix, the identity matrix has to be I_3:

$$I \cdot 3 \times 1 = 3 \times 1$$

$$R \times C \cdot 3 \times 1 = 3 \times 1$$

$$\boxed{3} \times \boxed{3 \cdot 3} \times \boxed{1} = \boxed{3} \times \boxed{1}$$

$$\boxed{3 \times 3} \cdot 3 \times 1 = 3 \times 1$$

Let's do an example where we find the identity matrix we'll need to multiply by another matrix.

Example

Choose the correct identity matrix for the given matrix, and then find the product of the identity matrix and the given matrix.

$$A = \begin{bmatrix} 4 & -6 \\ 1 & 1 \\ -2 & 9 \end{bmatrix}$$

We know that the dimensions of the identity matrix are always given by the number of rows in the other matrix, which means the identity matrix we need for matrix A will be I_3. We can prove it, too:

$$I \cdot 3 \times 2 = 3 \times 2$$

$$R \times C \cdot 3 \times 2 = 3 \times 2$$

$$\boxed{R} \times \boxed{3 \cdot 3} \times \boxed{2} = \boxed{3} \times \boxed{2}$$

$$\boxed{3 \times 3} \cdot 3 \times 2 = 3 \times 2$$

So the identity matrix is 3×3.

$$I_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Then the product of I_3 and matrix A is

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 4 & -6 \\ 1 & 1 \\ -2 & 9 \end{bmatrix}$$

$$= \begin{bmatrix} 4(1) + 1(0) - 2(0) & -6(1) + 1(0) + 9(0) \\ 4(0) + 1(1) - 2(0) & -6(0) + 1(1) + 9(0) \\ 4(0) + 1(0) - 2(1) & -6(0) + 1(0) + 9(1) \end{bmatrix}$$

$$= \begin{bmatrix} 4+0-0 & -6+0+0 \\ 0+1-0 & 0+1+0 \\ 0+0-2 & 0+0+9 \end{bmatrix}$$

$$= \begin{bmatrix} 4 & -6 \\ 1 & 1 \\ -2 & 9 \end{bmatrix}$$

As we expected, we get back to matrix A after multiplying it by the identity matrix I_3.

Properties of the identity matrix

When it comes to the identity matrix, it doesn't matter whether we multiply a matrix by the identity matrix, or multiply the identity matrix by a matrix; we'll get the original matrix either way. But the dimensions of the identity matrix may change, depending on whether it's the first or second matrix in the multiplication.

$IA = A$ I has the same number of columns as A has rows

$AI = A$ I has the same number of rows as A has columns

Transformations

Matrices can be extremely useful when it comes to describing what would otherwise be complex changes to two-dimensional space.

For instance, let's say we want to know what happens if we take every point in the coordinate plane, shift it up vertically by 2 units, and stretch it out horizontally by 5 units.

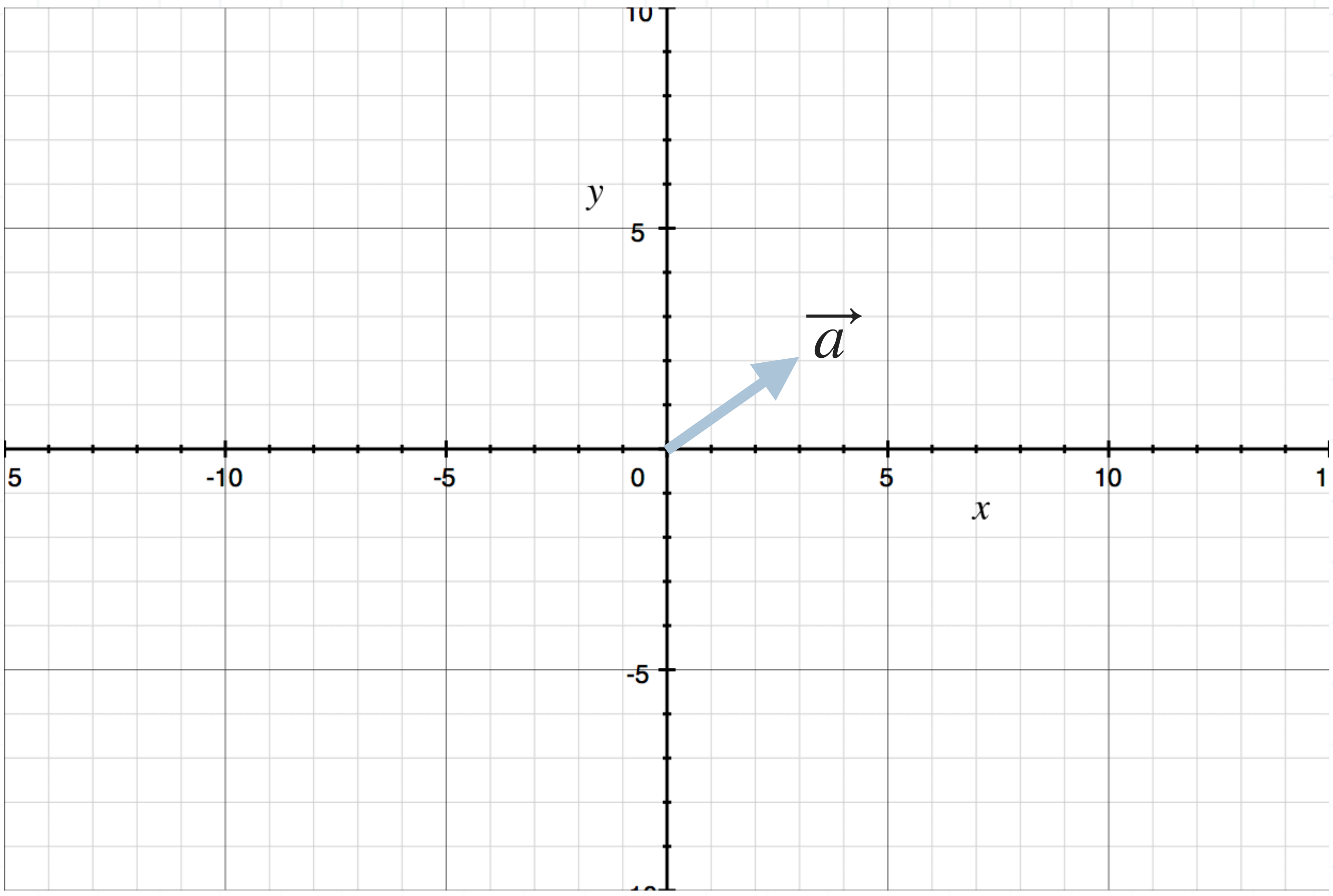
Up to now, we don't have a simple way to describe this change mathematically. But that's where matrices come in. They allow us to organize the transformation information into a transformation matrix, which will tell us exactly how every point in the plane should move.

Transformation of a single point

Let's say we have the point in space $(3,2)$. Keep in mind that for the purpose of this lesson, it's also not very different to think about that point as a vector that starts at the origin and extends out to the point $(3,2)$. Let's call the vector $\overrightarrow{a}$,

$$\overrightarrow{a}=\begin{bmatrix}3\\2\end{bmatrix}$$

and then sketch it in the plane as



We can apply a transformation matrix to the vector (or coordinate point), and change it into another vector (or coordinate point). Let's say we use the transformation matrix

$$M = \begin{bmatrix} -2 & 4 \\ 3 & 0 \end{bmatrix}$$

and apply it to the vector $\vec{a}$. Then the transformation of $\vec{a}$ by M will be the multiplication of $\vec{a}$ by M.

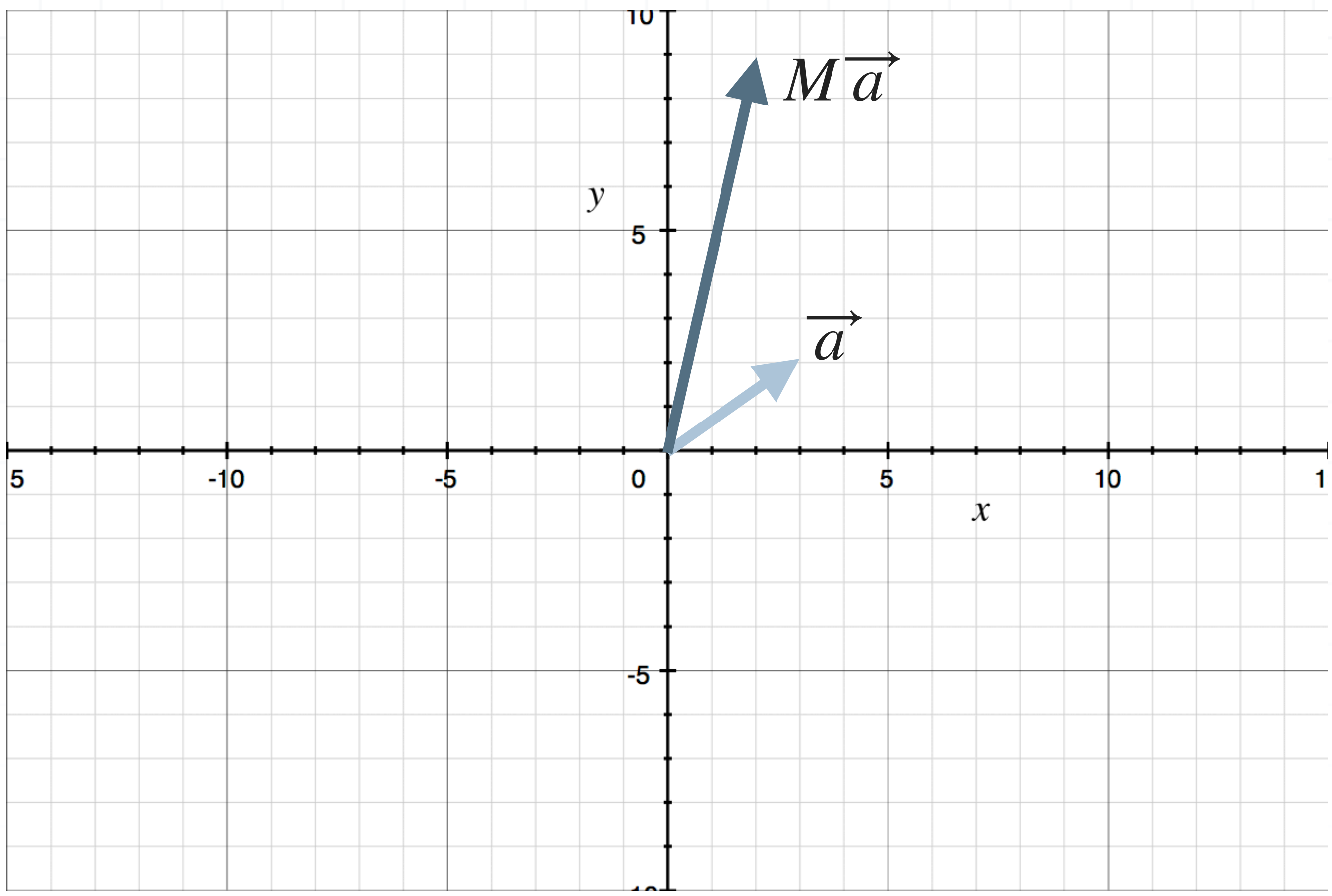
$$M\vec{a} = \begin{bmatrix} -2 & 4 \\ 3 & 0 \end{bmatrix} \begin{bmatrix} 3 \\ 2 \end{bmatrix}$$

$$M\vec{a} = \begin{bmatrix} -2(3) + 4(2) \\ 3(3) + 0(2) \end{bmatrix}$$

$$M\vec{a} = \begin{bmatrix} -6 + 8 \\ 9 + 0 \end{bmatrix}$$

$$M\vec{a} = \begin{bmatrix} 2 \\ 9 \end{bmatrix}$$

What this means is that the values in matrix M cause the vector $\vec{a} = (3,2)$ to transform into the vector $M\vec{a} = (2,9)$.



Transforming a figure

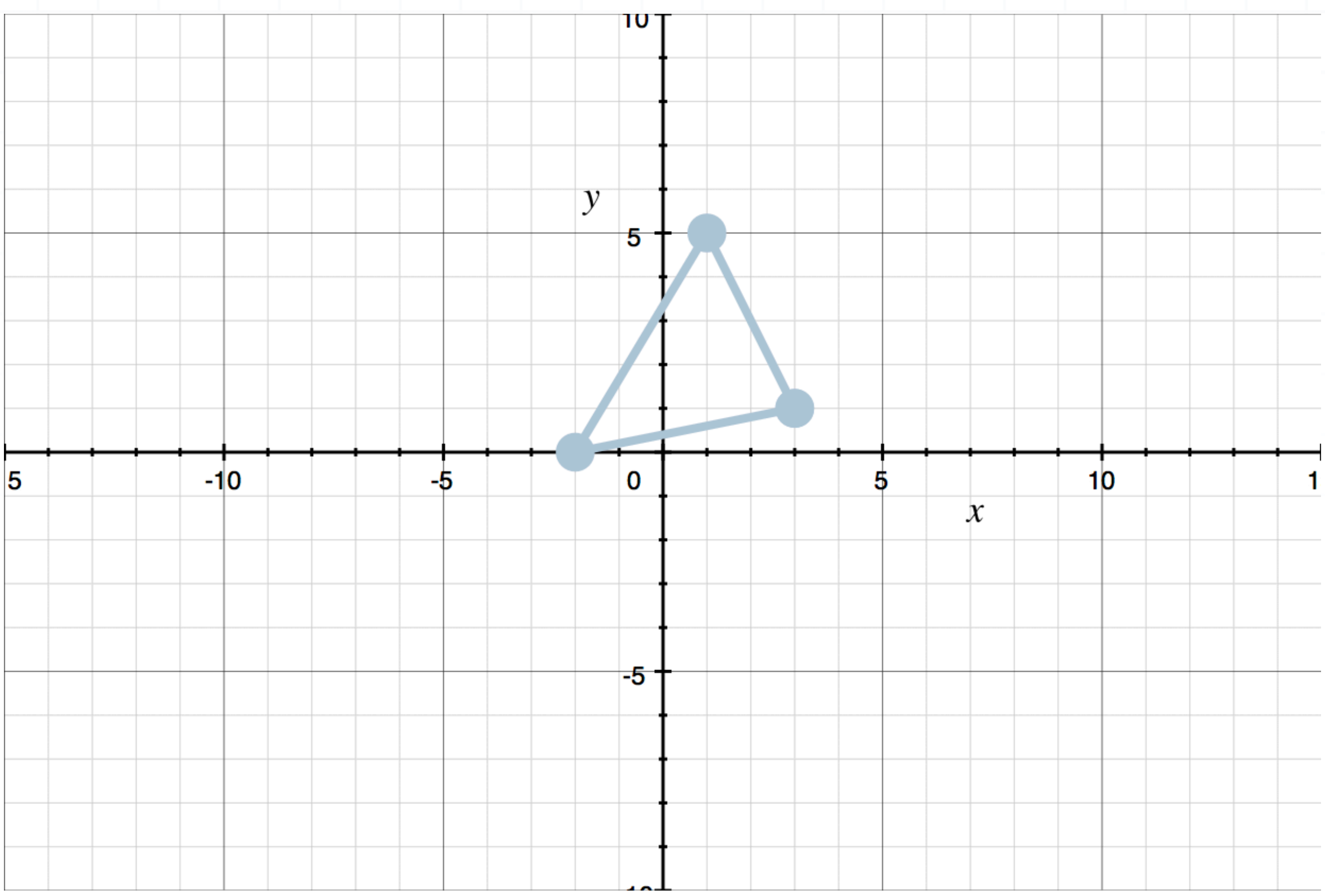
If instead of being given the single point or vector $\vec{a}$, we'd been given a set of points that represent the vertices of a figure, whether that figure is a triangle, quadrilateral, pentagon, or any other polygon, we can apply a transformation matrix to the set of points, and transform the figure into a different figure.

Example

A triangle is defined by its vertices $(3,1)$, $(1,5)$, and $(-2,0)$. Apply the transformation matrix M and give the vertices of the triangle after the transformation.

$$M = \begin{bmatrix} 1 & -3 \\ 4 & 0 \end{bmatrix}$$

The graph of the triangle with the given vertices is



First, pull the vertices of the triangle into a matrix. Put the x-values from the coordinate points into the first row, then put corresponding y-values into the second row, such that each column of the matrix represents the coordinate point of one vertex of the triangle.

$$\begin{bmatrix} 3 & 1 & -2 \\ 1 & 5 & 0 \end{bmatrix}$$

Then multiply the transformation matrix by the point-set matrix.

$$M \cdot \begin{bmatrix} 3 & 1 & -2 \\ 1 & 5 & 0 \end{bmatrix}$$

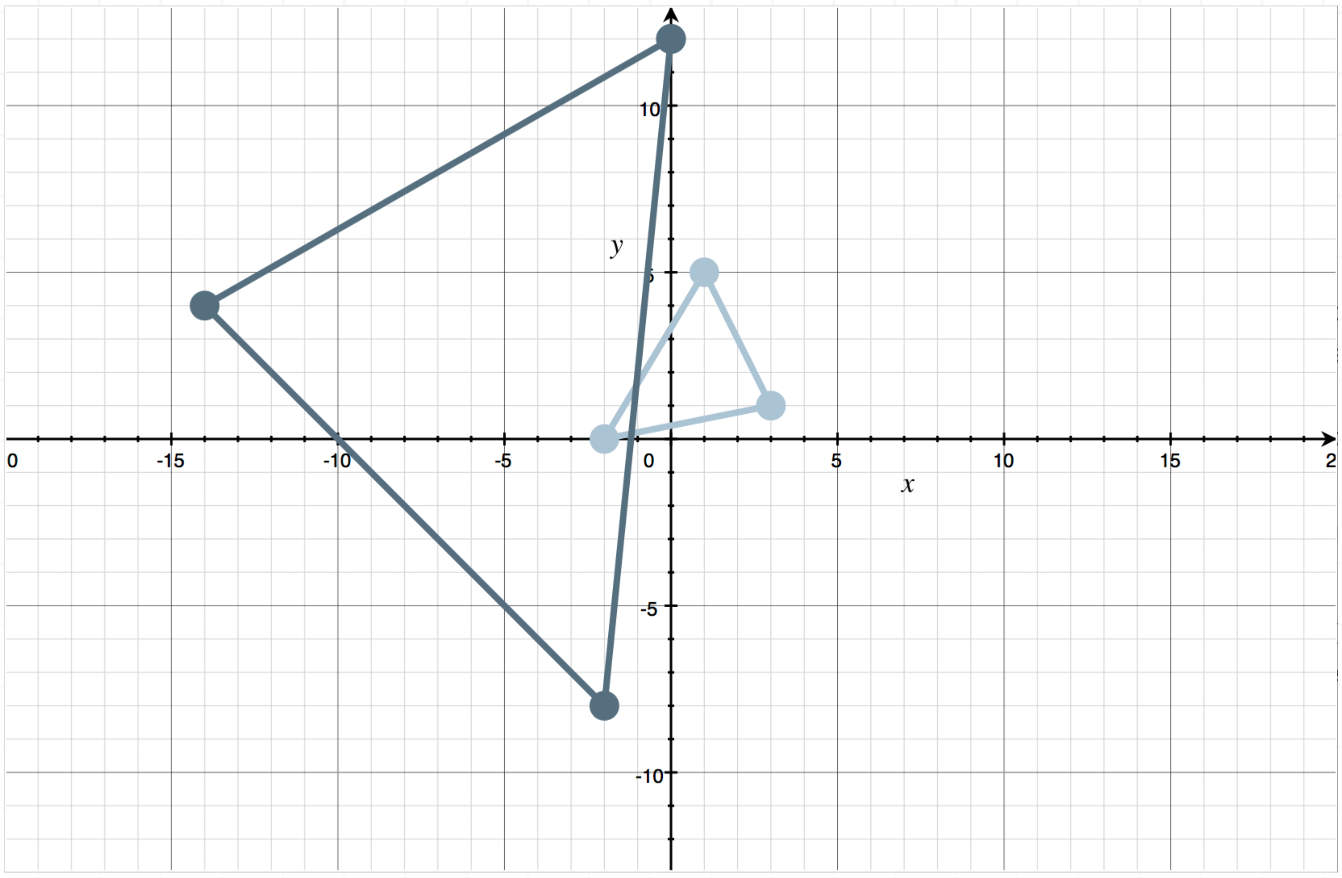
$$\begin{bmatrix} 1 & -3 \\ 4 & 0 \end{bmatrix} \begin{bmatrix} 3 & 1 & -2 \\ 1 & 5 & 0 \end{bmatrix}$$

$$\begin{bmatrix} 1(3) - 3(1) & 1(1) - 3(5) & 1(-2) - 3(0) \\ 4(3) + 0(1) & 4(1) + 0(5) & 4(-2) + 0(0) \end{bmatrix}$$

$$\begin{bmatrix} 3 - 3 & 1 - 15 & -2 - 0 \\ 12 + 0 & 4 + 0 & -8 + 0 \end{bmatrix}$$

$$\begin{bmatrix} 0 & -14 & -2 \\ 12 & 4 & -8 \end{bmatrix}$$

This transformed matrix gives us the vertices of the transformed triangle, which are $(0,12)$, $(-14,4)$, and $(-2,-8)$, and the graphs of the original and transformed triangles together are



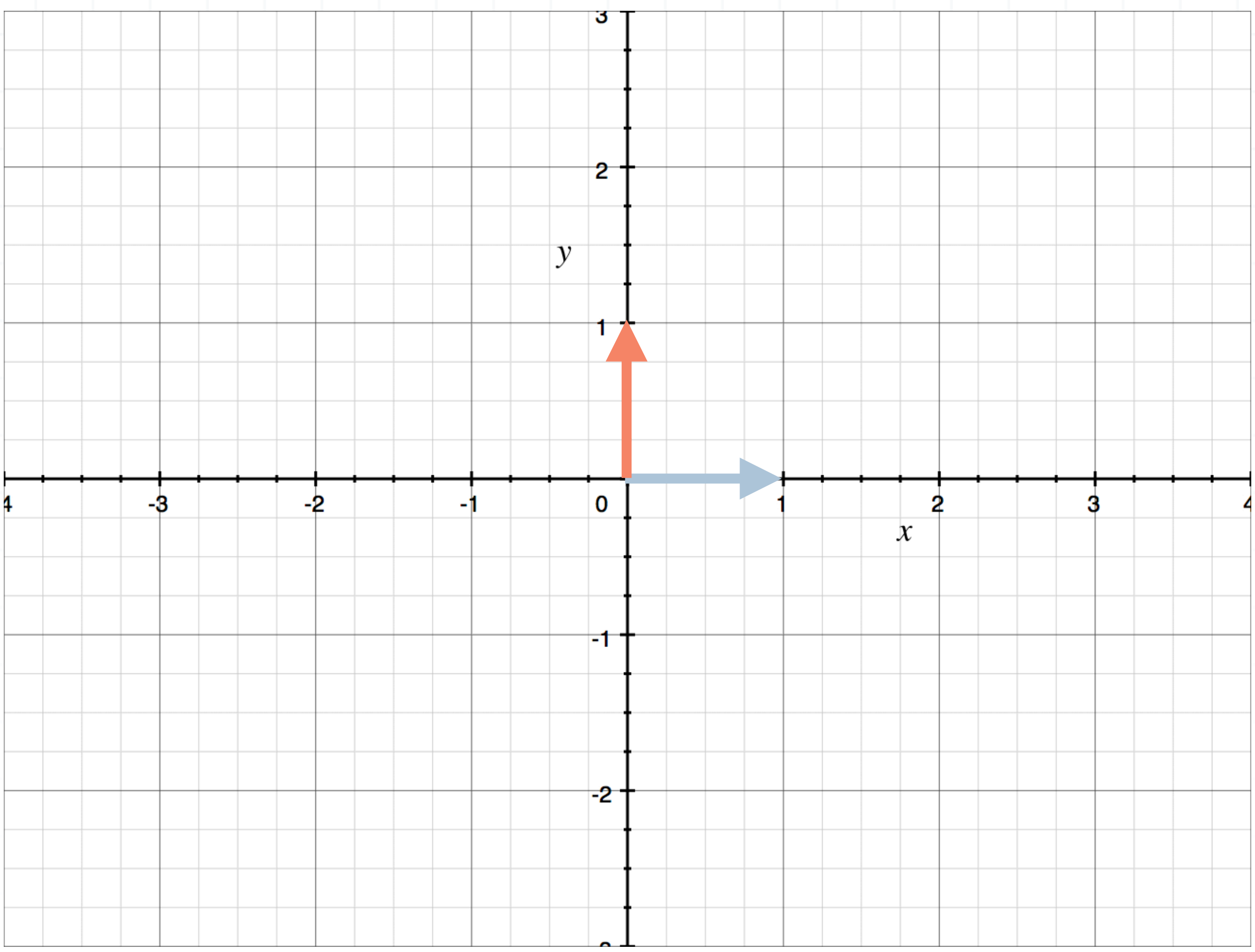
Understanding the transformation matrix

Think about transformations as a series of shifts, stretches, compressions, rotations, etc., that move a point from one spot to another. For instance, we know from the last example that the transformation matrix

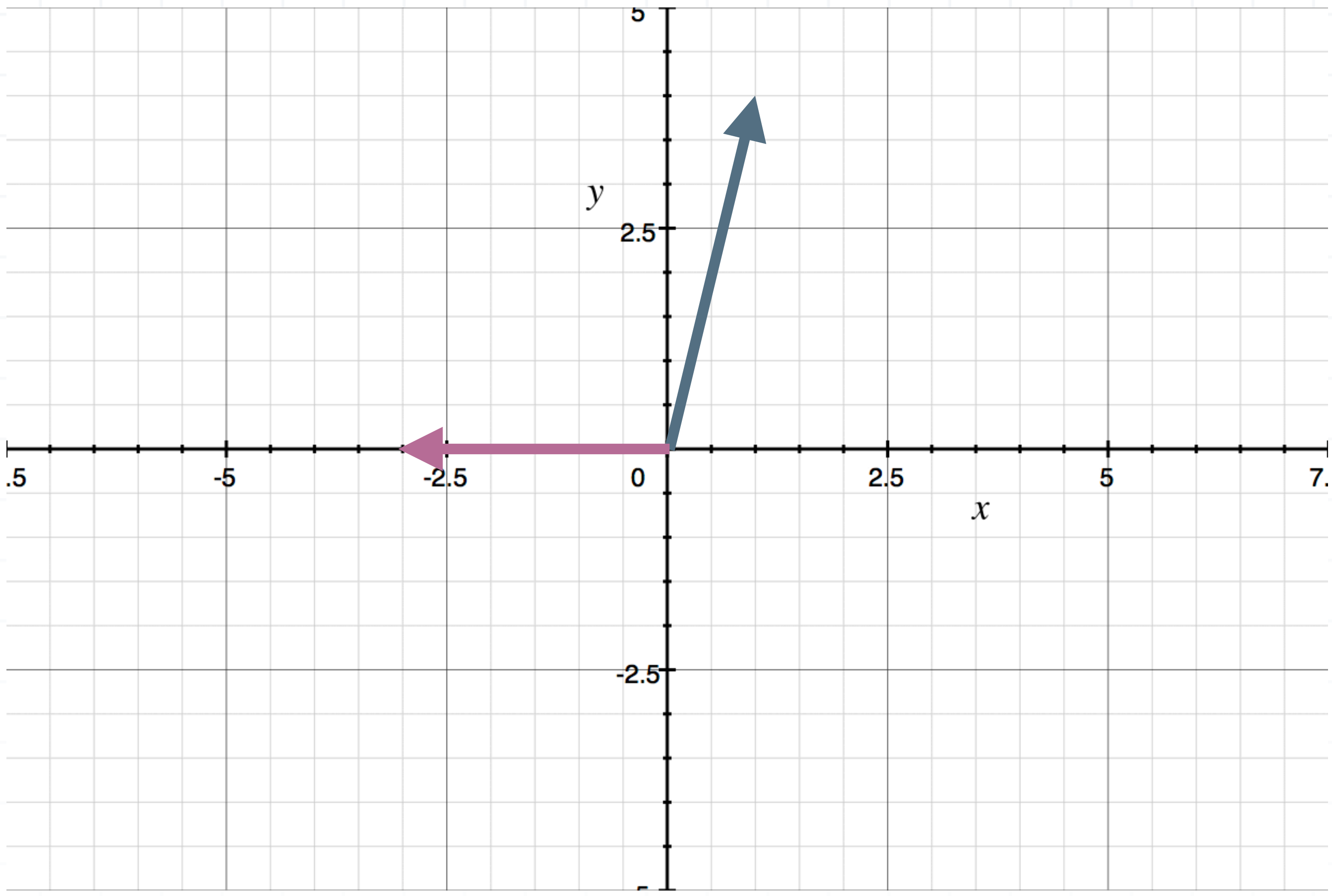
$$M = \begin{bmatrix} 1 & -3 \\ 4 & 0 \end{bmatrix}$$

transformed the light blue triangle into the dark blue triangle. Now we want to look at the entries in the transformation matrix M to see how they work together to change the light blue triangle into the dark blue triangle.

In a 2×2 transformation matrix, the first column (in this case $(1,4)$) tells us where the unit vector $(1,0)$ will land after the transformation. The second column (in this case $(-3,0)$) tells us where the unit vector $(0,1)$ will land after the transformation. In other words, given the unit vectors $(1,0)$ in light blue and $(0,1)$ in red,



the transformation matrix M changes the light blue vector $(1,0)$ into the dark blue vector $(1,4)$, and changes the red vector $(0,1)$ into the purple vector $(-3,0)$.

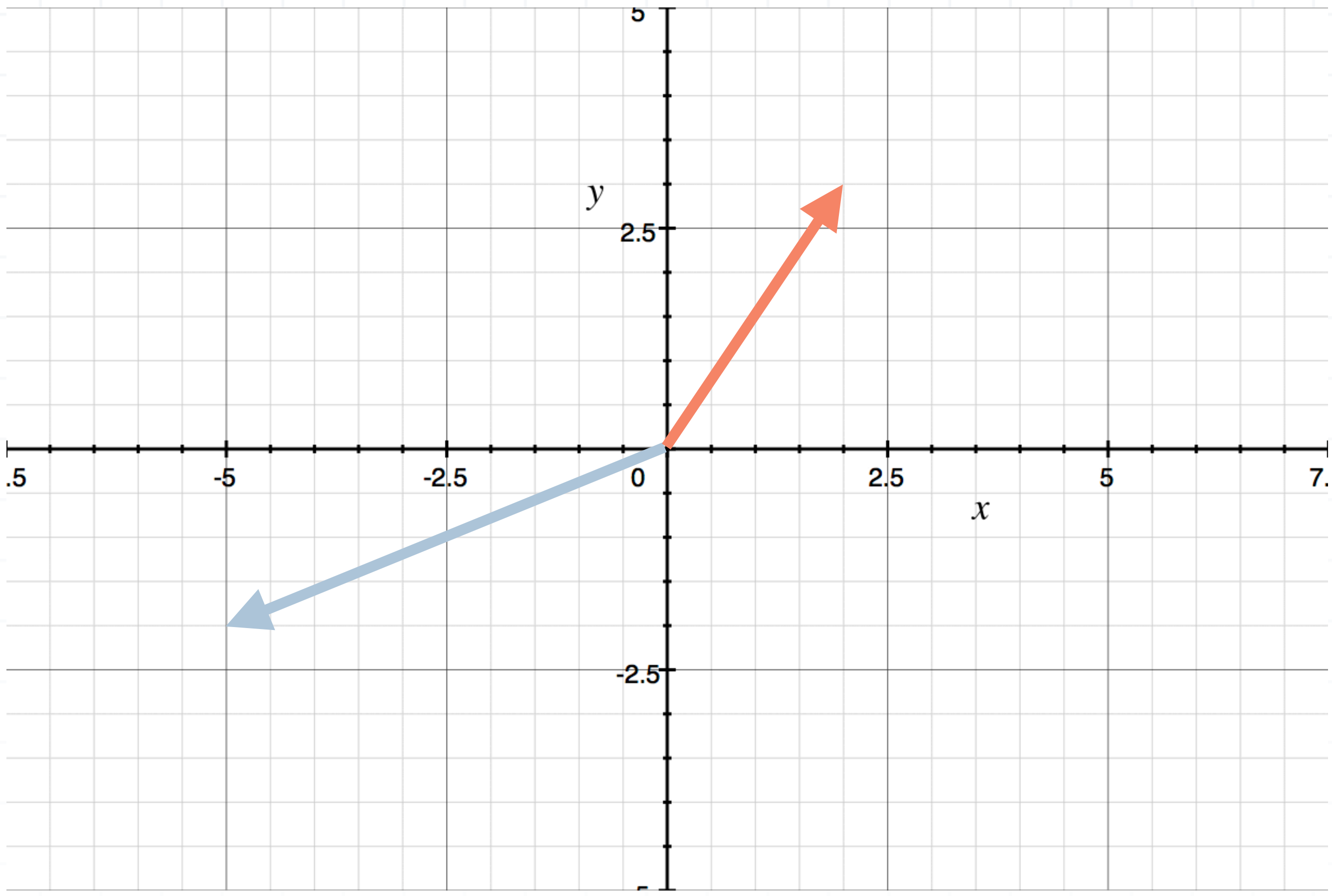


We already know how to take the transformation matrix and apply it to a point or vector, like we did when we transformed the light blue triangle into the dark blue triangle.

But we can also work backwards from a transformed figure to the original figure to determine what the transformation matrix must have been.

Example

The graph shows the light blue unit vector $(1,0)$ and the red unit vector $(0,1)$ after a transformation has been applied. Find the matrix that did the transformation.



The light blue vector now points to $(-5,-2)$. Since the first column of the transformation matrix represents where the unit vector $(1,0)$ lands after a transformation, we can fill in the first column of the transformation matrix.

$$\begin{bmatrix} -5 & \\ -2 & \end{bmatrix}$$

The red vector now points to $(2,3)$. Since the second column of a transformation matrix represents where the unit vector $(0,1)$ lands after a transformation, we can fill in the second column of the transformation matrix.

$$\begin{bmatrix} -5 & 2 \\ -2 & 3 \end{bmatrix}$$

This is the matrix that describes the transformation happening in coordinate space when $(1,0)$ moves to $(-5,-2)$ and when $(0,1)$ moves to $(2,3)$.

There's a really important conclusion to this last example problem. The transformation matrix we found doesn't just transform the individual points $(1,0)$ and $(0,1)$, it models the transformation of every point in the coordinate plane!

Therefore, armed with this transformation matrix, we can now figure out the transformed location of any other point. For instance, let's say we want to know what happens to $(5,3)$. Simply multiply that vector by the transformation matrix.

$$\begin{bmatrix} -5 & 2 \\ -2 & 3 \end{bmatrix} \begin{bmatrix} 5 \\ 3 \end{bmatrix}$$

$$\begin{bmatrix} -5(5) + 2(3) \\ -2(5) + 3(3) \end{bmatrix}$$

$$\begin{bmatrix} -25 + 6 \\ -10 + 9 \end{bmatrix}$$

$$\begin{bmatrix} -19 \\ -1 \end{bmatrix}$$

Under this transformation, $(5,3)$ will go to $(-19,-1)$. So we can use the transformation matrix to transform any point in coordinate space.

Matrix inverses, and invertible and singular matrices

We've talked about matrix addition, subtraction, and multiplication, and now in this lesson we'll work on matrix division. But we need to address a couple of other things first.

The determinant

The determinant of a matrix is a value we'll use frequently whenever we deal with matrices in any branch of math, so it's important to know how to find it. A matrix is always given in brackets, like

$$\begin{bmatrix} -2 & 4 \\ 3 & 0 \end{bmatrix}$$

but when we want to indicate the determinant instead, we use straight lines.

$$\begin{vmatrix} -2 & 4 \\ 3 & 0 \end{vmatrix}$$

To calculate the **determinant**, we multiply the value in the upper left by the value in the lower right, then subtract the product of the upper right and lower left. So the determinant for this matrix would be given by

$$\begin{vmatrix} -2 & 4 \\ 3 & 0 \end{vmatrix} = (-2)(0) - (4)(3)$$

$$\begin{vmatrix} -2 & 4 \\ 3 & 0 \end{vmatrix} = 0 - 12$$

$$\begin{vmatrix} -2 & 4 \\ 3 & 0 \end{vmatrix} = -12$$

Division as multiplication by the reciprocal

To build toward matrix division, we want to remember that dividing by some value is the same as multiplying by the reciprocal of that value. For instance, dividing by 4 is the same as multiplying by $1/4$. So if k is a real number, then we know that

$$k \cdot \frac{1}{k} = 1$$

If we call $1/k$ the inverse of k and instead write it as k^{-1}, then we could rewrite this equation as

$$kk^{-1} = 1$$

and read this as "k multiplied by the inverse of k is 1." What we want to know now is whether this is also true for matrices. If we divide matrix K by matrix K, or multiply matrix K by its inverse, do we get back to 1? In other words, we're trying to prove that

$$K \cdot \frac{1}{K} = I \text{ or } KK^{-1} = I$$

where I is the identity matrix, the matrix equivalent of 1.

Matrix inverses

Let's go ahead and give away the surprise up front: matrix division is a valid operation, and multiplying a matrix by its inverse will result in the identity matrix.

So how do we find the inverse of a matrix? Well, now that we know how to find the determinant of a matrix, the formula for the inverse matrix will actually be something we already know how to calculate. Given matrix M as,

$$M = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

its **inverse** is given by

$$M^{-1} = \frac{1}{|M|} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

$$= \frac{1}{\begin{vmatrix} a & b \\ c & d \end{vmatrix}} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

$$= \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

Notice that the formula for the inverse matrix is a fraction with a numerator of 1 and the determinant as the denominator, multiplied by another matrix.

The other matrix is called the **adjugate** of M, and the adjugate is the matrix in which the values a and d have been swapped, and the values b and c have been multiplied by -1.

Example

Find the inverse of matrix K, then find $K \cdot K^{-1}$ and $K^{-1} \cdot K$ to show that we found the correct inverse.

$$K = \begin{bmatrix} -2 & 4 \\ 3 & 0 \end{bmatrix}$$

To find the inverse of matrix K, we plug into the formula for the inverse matrix.

$$K^{-1} = \frac{1}{|K|} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

$$K^{-1} = \frac{1}{\begin{vmatrix} -2 & 4 \\ 3 & 0 \end{vmatrix}} \begin{bmatrix} 0 & -4 \\ -3 & -2 \end{bmatrix}$$

$$K^{-1} = \frac{1}{-2(0) - 4(3)} \begin{bmatrix} 0 & -4 \\ -3 & -2 \end{bmatrix}$$

$$K^{-1} = -\frac{1}{12} \begin{bmatrix} 0 & -4 \\ -3 & -2 \end{bmatrix}$$

$$K^{-1} = \begin{bmatrix} -\frac{0}{12} & \frac{4}{12} \\ \frac{3}{12} & \frac{2}{12} \end{bmatrix}$$

$$K^{-1} = \begin{bmatrix} 0 & \frac{1}{3} \\ \frac{1}{4} & \frac{1}{6} \end{bmatrix}$$

This is the inverse of K, but we can prove it to ourselves by multiplying K by its inverse. If we've done our math right, we should get the identity matrix when we multiply them.

$$\begin{bmatrix} -2 & 4 \\ 3 & 0 \end{bmatrix} \begin{bmatrix} 0 & \frac{1}{3} \\ \frac{1}{4} & \frac{1}{6} \end{bmatrix} = \begin{bmatrix} -2(0) + 4\left(\frac{1}{4}\right) & -2\left(\frac{1}{3}\right) + 4\left(\frac{1}{6}\right) \\ 3(0) + 0\left(\frac{1}{4}\right) & 3\left(\frac{1}{3}\right) + 0\left(\frac{1}{6}\right) \end{bmatrix}$$

$$\begin{bmatrix} -2 & 4 \\ 3 & 0 \end{bmatrix} \begin{bmatrix} 0 & \frac{1}{3} \\ \frac{1}{4} & \frac{1}{6} \end{bmatrix} = \begin{bmatrix} 0 + 1 & -\frac{2}{3} + \frac{4}{6} \\ 0 + 0 & 1 + 0 \end{bmatrix}$$

$$\begin{bmatrix} -2 & 4 \\ 3 & 0 \end{bmatrix} \begin{bmatrix} 0 & \frac{1}{3} \\ \frac{1}{4} & \frac{1}{6} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$KK^{-1} = I_2$$

When we multiplied K by its inverse, we get the identity matrix. We also want to make the point that we can multiply in the other direction, and we still get the identity matrix.

$$\begin{bmatrix} 0 & \frac{1}{3} \\ \frac{1}{4} & \frac{1}{6} \end{bmatrix} \begin{bmatrix} -2 & 4 \\ 3 & 0 \end{bmatrix} = \begin{bmatrix} 0(-2) + \frac{1}{3}(3) & 0(4) + \frac{1}{3}(0) \\ \frac{1}{4}(-2) + \frac{1}{6}(3) & \frac{1}{4}(4) + \frac{1}{6}(0) \end{bmatrix}$$

$$\begin{bmatrix} 0 & \frac{1}{3} \\ \frac{1}{4} & \frac{1}{6} \end{bmatrix} \begin{bmatrix} -2 & 4 \\ 3 & 0 \end{bmatrix} = \begin{bmatrix} 0+1 & 0+0 \\ -\frac{1}{2}+\frac{1}{2} & 1+0 \end{bmatrix}$$

$$\begin{bmatrix} 0 & \frac{1}{3} \\ \frac{1}{4} & \frac{1}{6} \end{bmatrix} \begin{bmatrix} -2 & 4 \\ 3 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$K^{-1}K = I_2$$

This example shows how to use the formula to find the inverse matrix, and proves that multiplying by the inverse matrix is commutative. Whether we calculate KK^{-1} or $K^{-1}K$, we get back to the identity matrix either way.

Invertible and singular matrices

Not every matrix has an inverse. Given the formula for the inverse matrix,

$$K^{-1} = \frac{1}{|K|} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

we can probably spot right away that, because a fraction is undefined when its denominator is 0, we have to say that $|K| \neq 0$. In other words, if the determinant is 0, then the inverse matrix is undefined. To be specific, if

$$K = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

then the inverse of K is undefined when $ad - bc = 0$, or when $ad = bc$. If we divide both sides by b and d, we get

$$\frac{a}{b} = \frac{c}{d}$$

So if the ratio of a to b (the values in the first row of matrix K) is equivalent to the ratio of c to d (the values in the second row of matrix K), then we know right away that the matrix K doesn't have a defined inverse. If the matrix doesn't have an inverse, we call it a **singular matrix**. When the matrix does have an inverse, we say that the matrix is **invertible**.

Let's do an example where we determine whether the matrix is singular or invertible.

Example

Say whether each matrix is invertible or singular.

$$M = \begin{bmatrix} 1 & -3 \\ 3 & 5 \end{bmatrix}$$

$$L = \begin{bmatrix} -6 & 2 \\ 3 & -1 \end{bmatrix}$$

For each matrix, we'll look at whether or not the ratio of a to b is equal to the ratio of c to d. For matrix M, we get

$$\frac{1}{-3} = -\frac{1}{3} \quad \neq \quad \frac{3}{5}$$

Because these aren't equivalent ratios, matrix M is invertible, which means it has a defined inverse. For matrix L, we get

$$\frac{-6}{2} = -3 \quad = \quad \frac{3}{-1} = -3$$

Because these are equivalent ratios, matrix L is not invertible, which means it doesn't have a defined inverse, and therefore that it's a singular matrix.

Solving systems with inverse matrices

We learned in Algebra how to solve systems of linear equations using substitution, elimination, and graphing, and we learned earlier how to apply Gauss-Jordan elimination to a matrix to solve a system.

Now we want to talk about how to solve systems using inverse matrices. To walk through this, let's use a simple system.

$$3x - 4y = 6$$

$$-2x + y = -9$$

And for the sake of illustration, let's compare this to the generic system

$$ax + by = f$$

$$cx + dy = g$$

We can always represent a linear system like this as the coefficient matrix, multiplied by the column vector (x, y), set equal to the column vector (f, g).

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} f \\ g \end{bmatrix}$$

Or, if we call the coefficient matrix M, rename the column vector (x, y) as $\vec{a}$, and rename the column vector (f, g) as $\vec{b}$, then we can say that this equation is equivalent to

$$M\vec{a} = \vec{b}$$

If we multiply both sides by the inverse of M, M^{-1}, we get

$$M^{-1}M\vec{a} = M^{-1}\vec{b}$$

$$I\vec{a} = M^{-1}\vec{b}$$

$$\vec{a} = M^{-1}\vec{b}$$

The reason we went through these steps to multiply by the inverse is because we now have an equation that's solved for $\vec{a}$, which remember is representing the column vector (x, y), which really is just the solution to the system. In other words, this equation is saying that we can find the solution to the system simply by multiplying the inverse matrix M^{-1} by the column vector (f, g)!

Furthermore, as long as we keep M, and therefore M^{-1} the same, we can substitute any values that we'd like for f and g (the constants on the right side of the system of equations), and $\vec{a} = M^{-1}\vec{b}$ will give us an immediate solution set for (x, y).

Right now that doesn't necessarily feel super useful, but as we use matrices in more advanced ways, it'll become extremely valuable to be able to change the values that make up $\vec{b}$, and immediately get the solution set (x, y) that comes from $\vec{a}$.

Let's do an example where we use the inverse matrix to solve the system.

Example

Use an inverse matrix to solve the system.

$$3x - 4y = 6$$

$$-2x + y = -9$$

Start by transferring the system into a matrix equation.

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} f \\ g \end{bmatrix}$$

$$\begin{bmatrix} 3 & -4 \\ -2 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 6 \\ -9 \end{bmatrix}$$

Find the inverse of the coefficient matrix.

$$M = \begin{bmatrix} 3 & -4 \\ -2 & 1 \end{bmatrix}$$

$$M^{-1} = \frac{1}{3(1) - (-2)(-4)} \begin{bmatrix} 1 & 4 \\ 2 & 3 \end{bmatrix}$$

$$M^{-1} = -\frac{1}{5} \begin{bmatrix} 1 & 4 \\ 2 & 3 \end{bmatrix}$$

$$M^{-1} = \begin{bmatrix} -\frac{1}{5} & -\frac{4}{5} \\ -\frac{2}{5} & -\frac{3}{5} \end{bmatrix}$$

Then we can say that the solution to the system is

$$\overrightarrow{a} = M^{-1}\overrightarrow{b}$$

$$\overrightarrow{a} = \begin{bmatrix} -\frac{1}{5} & -\frac{4}{5} \\ -\frac{2}{5} & -\frac{3}{5} \end{bmatrix} \begin{bmatrix} 6 \\ -9 \end{bmatrix}$$

$$\vec{a} = \begin{bmatrix} -\frac{1}{5}(6) - \frac{4}{5}(-9) \\ -\frac{2}{5}(6) - \frac{3}{5}(-9) \end{bmatrix}$$

$$\vec{a} = \begin{bmatrix} -\frac{6}{5} + \frac{36}{5} \\ -\frac{12}{5} + \frac{27}{5} \end{bmatrix}$$

$$\vec{a} = \begin{bmatrix} \frac{30}{5} \\ \frac{15}{5} \end{bmatrix}$$

$$\vec{a} = \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 6 \\ 3 \end{bmatrix}$$

Using this process with the inverse matrix, we conclude that $x = 6$ and $y = 3$.

Representing the system graphically

We can also express the same system from the example,

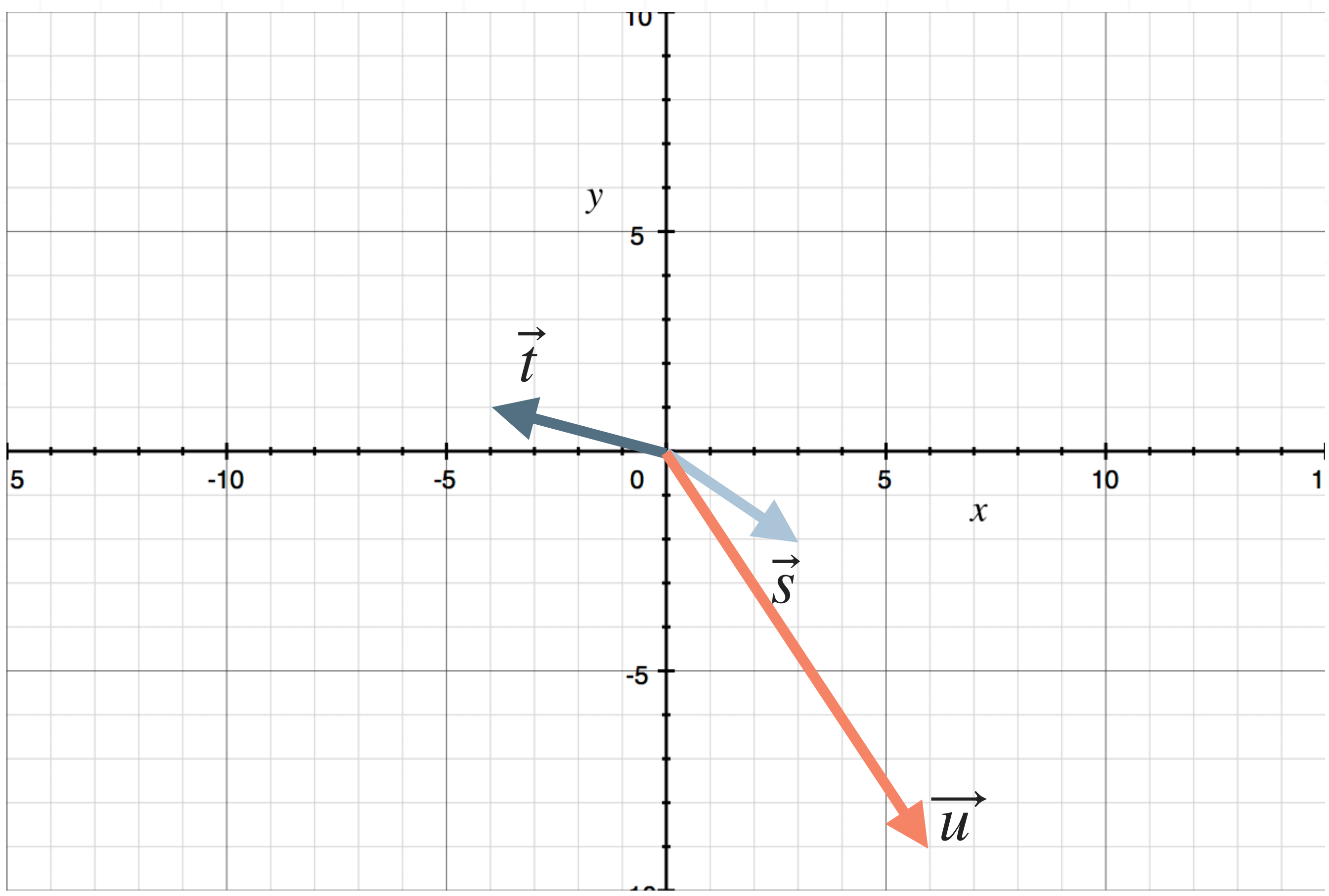
$$3x - 4y = 6$$

$$-2x + y = -9$$

as this matrix equation:

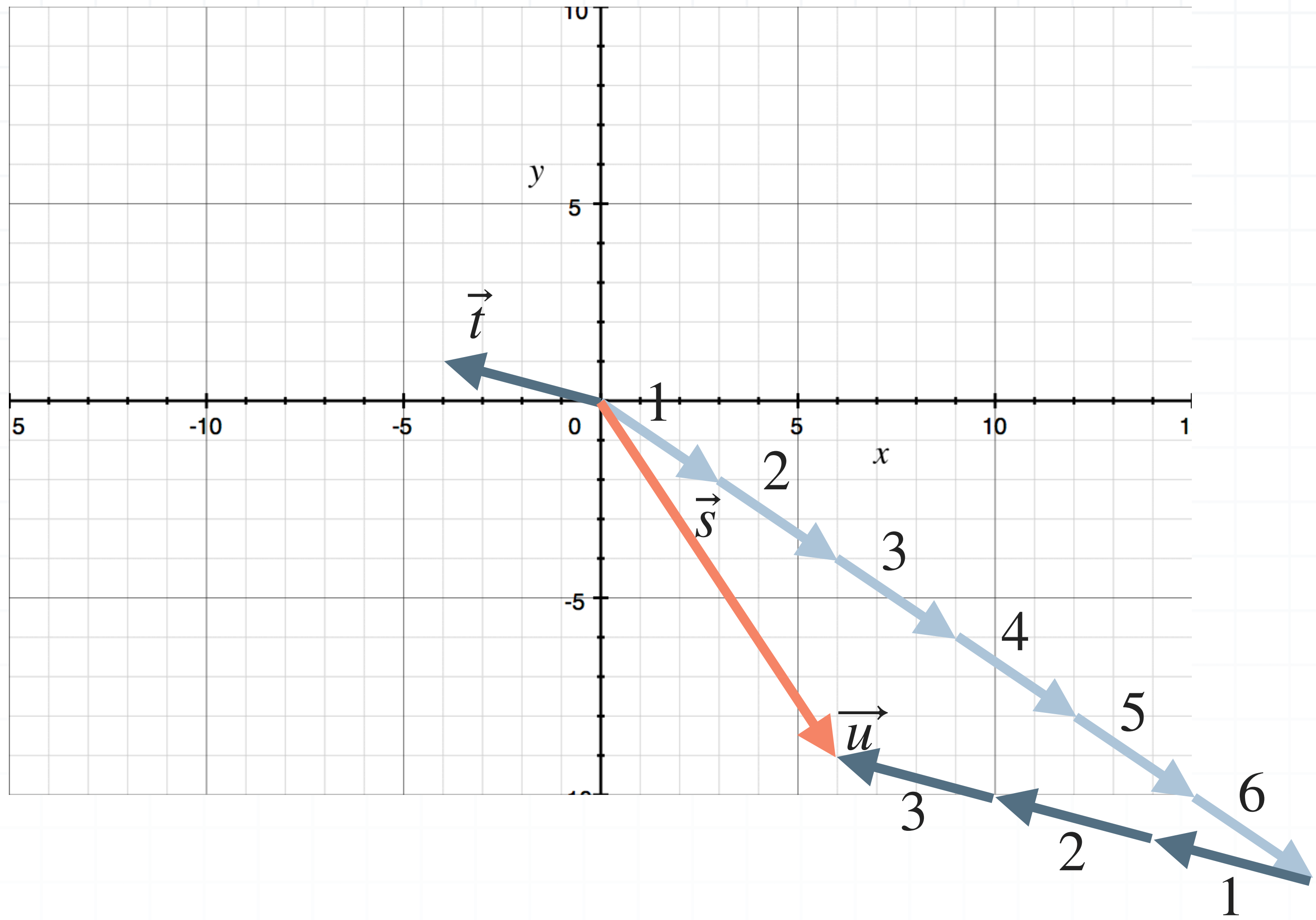
$$\begin{bmatrix} 3 \\ -2 \end{bmatrix} x + \begin{bmatrix} -4 \\ 1 \end{bmatrix} y = \begin{bmatrix} 6 \\ -9 \end{bmatrix}$$

Graphically, we can think about this equation as the vector $\vec{s} = (3, -2)$ for x and the vector $\vec{t} = (-4,1)$ for y, and the resulting vector $\vec{u} = (6, -9)$. If we put these vectors in the same coordinate plane, we get



We already know from the example that $x = 6$ and $y = 3$. What that solution tells us graphically is that we need to put 6 of the x vectors and 3 of the y vectors together, and we'll end up at the same spot as the resulting vector $\vec{u} = (6, -9)$.

If we sketch out that linear combination, we see that the solution set we found earlier does actually satisfy the system.


10
y
5
$\vec{t}$
1
5
-10
-5
0
5
10
x
2
$\vec{s}$
3
-5
4
5
$\vec{u}$
6
3
2
1

Solving systems with Cramer's Rule

Cramer's Rule is a simple rule that lets us use determinants to solve a system of equations. It tells us that we can solve for any variable in the system by calculating

$$\frac{D_v}{D}$$

where D_v is the determinant of the coefficient matrix, with the answer column values substituted into the column representing the variable for which we're trying to solve, and where D is the determinant of the coefficient matrix.

Which means that, if we want to find the value of x, we need to find D_x/D, and if we want to find the value of y, we need to find D_y/D. All that sounds tricky, but let's look at an example to break it down.

Example

Solve for x in the system.

$$2x - 3y = 5$$

$$3x + 12y = -8$$

Because we're looking for the value of x, we want to find D_x/D. We need to start with the coefficient matrix for the system,

$$\begin{bmatrix} 2 & -3 \\ 3 & 12 \end{bmatrix}$$

The answer column matrix is the matrix of constants from the right side of the system,

$$\begin{bmatrix} 5 \\ -8 \end{bmatrix}$$

D_x is the determinant of the coefficient matrix with the answer column matrix substituted into the x-column,

$$D_x = \begin{vmatrix} 5 & -3 \\ -8 & 12 \end{vmatrix}$$

D is the determinant of the coefficient matrix,

$$D = \begin{vmatrix} 2 & -3 \\ 3 & 12 \end{vmatrix}$$

Putting these values together, Cramer's Rule tells us that the value of x in the system is

$$x = \frac{D_x}{D} = \frac{\begin{vmatrix} 5 & -3 \\ -8 & 12 \end{vmatrix}}{\begin{vmatrix} 2 & -3 \\ 3 & 12 \end{vmatrix}}$$

$$x = \frac{(5)(12) - (-3)(-8)}{(2)(12) - (-3)(3)}$$

$$x = \frac{36}{33} = \frac{12}{11}$$

Let's do another example where we use Cramer's Rule to solve for y.

Example

Use Cramer's Rule to solve for the value of y that satisfies the system.

$$3x - 2y = 7$$

$$5x - 8y = 21$$

The coefficient matrix is

$$\begin{bmatrix} 3 & -2 \\ 5 & -8 \end{bmatrix}$$

The answer column matrix is

$$\begin{bmatrix} 7 \\ 21 \end{bmatrix}$$

Then D_y is what we get when we substitute the answer column matrix into the second column of the coefficient matrix, and then take the determinant of the result.

$$D_y = \begin{vmatrix} 3 & 7 \\ 5 & 21 \end{vmatrix}$$

The determinant of the coefficient matrix is

$$D = \begin{vmatrix} 3 & -2 \\ 5 & -8 \end{vmatrix}$$

Putting these values together, Cramer's Rule tells us that the value of y in the system is

$$y = \frac{D_y}{D} = \frac{\begin{vmatrix} 3 & 7 \\ 5 & 21 \end{vmatrix}}{\begin{vmatrix} 3 & -2 \\ 5 & -8 \end{vmatrix}}$$

$$y = \frac{(3)(21) - (7)(5)}{(3)(-8) - (-2)(5)}$$

$$y = \frac{28}{-14} = -2$$

Fraction decomposition

In this section, we want to learn all about **partial fractions decomposition**, which is a tool we use to break down and rewrite rational functions.

Remember that **rational functions** are fractions in which both the numerator and denominator are polynomial functions. For instance, the function $f(x)$ is a rational function,

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

and we can use partial fractions decomposition to rewrite $f(x)$ as

$$f(x) = \frac{5}{x-2} - \frac{3}{x-1}$$

We can work backwards to prove that this decomposition is equivalent to the original function by finding a common denominator,

$$f(x) = \frac{5(x-1)}{(x-2)(x-1)} - \frac{3(x-2)}{(x-1)(x-2)}$$

combining the fractions,

$$f(x) = \frac{5(x-1)-3(x-2)}{(x-1)(x-2)}$$

and then simplifying.

$$f(x) = \frac{5x-5-3x+6}{x^2-2x-x+2}$$

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

Let's do another example where we work backwards from the decomposition to find the original ration function.

Example

Rewrite the function as one fraction.

$$f(x) = -\frac{6}{x+3} - \frac{1}{x-2}$$

We can work backwards to find the original function associated with this decomposition by finding a common denominator,

$$f(x) = -\frac{6(x-2)}{(x+3)(x-2)} - \frac{1(x+3)}{(x-2)(x+3)}$$

combining the fractions,

$$f(x) = \frac{-6(x-2) - (x+3)}{(x-2)(x+3)}$$

and then simplifying.

$$f(x) = \frac{-6x + 12 - x - 3}{x^2 + 3x - 2x - 6}$$

$$f(x) = \frac{9 - 7x}{x^2 + x - 6}$$

As we can see, it's fairly straightforward to get from the decomposition back to the original function. But we don't know yet how to get from the original function to the decomposition. That's where partial fractions decomposition comes in.

And we'll use these decompositions throughout Calculus and more advanced math courses, so it's a technique we want to become familiar with.

Factors and the decomposition

Our first step in any decomposition is to factor the denominator of the rational function as completely as possible. In other words, we want to make sure that none of the factors in the denominator can be broken down any further.

Once the denominator is factored, we need to identify the combination of factors we're dealing with. There are four types of factors we can find: distinct linear factors, repeated linear factors, distinct quadratic factors, and repeated quadratic factors.

Function	**Factored form**	**Factor type**
$f(x)=\dfrac{2x+1}{x^2-3x+2}$	$f(x)=\dfrac{2x+1}{(x-1)(x-2)}$	Distinct linear
$f(x)=\dfrac{3x-2}{x^2+2x+1}$	$f(x)=\dfrac{3x-2}{(x+1)(x+1)}$	Repeated linear

$$f(x) = \frac{x+4}{x^4+4x^2+3} \qquad f(x) = \frac{x+4}{(x^2+3)(x^2+1)}$$ Distinct quadratic

$$f(x) = \frac{x^3-2x^2}{x^4+2x^2+1} \qquad f(x) = \frac{x^3-2x^2}{(x^2+1)(x^2+1)}$$ Repeated quadratic

Linear factors are first degree factors, while **quadratic factors** are second degree factors. In other words, we see x^1 terms in linear factors, and x^2 terms in quadratic factors. We call factors **distinct** when they're different from one another, and we call factors **repeated** when they're the same.

So $(x-1)(x-2)$ is a set of distinct linear factors because each factor is linear, and $(x-1)$ is different than $(x-2)$. And $(x^2+1)(x^2+1)$ is a set of repeated quadratic factors because each factor is quadratic, and (x^2+1) is the same as (x^2+1).

For linear factors in our decomposition, we'll use a single constant A for the numerator. But for quadratic factors in our decomposition, we'll use a linear factor $Ax+B$ for the numerator, where A and B are constants.

Factored form	**Decomposition**
$f(x) = \frac{2x+1}{(x-1)(x-2)}$	$\frac{A}{x-1} + \frac{B}{x-2}$
$f(x) = \frac{x+4}{(x^2+3)(x^2+1)}$	$\frac{Ax+B}{x^2+3} + \frac{Cx+D}{x^2+1}$

Notice how we separated each factor into its own fraction, and only included each factor one time. For each linear factor in the decomposition, we'll need to choose a new constant, A, B, C, etc. And for each quadratic factor in the decomposition, we'll need to choose new constants, $Ax+B$,

$Cx+D$, $Ex+F$, etc. The numerator of the original rational function is irrelevant to the decomposition; it won't affect any part of the decomposition.

Let's do an example where we build the decomposition of a rational function.

Example

Build the decomposition of each rational function.

$$f(x)=\frac{3x}{(1-x)(x^2+2x-8)}$$

$$g(x)=\frac{x^2+6x+1}{(x^2+2)(x^2+3)(x^2+5)}$$

The denominator of the function $f(x)$ can be factored further, because x^2+2x-8 can be factored as $(x+4)(x-2)$.

$$f(x)=\frac{3x}{(1-x)(x+4)(x-2)}$$

Now the denominator of $f(x)$ can't be factored any further, and it's the product of three distinct linear factors.

Therefore, the decomposition of $f(x)$ should include three fractions. The denominator of each of the three fractions will be one of the factors from the original denominator, and the numerator of each of the three fractions will be a distinct constant, A, B, C, etc. So the decomposition will be

$$f(x) = \frac{A}{1-x} + \frac{B}{x+4} + \frac{C}{x-2}$$

The denominator of the function $g(x)$ is the product of three distinct quadratic factors.

Therefore, the decomposition of $g(x)$ should include three fractions. The denominator of each of the three fractions will be one of the factors from the original denominator, and the numerator of each of the three fractions will be a distinct expression, $Ax+B$, $Cx+D$, $Ex+F$, etc. So the decomposition will be

$$g(x) = \frac{Ax+B}{x^2+2} + \frac{Cx+D}{x^2+3} + \frac{Ex+F}{x^2+5}$$

Distinct linear factors

Now that we understand the general idea behind partial fractions decomposition, and we've seen how to set up a decomposition equation, let's look in more detail at how to use partial fractions decomposition when the denominator of our rational function is the product of distinct linear factors.

Distinct linear factors

Remember that factors are **distinct** when they're different. So $(x-1)(x-2)$ represents a set of distinct factors, but $(x-1)(x-1)$ is a non-distinct factor set. When a factor set is non-distinct, we call the factors **repeated**, but we'll save repeated factors for later.

Factors are linear when they contain first degree x terms. In other words, a **linear factor** only contains x variables that are raised to the first power, so we only see x^1, or just x. Otherwise, if a factor contains an x^2 term, we call it a **quadratic factor**, but we'll save quadratic factors for later.

The factor x by itself is a linear factor, so something like $x(x-1)(x-2)$ is a set of distinct linear factors.

If we see a constant coefficient in front of our factor set, we can pull that out of the fraction. For instance, given the function

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

we can pull the 2 coefficient out of the fraction to rewrite the function as

$$f(x) = \frac{1}{2}\left(\frac{2x+1}{x(x-1)(x-2)}\right)$$

Then we can use a partial fractions decomposition just on what we have left inside the parentheses. We just can't forget about the $1/2$ when we give our final answer.

Setting up the decomposition

Once we've factored the denominator of the rational function as completely as possible, and determined that the denominator is the product of distinct linear factors, we can set up the decomposition equation.

The left side of the decomposition equation will always be the factored form of the original rational function. The right side of the decomposition equation will be a sum of separate fractions, where the denominator of each fraction is one of the distinct linear factors, and the numerator is a unique constant, A, B, C, etc.

For instance, we would set up the decomposition of

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

as

$$\frac{2x+1}{x(x-1)(x-2)} = \frac{A}{x} + \frac{B}{x-1} + \frac{C}{x-2}$$

On the right side of the equation, we've separated each distinct linear factor into its own fraction. While the distinct linear factors serve as the denominator of each fraction, we insert distinct unknowns A, B, C, etc. into the numerator of each fraction.

Let's do another example where we set up the decomposition equation.

Example

Set up the partial fractions decomposition equation for the function.

$$f(x) = \frac{6x - 5}{2x^3 + 5x^2 - 12x}$$

We have to start by factoring the denominator. We'll pull out a common factor of x first,

$$f(x) = \frac{6x - 5}{x(2x^2 + 5x - 12)}$$

then we'll factor the remaining trinomial.

$$f(x) = \frac{6x - 5}{x(2x - 3)(x + 4)}$$

The denominator is the product of three distinct linear factors, so we'll break the denominator into the sum of three fractions, each one with one of the factors as its denominator.

$$f(x) = \frac{}{x} + \frac{}{2x - 3} + \frac{}{x + 4}$$

Because we have linear factors, we need a single unknown constant in each numerator,

$$f(x) = \frac{A}{x} + \frac{B}{2x-3} + \frac{C}{x+4}$$

then we set this decomposition equal to the factored form of the original function to get the partial fractions decomposition equation.

$$\frac{6x-5}{x(2x-3)(x+4)} = \frac{A}{x} + \frac{B}{2x-3} + \frac{C}{x+4}$$

How to solve for the constants

Once we've set up the partial fractions decomposition equation, all that's left to do is solve for the constants we added in, A, B, C, etc.

When we're dealing with only distinct linear factors, we'll use the same method to solve for each constant.

1. Set each distinct linear factor equal to 0, and solve the equation for x.

2. If we're trying to solve for A, remove the distinct linear factor associated with A from the factored form of the original function.

3. Evaluate the new function at the value of x that we solved for in Step 1.

Let's do an example with distinct linear factors so that we can see these steps in action.

Example

Rewrite the function as its partial fractions decomposition.

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

These are distinct linear factors, so we'll set up the decomposition as

$$\frac{2x+1}{(x-1)(x-2)} = \frac{A}{x-1} + \frac{B}{x-2}$$

To solve for A, we'll set $x - 1 = 0$ to find $x = 1$. We'll remove the $x - 1$ factor from the original factored function,

$$\frac{2x+1}{x-2}$$

then we'll evaluate this modified function at the $x = 1$ value we just found.

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

To solve for B, we'll set $x - 2 = 0$ to find $x = 2$. We'll remove the $x - 2$ factor from the original factored function,

$$\frac{2x+1}{x-1}$$

then we'll evaluate this modified function at the $x = 2$ value we just found.

$$\frac{2x+1}{x-1} \to \frac{2(2)+1}{2-1} \to \frac{5}{1} \to 5$$

Plugging $A = -3$ and $B = 5$ back into the partial fractions decomposition gives

$$f(x) = \frac{A}{x-1} + \frac{B}{x-2}$$

$$f(x) = -\frac{3}{x-1} + \frac{5}{x-2}$$

$$f(x) = \frac{5}{x-2} - \frac{3}{x-1}$$

This function is equivalent to the original rational function we started with.

Repeated linear factors

Let's look in more detail at how to use partial fractions decomposition when the denominator of our rational function is the product of repeated linear factors.

Repeated linear factors

Remember that factors are **repeated** when they're equivalent. So $(x-1)(x-1)$ and $(x+2)(x+2)(x+2)$ are sets of repeated factors. We have to realize that x^3 is also a set of repeated factors, because x^3 is a factor of x repeated three times, $x \cdot x \cdot x$. The factor set $(x+1)^4$ is also repeated, because we can rewrite it as $(x+1)(x+1)(x+1)(x+1)$.

Factors are linear when they contain first degree x terms. In other words, a **linear factor** only contains x variables that are raised to the first power, so we only see x^1, or just x. Otherwise, if a factor contains an x^2 term, we call it a **quadratic factor**, but we'll save quadratic factors for later.

Setting up the decomposition

Once we've factored the denominator of the rational function as completely as possible, and determined that the denominator is the product of repeated linear factors, we can set up the decomposition equation.

The left side of the decomposition equation will always be the factored form of the original rational function. The right side of the decomposition equation will be a sum of separate fractions, where the denominators are the descending powers of the repeated linear factors, and the numerators are unique constants, A, B, C, etc.

For instance, we would set up the decomposition of

$$f(x) = \frac{x-3}{(x-1)(x-1)(x-1)}$$

as

$$\frac{x-3}{(x-1)(x-1)(x-1)} = \frac{A}{x-1} + \frac{B}{(x-1)^2} + \frac{C}{(x-1)^3}$$

In other words, the original denominator was

$$(x-1)(x-1)(x-1)$$

$$(x-1)^3$$

Because we have a linear factor repeated three times, we need to include fractions in our decomposition for the third power, second power, and first power of that linear factor. If the original denominator had been $(x-1)^6$ power, we would have needed to include fractions in our decomposition for the sixth, fifth, fourth, third, second, and first powers,

$$\frac{x-3}{(x-1)^6} = \frac{A}{x-1} + \frac{B}{(x-1)^2} + \frac{C}{(x-1)^3} + \frac{D}{(x-1)^4} + \frac{E}{(x-1)^5} + \frac{F}{(x-1)^6}$$

Let's do another example where we set up the decomposition equation.

Example

Set up the partial fractions decomposition equation for the function.

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

The denominator is the product of one distinct linear factor, repeated four times, so we'll break the denominator into the sum of four fractions, one for each of the fourth, third, second, and first powers of the distinct linear factor.

$$f(x) = \frac{\quad}{x + 2} + \frac{\quad}{(x + 2)^2} + \frac{\quad}{(x + 2)^3} + \frac{\quad}{(x + 2)^4}$$

Because we have linear factors, we need a single unknown constant in each numerator,

$$f(x) = \frac{A}{x + 2} + \frac{B}{(x + 2)^2} + \frac{C}{(x + 2)^3} + \frac{D}{(x + 2)^4}$$

then we set this decomposition equal to the factored form of the original function to get the partial fractions decomposition equation.

$$\frac{x^2 + x + 1}{(x + 2)^4} = \frac{A}{x + 2} + \frac{B}{(x + 2)^2} + \frac{C}{(x + 2)^3} + \frac{D}{(x + 2)^4}$$

How to solve for the constants

Once we've set up the partial fractions decomposition equation, all that's left to do is solve for the constants we added in, A, B, C, etc.

When we're dealing with only repeated linear factors, we'll use the following steps to solve for each constant.

1. Combine the fractions on the right side of the equation by finding a common denominator equivalent to the function's original denominator.

2. Once the denominators on both sides are equivalent, set the numerators equal to each other.

3. Evaluate the equation at $x = k$, where k is the value that makes the repeated factor equal to 0. This will solve for the constant associated with the first-power factor.

4. Simplify the remaining equation and equate coefficients to build a system of equations that can be solved for the other constants.

Let's do an example with repeated linear factors so that we can see these steps in action.

Example

Rewrite the function as its partial fractions decomposition.

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

The denominator of this original rational function is $(x+1)^2$. Because the linear factor $x+1$ is raised to the power of 2, we'll need to include the second and first powers of the factor in the decomposition.

$$\frac{3x-2}{(x+1)^2}=\frac{A}{x+1}+\frac{B}{(x+1)^2}$$

Now that we have the decomposition equation, we'll combine the fractions on the right side by finding a common denominator.

$$\frac{3x-2}{(x+1)^2}=\frac{A(x+1)}{(x+1)(x+1)}+\frac{B}{(x+1)^2}$$

$$\frac{3x-2}{(x+1)^2}=\frac{A(x+1)}{(x+1)^2}+\frac{B}{(x+1)^2}$$

$$\frac{3x-2}{(x+1)^2}=\frac{A(x+1)+B}{(x+1)^2}$$

Because the denominators are equivalent, that means the numerators must also be equivalent, so we'll set the numerators equal to one another.

$$3x-2=A(x+1)+B$$

The linear factor $x+1$ is equal to 0 when $x=-1$, so we'll evaluate this equation at $x=-1$ in order to solve for B.

$$3(-1)-2=A(-1+1)+B$$

$$-3-2=B$$

$$B=-5$$

We'll plug this back into the numerator equation and then solve for A.

$$3x - 2 = A(x + 1) - 5$$

$$3x + 3 = A(x + 1)$$

$$3(x + 1) = A(x + 1)$$

$$3 = A$$

Plugging $A = 3$ and $B = -5$ back into the partial fractions decomposition gives

$$f(x) = \frac{3}{x + 1} - \frac{5}{(x + 1)^2}$$

This function is equivalent to the original rational function we started with.

Let's do one more example with a higher degree denominator.

Example

Rewrite the function as its partial fractions decomposition.

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

The denominator of this original rational function is $(x+2)^4$. Because the linear factor $x+2$ is raised to the power of 4, we'll need to include the fourth, third, second, and first powers of the factor in the decomposition.

$$\frac{x^2+x+1}{(x+2)^4}=\frac{A}{x+2}+\frac{B}{(x+2)^2}+\frac{C}{(x+2)^3}+\frac{D}{(x+2)^4}$$

Now that we have the decomposition equation, we'll combine the fractions on the right side by finding a common denominator.

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

$$+\frac{C(x+2)}{(x+2)^3(x+2)}+\frac{D}{(x+2)^4}$$

$$\frac{x^2+x+1}{(x+2)^4}=\frac{A(x+2)^3}{(x+2)^4}+\frac{B(x+2)^2}{(x+2)^4}+\frac{C(x+2)}{(x+2)^4}+\frac{D}{(x+2)^4}$$

$$\frac{x^2+x+1}{(x+2)^4}=\frac{A(x+2)^3+B(x+2)^2+C(x+2)+D}{(x+2)^4}$$

Because the denominators are equivalent, that means the numerators must also be equivalent, so we'll set the numerators equal to one another.

$$x^2+x+1=A(x+2)^3+B(x+2)^2+C(x+2)+D$$

The linear factor $x+2$ is equal to 0 when $x=-2$, so we'll evaluate this equation at $x=-2$ in order to solve for D.

$$(-2)^2-2+1=A(-2+2)^3+B(-2+2)^2+C(-2+2)+D$$

$$4-2+1=D$$

$$D = 3$$

We'll plug this back into the numerator equation,

$$x^2 + x + 1 = A(x + 2)^3 + B(x + 2)^2 + C(x + 2) + 3$$

and then simplify the right side, collecting like terms.

$$x^2 + x + 1 = A(x^3 + 6x^2 + 12x + 8) + B(x^2 + 4x + 4) + C(x + 2) + 3$$

$$x^2 + x + 1 = Ax^3 + 6Ax^2 + 12Ax + 8A + Bx^2 + 4Bx + 4B + Cx + 2C + 3$$

$$x^2 + x + 1 = Ax^3 + (6Ax^2 + Bx^2) + (12Ax + 4Bx + Cx) + (8A + 4B + 2C + 3)$$

$$x^2 + x + 1 = Ax^3 + (6A + B)x^2 + (12A + 4B + C)x + (8A + 4B + 2C + 3)$$

The coefficients on x^3 are 0 and A, the coefficients on x^2 are 1 and $6A + B$, the coefficients on x are 1 and $12A + 4B + C$, and the constants are 1 and $8A + 4B + 2C + 3$, so we get the system of equations

$$A = 0$$

$$6A + B = 1$$

$$12A + 4B + C = 1$$

$$8A + 4B + 2C + 3 = 1$$

or $A = 0$ with

$$B = 1$$

$$4B + C = 1$$

$$4B + 2C = -2$$

or $A = 0$ and $B = 1$ with

$$4 + C = 1$$

$$4 + 2C = -2$$

or $A = 0$, $B = 1$, $C = -3$, and $D = 3$. Plugging these values back into the partial fractions decomposition gives

$$f(x) = \frac{0}{x+2} + \frac{1}{(x+2)^2} + \frac{-3}{(x+2)^3} + \frac{3}{(x+2)^4}$$

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

This function is equivalent to the original rational function we started with.

Distinct quadratic factors

Let's look in more detail at how to use partial fractions decomposition when the denominator of our rational function is the product of distinct quadratic factors.

Distinct quadratic factors

Remember that factors are distinct when they're different. So $(x-1)(x-2)$ represents a set of distinct factors, but $(x-1)(x-1)$ is a non-distinct factor set.

Factors are linear when they contain first degree x terms, but they're quadratic when they contain a second degree x term. So $(x-1)(x-1)$ is a set of linear factors, while $(x^2+1)(x^2+3)$ is a set of quadratic factors.

Setting up the decomposition

Once we've factored the denominator of the rational function as completely as possible, and determined that the denominator is the product of distinct quadratic factors, we can set up the decomposition equation.

The left side of the decomposition equation will always be the factored form of the original rational function. The right side of the decomposition equation will be a sum of separate fractions, where the denominator of each fraction is one of the distinct quadratic factors, and the numerator is a unique expression in the form $Ax+B$, $Cx+D$, $Ex+F$, etc.

For instance, we'd set up the decomposition of

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

as

$$\frac{x^2 + 2x - 3}{(x^2 + 1)(x^2 + 3)} = \frac{Ax + B}{x^2 + 1} + \frac{Cx + D}{x^2 + 3}$$

How to solve for the constants

Once we've set up the partial fractions decomposition equation, all that's left to do is solve for the constants we added in, A, B, C, etc.

When we're dealing with only distinct quadratic factors, we'll use the same method to solve for each constant.

1. Combine the fractions on the right side of the equation by finding a common denominator equivalent to the function's original denominator.

2. Once the denominators on both sides are equivalent, set the numerators equal to each other.

3. Simplify the equation and equate coefficients to build a system of equations that can be solved for the constants.

Let's do an example with distinct quadratic factors so that we can see these steps in action.

Example

Rewrite the function as its partial fractions decomposition.

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

These are distinct quadratic factors, so we'll set up the decomposition as

$$\frac{x+4}{(x^2+3)(x^2+1)} = \frac{Ax+B}{x^2+3} + \frac{Cx+D}{x^2+1}$$

We'll combine the fractions on the right side of the equation by finding a common denominator.

$$\frac{x+4}{(x^2+3)(x^2+1)} = \frac{(Ax+B)(x^2+1)}{(x^2+3)(x^2+1)} + \frac{(Cx+D)(x^2+3)}{(x^2+3)(x^2+1)}$$

$$\frac{x+4}{(x^2+3)(x^2+1)} = \frac{(Ax+B)(x^2+1)+(Cx+D)(x^2+3)}{(x^2+3)(x^2+1)}$$

Once the denominators on both sides are equivalent, we'll set the numerators equal to each other.

$$x+4 = (Ax+B)(x^2+1)+(Cx+D)(x^2+3)$$

Multiply out the right side of this numerator equation,

$$x+4 = Ax^3+Ax+Bx^2+B+Cx^3+3Cx+Dx^2+3D$$

then group like terms and factor.

$$x+4=(A+C)x^3+(B+D)x^2+(A+3C)x+(B+3D)$$

If we think about this equation as

$$0x^3+0x^2+1x+4=(A+C)x^3+(B+D)x^2+(A+3C)x+(B+3D)$$

then we can equate coefficients to build a system of equations.

$A+C=0$ $\quad\quad$ $B+D=0$

$A+3C=1$ $\quad\quad$ $B+3D=4$

Solving the two equations on the left as a system gives $A=-1/2$ and $C=1/2$, while solving the two equations on the right as a system gives $B=-2$ and $D=2$. Plugging these back into the partial fractions decomposition gives

$$f(x)=\frac{Ax+B}{x^2+3}+\frac{Cx+D}{x^2+1}$$

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

$$f(x)=-\frac{1}{2}\left(\frac{x+4}{x^2+3}\right)+\frac{1}{2}\left(\frac{x+4}{x^2+1}\right)$$

$$f(x)=\frac{1}{2}\left(\frac{x+4}{x^2+1}\right)-\frac{1}{2}\left(\frac{x+4}{x^2+3}\right)$$

Repeated quadratic factors

Let's look in more detail at how to use partial fractions decomposition when the denominator of our rational function is the product of repeated quadratic factors.

Repeated quadratic factors

Remember that factors are repeated when they're equivalent. So $(x^2+1)(x^2+1)$ and $(2x^2+5)(2x^2+5)(2x^2+5)$ are sets of repeated factors. The factor set $(3+2x^2)^4$ is also repeated, because we can rewrite it as $(3+2x^2)(3+2x^2)(3+2x^2)(3+2x^2)$.

Factors are linear when they contain first degree x terms, but they're quadratic when they contain a second degree x term. So $(x-1)(x-1)$ is a set of linear factors, while $(x^2+1)(x^2+3)$ is a set of quadratic factors.

Setting up the decomposition

Once we've factored the denominator of the rational function as completely as possible, and determined that the denominator is the product of repeated quadratic factors, we can set up the decomposition equation.

The left side of the decomposition equation will always be the factored form of the original rational function. The right side of the decomposition equation will be a sum of separate fractions, where the denominators are the descending powers of the repeated quadratic factors, and the numerators are expressions in the form $Ax+B$, $Cx+D$, $Ex+F$, etc.

For instance, we'd set up the decomposition of

$$f(x)=\frac{2x-3}{(2x^2+5)(2x^2+5)(2x^2+5)}$$

as

$$\frac{2x-3}{(2x^2+5)(2x^2+5)(2x^2+5)}=\frac{Ax+B}{2x^2+5}+\frac{Cx+D}{(2x^2+5)^2}+\frac{Ex+F}{(2x^2+5)^3}$$

In other words, the original denominator was

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

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

Because we have a quadratic factor repeated three times, we need to include fractions in our decomposition for the third power, second power, and first power of that quadratic factor. If the original denominator had been $(2x^2+5)^5$ power, we would have needed to include fractions in our decomposition for the fifth, fourth, third, second, and first powers,

$$\frac{2x-3}{(2x^2+5)^5}=\frac{Ax+B}{2x^2+5}+\frac{Cx+D}{(2x^2+5)^2}+\frac{Ex+F}{(2x^2+5)^3}+\frac{Gx+H}{(2x^2+5)^4}+\frac{Ix+J}{(2x^2+5)^5}$$

How to solve for the constants

Once we've set up the partial fractions decomposition equation, all that's left to do is solve for the constants we added in, A, B, C, etc.

When we're dealing with only repeated quadratic factors, we'll use the same method to solve for each constant.

1. Combine the fractions on the right side of the equation by finding a common denominator equivalent to the function's original denominator.

2. Once the denominators on both sides are equivalent, set the numerators equal to each other.

3. Simplify the equation and equate coefficients to build a system of equations that can be solved for the constants.

Let's do an example with repeated quadratic factors so that we can see these steps in action.

Example

Rewrite the function as its partial fractions decomposition.

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

These are repeated quadratic factors, so we'll set up the decomposition as

$$\frac{x^3 - 2x^2}{(x^2+1)(x^2+1)} = \frac{Ax+B}{x^2+1} + \frac{Cx+D}{(x^2+1)^2}$$

We'll combine the fractions on the right side of the equation by finding a common denominator equivalent to the function's original denominator.

$$\frac{x^3-2x^2}{(x^2+1)(x^2+1)}=\frac{(Ax+B)(x^2+1)}{(x^2+1)(x^2+1)}+\frac{Cx+D}{(x^2+1)^2}$$

$$\frac{x^3-2x^2}{(x^2+1)(x^2+1)}=\frac{(Ax+B)(x^2+1)}{(x^2+1)^2}+\frac{Cx+D}{(x^2+1)^2}$$

$$\frac{x^3-2x^2}{(x^2+1)(x^2+1)}=\frac{(Ax+B)(x^2+1)+Cx+D}{(x^2+1)^2}$$

Now that the denominators are equivalent, we can set the numerators equal to one another.

$$x^3-2x^2=(Ax+B)(x^2+1)+Cx+D$$

Expand the right side of the equation, then collect like terms and factor.

$$x^3-2x^2=Ax^3+Ax+Bx^2+B+Cx+D$$

$$x^3-2x^2=Ax^3+Bx^2+(A+C)x+(B+D)$$

If we equate coefficients from the left and right sides, we get a system of equations.

$$A=1$$

$$B=-2$$

$$A+C=0$$

$$B+D=0$$

Solving this system gives $A=1$, $B=-2$, $C=-1$, and $D=2$. Plugging these values back into the partial fractions decomposition gives

$$f(x) = \frac{Ax + B}{x^2 + 1} + \frac{Cx + D}{(x^2 + 1)^2}$$

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

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

Mixed factors

We've looked at how to build the partial fractions decomposition when we have the product of distinct linear factors, repeated linear factors, distinct quadratic factors, and repeated quadratic factors.

Up to now, we've kept each of these four factor types separated from each other, but it's actually really common to find these factors mixed together as part of the same denominator of the rational function.

For instance, we might see a function with one distinct linear factor and one distinct quadratic factor,

$$f(x)=\frac{x^2-2x-5}{(x-1)(x^2+9)}$$

or we might see a function with a distinct linear factor, a set of repeated linear factors, and a set of repeated quadratic factors,

$$f(x)=\frac{3x^2+2}{x(x-2)^2(x^2+3)^3}$$

These are just examples, but the important thing to remember is that we can have combination of any of the four factor types.

Setting up the decomposition

When we have mixed factors, our process for setting up the decomposition will still be the same as the process we've used previously with each factor type individually.

- Any distinct linear factors each get their own fraction, and we'll pair each one with a unique constant, A, B, C, etc.

- Any repeated linear factor set needs to be broken into a set of fractions, each with a different power of the linear factor, from the first power of the factor up to the power of the original factor set, and we'll pair each one with a unique constant, A, B, C, etc.

- Any distinct quadratic factors each get their own fraction, and we'll pair each one with a unique expression, $Ax+B$, $Cx+D$, $Ex+F$, etc.

- Any repeated quadratic factor set needs to be broken into a set of fractions, each with a different power of the quadratic factor, from the first power of the factor up to the power of the original factor set, and we'll pair each one with a unique expression, $Ax+B$, $Cx+D$, $Ex+F$, etc.

Let's do an example where we set up the decomposition equation.

Example

Set up the decomposition equation for the rational function.

$$f(x)=\frac{2x-3}{x(x+2)(x-3)^3(2x^2+5)^2}$$

The denominator of the function is already factored as much as it can be. We can identify that x and $x+2$ are distinct linear factors, so we'll start setting up the decomposition as

$$\frac{A}{x}+\frac{B}{x+2}$$

The part of the decomposition for the repeated linear factor set $(x-3)^3$, because we have three factors of $x-3$, needs to include the third, second, and first powers of that factor. Adding that to our decomposition gives

$$\frac{A}{x}+\frac{B}{x+2}+\frac{C}{x-3}+\frac{D}{(x-3)^2}+\frac{E}{(x-3)^3}$$

The part of the decomposition for the repeated quadratic factor set $(2x^2+5)^2$, because we have two factors of $2x^2+5$, needs to include the second and first powers of that factor. Adding that to our decomposition gives

$$\frac{A}{x}+\frac{B}{x+2}+\frac{C}{x-3}+\frac{D}{(x-3)^2}+\frac{E}{(x-3)^3}+\frac{Fx+G}{2x^2+5}+\frac{Hx+I}{(2x^2+5)^2}$$

So the decomposition equation for this function is

$$\frac{2x-3}{x(x+2)(x-3)^3(2x^2+5)^2}$$

$$=\frac{A}{x}+\frac{B}{x+2}+\frac{C}{x-3}+\frac{D}{(x-3)^2}+\frac{E}{(x-3)^3}+\frac{Fx+G}{2x^2+5}+\frac{Hx+I}{(2x^2+5)^2}$$

How to solve for the constants

Once we've set up the partial fractions decomposition equation, all that's left to do is solve for the constants we added in, A, B, C, etc.

When we're dealing with mixed factors, we'll start by solving for the constants associated with the distinct linear factors. For all the other constants, we'll follow these steps:

1. Combine the fractions on the right side of the equation by finding a common denominator equivalent to the function's original denominator.

2. Once the denominators on both sides are equivalent, set the numerators equal to each other.

3. Simplify the equation and equate coefficients to build a system of equations that can be solved for the constants.

Let's do an example with mixed factors so that we can see these steps in action.

Example

Rewrite the function as its partial fractions decomposition.

$$f(x) = \frac{4x^4 - 23x^3 + 36x^2 - 48x + 18}{x(x-3)^2(x^2+2)}$$

The denominator is already factored as far as it can be, and we have a distinct linear factor, a set of repeated linear factors, and a distinct quadratic factor. So we'll set up the partial fractions decomposition equation as

$$\frac{4x^4-23x^3+36x^2-48x+18}{x(x-3)^2(x^2+2)}=\frac{A}{x}+\frac{B}{x-3}+\frac{C}{(x-3)^2}+\frac{Dx+E}{x^2+2}$$

To solve for A, we'll remove the factor x from the left side, then evaluate what remains on the left side at $x=0$.

$$\frac{4x^4-23x^3+36x^2-48x+18}{(x-3)^2(x^2+2)}$$

$$\to\frac{4(0)^4-23(0)^3+36(0)^2-48(0)+18}{(0-3)^2(0^2+2)}\to\frac{18}{9(2)}\to 1=A$$

To solve for the other constants, we'll find a common denominator on the right side, then combine the fractions.

$$\frac{4x^4-23x^3+36x^2-48x+18}{x(x-3)^2(x^2+2)}=\frac{1}{x}+\frac{B}{x-3}+\frac{C}{(x-3)^2}+\frac{Dx+E}{x^2+2}$$

$$\frac{4x^4-23x^3+36x^2-48x+18}{x(x-3)^2(x^2+2)}=\frac{(x-3)^2(x^2+2)}{x(x-3)^2(x^2+2)}+\frac{Bx(x-3)(x^2+2)}{x(x-3)^2(x^2+2)}$$

$$+\frac{Cx(x^2+2)}{x(x-3)^2(x^2+2)}+\frac{(Dx+E)x(x-3)^2}{x(x-3)^2(x^2+2)}$$

$$\frac{4x^4-23x^3+36x^2-48x+18}{x(x-3)^2(x^2+2)}$$

$$= \frac{(x-3)^2(x^2+2) + Bx(x-3)(x^2+2) + Cx(x^2+2) + (Dx+E)x(x-3)^2}{x(x-3)^2(x^2+2)}$$

Because the denominators are equivalent, we can set the numerators equal to one another.

$$4x^4 - 23x^3 + 36x^2 - 48x + 18$$

$$= (x-3)^2(x^2+2) + Bx(x-3)(x^2+2) + Cx(x^2+2) + (Dx+E)x(x-3)^2$$

$$4x^4 - 23x^3 + 36x^2 - 48x + 18$$

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

$$+Cx^3 + 2Cx + (Dx^2 + Ex)(x^2 - 6x + 9)$$

$$4x^4 - 23x^3 + 36x^2 - 48x + 18$$

$$= x^4 + 2x^2 - 6x^3 - 12x + 9x^2 + 18 + Bx^4 + 2Bx^2 - 3Bx^3 - 6Bx$$

$$+Cx^3 + 2Cx + Dx^4 - 6Dx^3 + 9Dx^2 + Ex^3 - 6Ex^2 + 9Ex$$

$$4x^4 - 23x^3 + 36x^2 - 48x + 18$$

$$= (1 + B + D)x^4 + (-6 - 3B + C - 6D + E)x^3$$

$$+(2 + 9 + 2B + 9D - 6E)x^2 + (-12 - 6B + 2C + 9E)x + 18$$

We can equate coefficients on either side of the equation to build a system of equations.

$$1 + B + D = 4$$

$$-6 - 3B + C - 6D + E = -23$$

$$2 + 9 + 2B + 9D - 6E = 36$$

$$-12 - 6B + 2C + 9E = -48$$

$$18 = 18$$

Simplifying the first four equations of the system gives

$$B + D = 3$$

$$3B - C + 6D - E = 17$$

$$2B + 9D - 6E = 25$$

$$6B - 2C - 9E = 36$$

We'll use a matrix to solve for the values of B, C, D, and E.

$$\begin{bmatrix} 1 & 0 & 1 & 0 & | & 3 \\ 3 & -1 & 6 & -1 & | & 17 \\ 2 & 0 & 9 & -6 & | & 25 \\ 6 & -2 & 0 & -9 & | & 36 \end{bmatrix} \to \begin{bmatrix} 1 & 0 & 1 & 0 & | & 3 \\ 0 & -1 & 3 & -1 & | & 8 \\ 2 & 0 & 9 & -6 & | & 25 \\ 6 & -2 & 0 & -9 & | & 36 \end{bmatrix}$$

$$\to \begin{bmatrix} 1 & 0 & 1 & 0 & | & 3 \\ 0 & -1 & 3 & -1 & | & 8 \\ 0 & 0 & 7 & -6 & | & 19 \\ 6 & -2 & 0 & -9 & | & 36 \end{bmatrix} \to \begin{bmatrix} 1 & 0 & 1 & 0 & | & 3 \\ 0 & -1 & 3 & -1 & | & 8 \\ 0 & 0 & 7 & -6 & | & 19 \\ 0 & -2 & -6 & -9 & | & 18 \end{bmatrix}$$

$$\to \begin{bmatrix} 1 & 0 & 1 & 0 & | & 3 \\ 0 & 1 & -3 & 1 & | & -8 \\ 0 & 0 & 7 & -6 & | & 19 \\ 0 & -2 & -6 & -9 & | & 18 \end{bmatrix} \to \begin{bmatrix} 1 & 0 & 1 & 0 & | & 3 \\ 0 & 1 & -3 & 1 & | & -8 \\ 0 & 0 & 7 & -6 & | & 19 \\ 0 & 0 & -12 & -7 & | & 2 \end{bmatrix}$$

$$\to \begin{bmatrix} 1 & 0 & 1 & 0 & | & 3 \\ 0 & 1 & -3 & 1 & | & -8 \\ 0 & 0 & 1 & -\frac{6}{7} & | & \frac{19}{7} \\ 0 & 0 & -12 & -7 & | & 2 \end{bmatrix} \to \begin{bmatrix} 1 & 0 & 0 & \frac{6}{7} & | & \frac{2}{7} \\ 0 & 1 & -3 & 1 & | & -8 \\ 0 & 0 & 1 & -\frac{6}{7} & | & \frac{19}{7} \\ 0 & 0 & -12 & -7 & | & 2 \end{bmatrix}$$

$$\to \begin{bmatrix} 1 & 0 & 0 & \frac{6}{7} & | & \frac{2}{7} \\ 0 & 1 & 0 & -\frac{11}{7} & | & \frac{1}{7} \\ 0 & 0 & 1 & -\frac{6}{7} & | & \frac{19}{7} \\ 0 & 0 & -12 & -7 & | & 2 \end{bmatrix} \to \begin{bmatrix} 1 & 0 & 0 & \frac{6}{7} & | & \frac{2}{7} \\ 0 & 1 & 0 & -\frac{11}{7} & | & \frac{1}{7} \\ 0 & 0 & 1 & -\frac{6}{7} & | & \frac{19}{7} \\ 0 & 0 & 0 & -\frac{121}{7} & | & \frac{242}{7} \end{bmatrix}$$

$$\to \begin{bmatrix} 1 & 0 & 0 & \frac{6}{7} & | & \frac{2}{7} \\ 0 & 1 & 0 & -\frac{11}{7} & | & \frac{1}{7} \\ 0 & 0 & 1 & -\frac{6}{7} & | & \frac{19}{7} \\ 0 & 0 & 0 & 1 & | & -2 \end{bmatrix} \to \begin{bmatrix} 1 & 0 & 0 & 0 & | & 2 \\ 0 & 1 & 0 & -\frac{11}{7} & | & \frac{1}{7} \\ 0 & 0 & 1 & -\frac{6}{7} & | & \frac{19}{7} \\ 0 & 0 & 0 & 1 & | & -2 \end{bmatrix}$$

$$\to \begin{bmatrix} 1 & 0 & 0 & 0 & | & 2 \\ 0 & 1 & 0 & 0 & | & -3 \\ 0 & 0 & 1 & -\frac{6}{7} & | & \frac{19}{7} \\ 0 & 0 & 0 & 1 & | & -2 \end{bmatrix} \to \begin{bmatrix} 1 & 0 & 0 & 0 & | & 2 \\ 0 & 1 & 0 & 0 & | & -3 \\ 0 & 0 & 1 & 0 & | & 1 \\ 0 & 0 & 0 & 1 & | & -2 \end{bmatrix}$$

Pairing $A = 1$ with the values from this matrix, $B = 2$, $C = -3$, $D = 1$, and $E = -2$, we can plug into the partial fractions decomposition equation.

$$f(x) = \frac{A}{x} + \frac{B}{x-3} + \frac{C}{(x-3)^2} + \frac{Dx+E}{x^2+2}$$

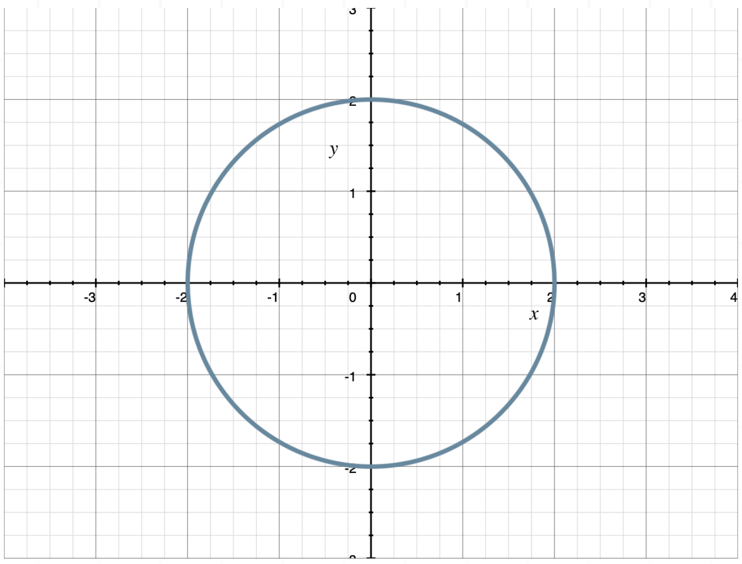
$$f(x) = \frac{1}{x} + \frac{2}{x-3} + \frac{-3}{(x-3)^2} + \frac{1x + (-2)}{x^2+2}$$

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

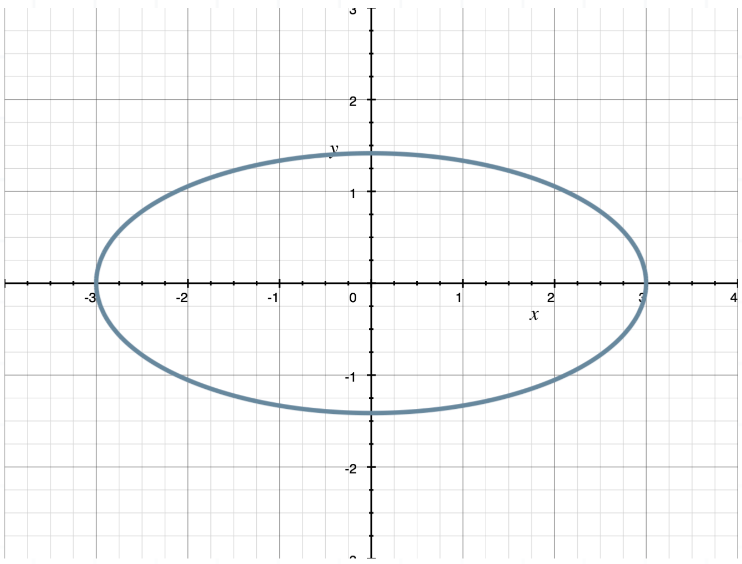
Identifying conic sections

Conic sections are circles, ellipses, parabolas, and hyperbolas.

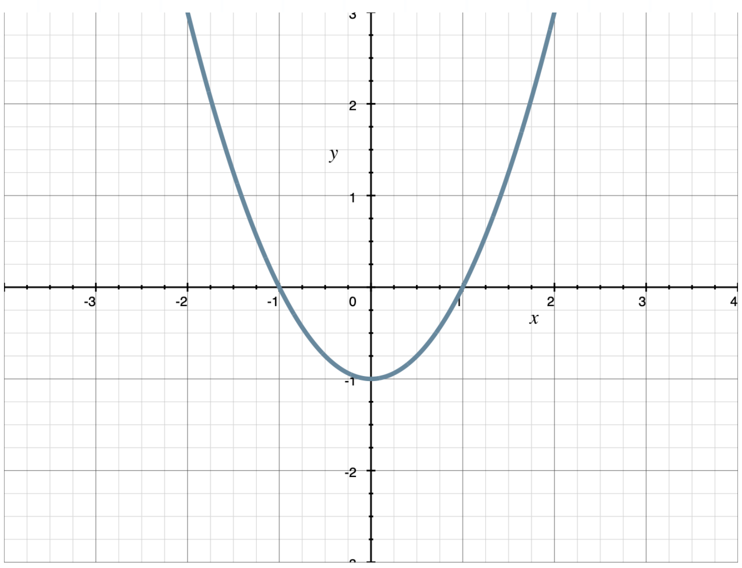
Example circle:



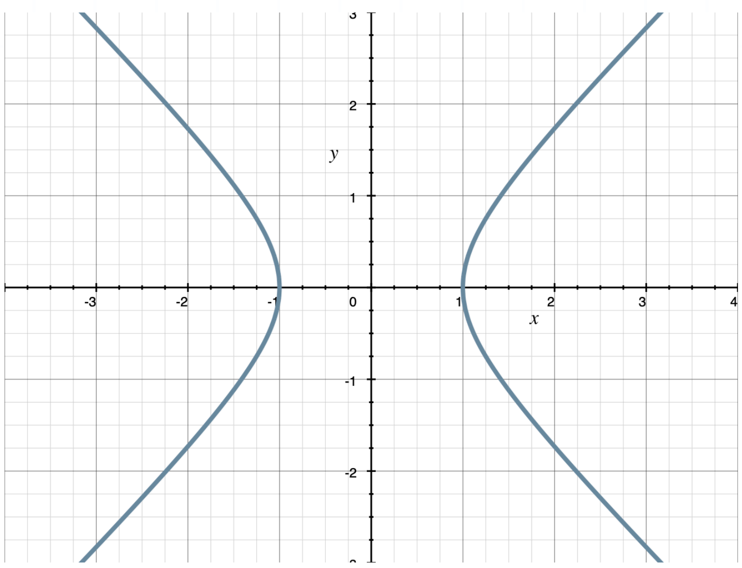
Example ellipse:



Example parabola:



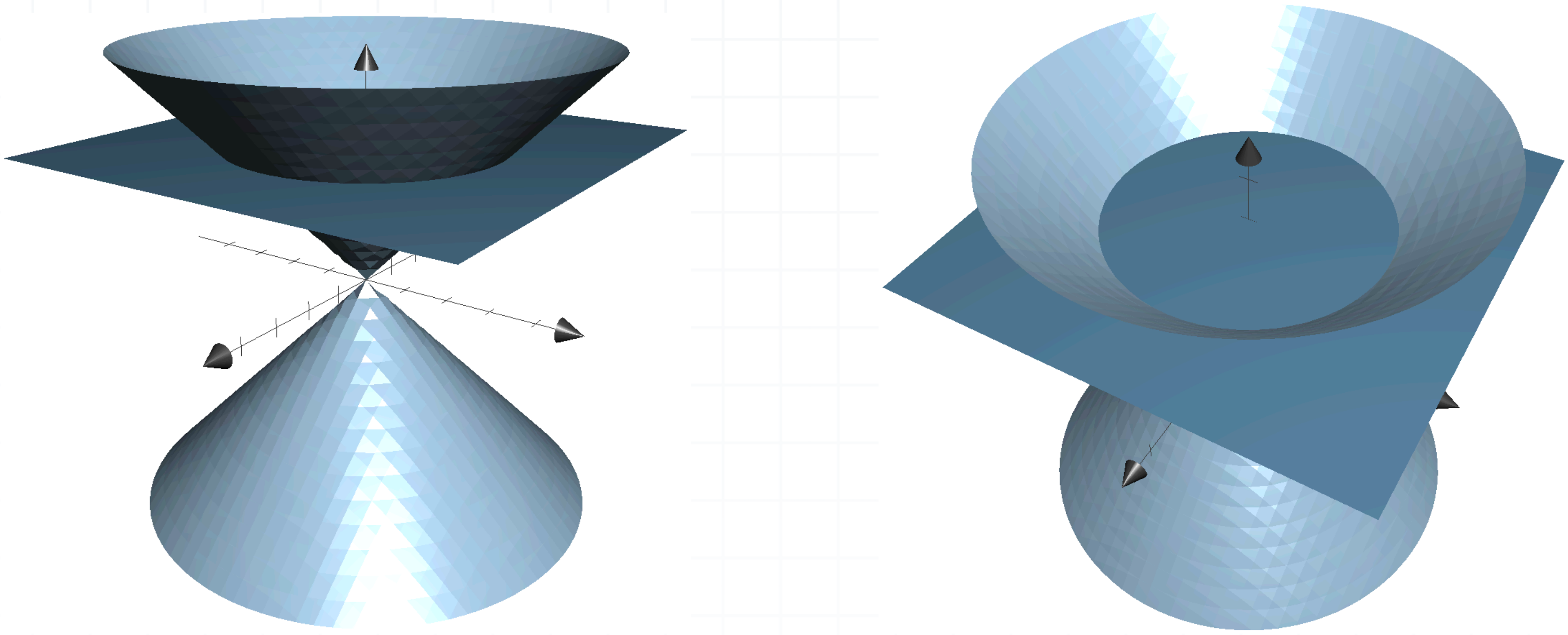
Example hyperbola:



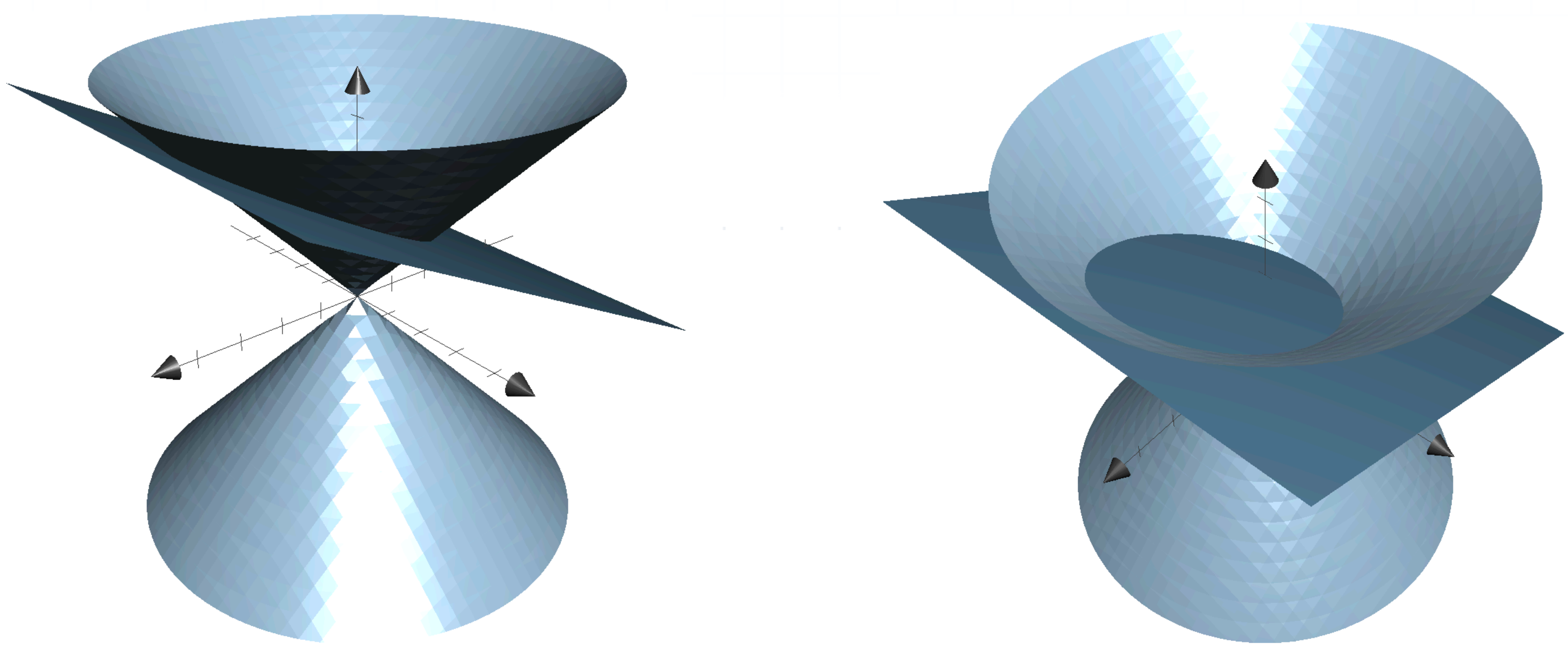
These shapes are called "conic sections" because they're created from the intersection of a plane and a cone.

Slicing the cone

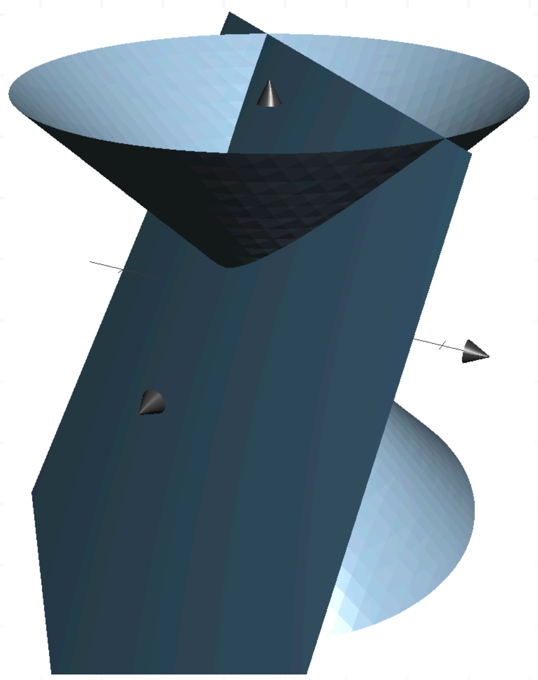
For instance, if we imagine slicing a cone with a perfectly horizontal plane, we can see how the intersection of the plane and the cone is a circle.

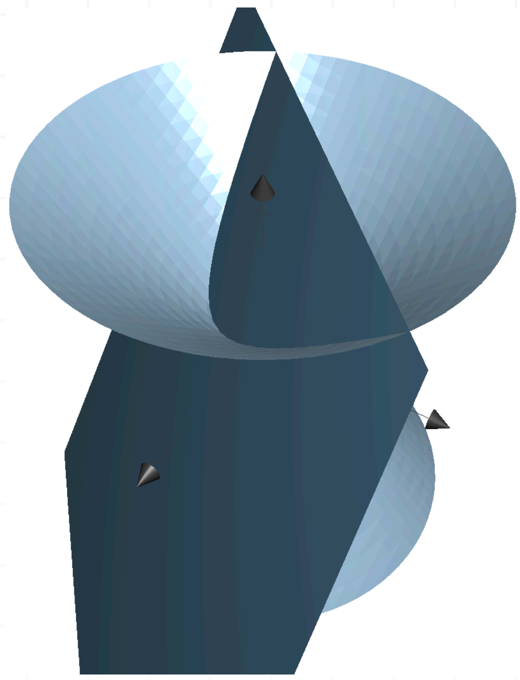


If we tilt that plane slightly, the circle starts to skew, and the intersection of the plane with the cone becomes an ellipse.

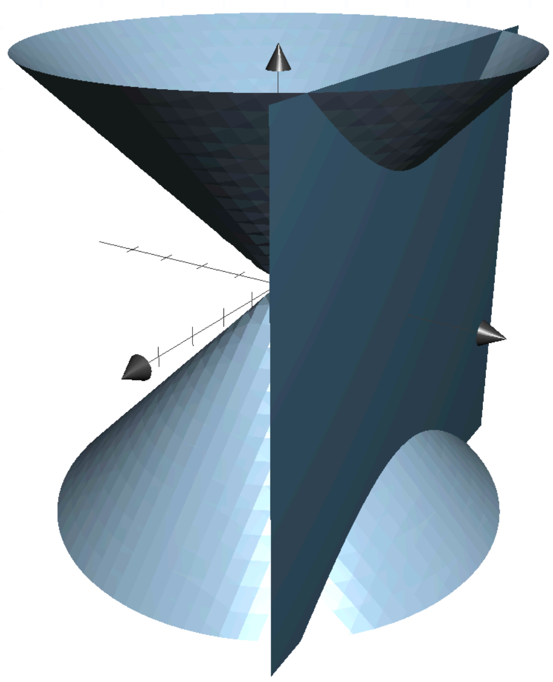


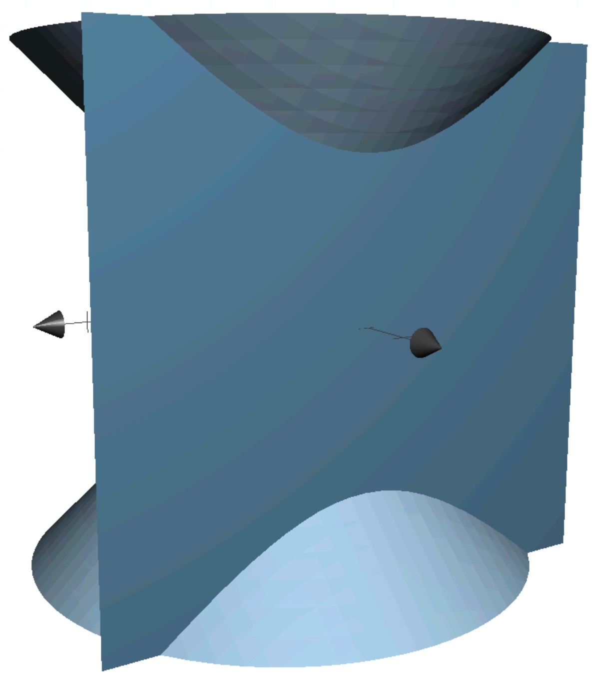
If we continue tilting the plane even more, eventually it'll intersect only one side of the cone, and the intersection of the plane and the cone will form a parabola.





Finally, if we tilt the plane so much that it intersects both the top and bottom parts of the cone, the intersection of the plane with the cone becomes a hyperbola.





Equations of conic sections

Conic sections are easy to spot from their equations. Remember that the equation of a line is an equation in which x and y are both linear variables (they're both raised to the power of 1). For instance, the line $y = x + 3$ has a y^1 term and an x^1 term.

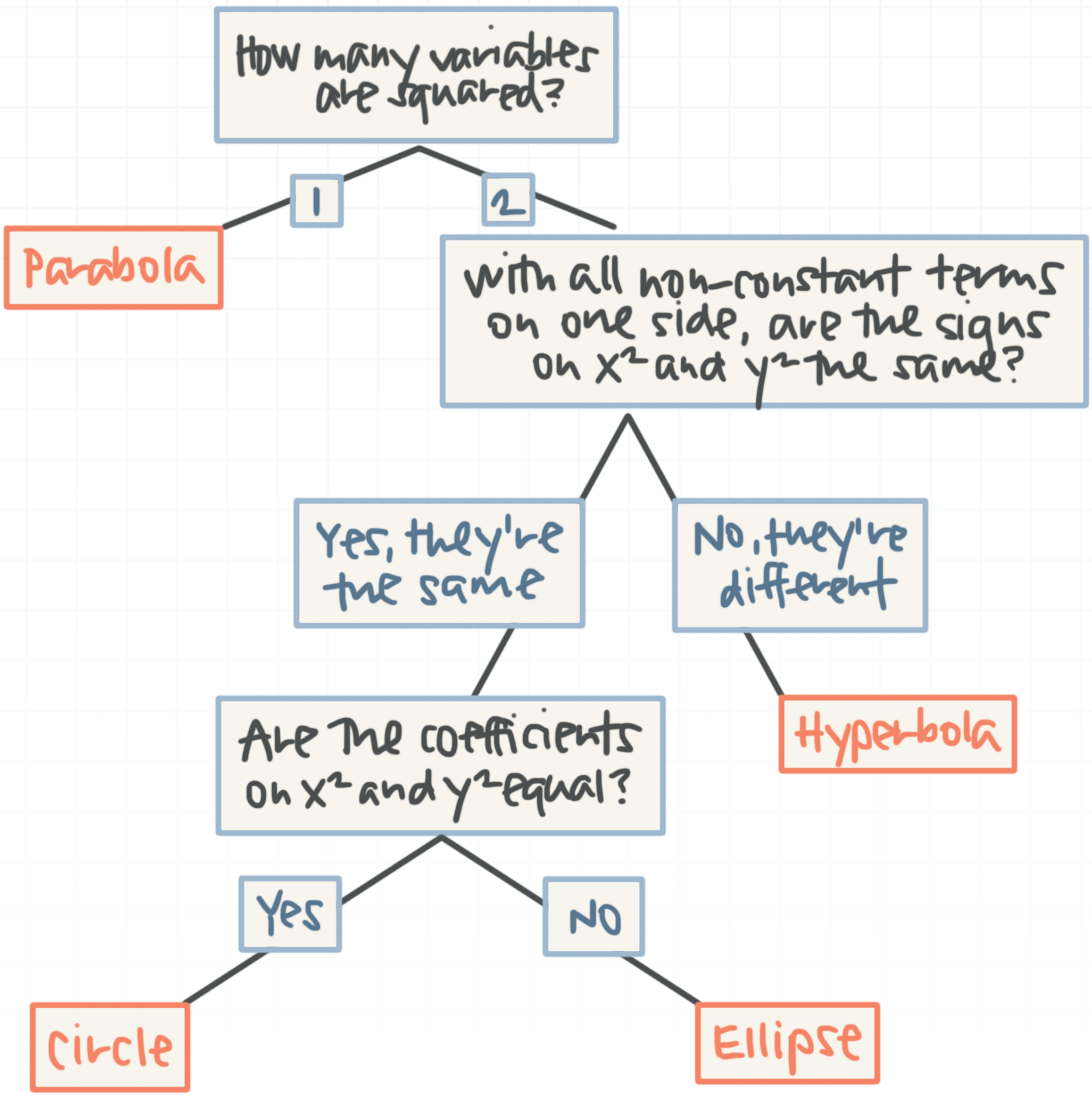
If we instead raise one of the variables to the power of 2, the equation becomes quadratic, and it's the equation of a **parabola**. So if the equation includes either y^2 and x, or x^2 and y, the equation represents a parabola.

If both variables are squared, such that we have x^2 and y^2 in the equation, then the equation represents a circle, an ellipse, or a hyperbola. To figure out which figure is represented, collect all the non-constant terms on one side, leaving the constant on the other side.

If the x^2 and y^2 terms have the same sign, then the equation represents a circle or an ellipse, but if they have a different sign, then the equation represents a **hyperbola**.

If the coefficients on the x^2 and y^2 terms are unequal, then the equation represents an **ellipse**, but if the coefficients on the x^2 and y^2 terms are equal, then the equation represents a **circle**.

Here's a flowchart that summarizes this information:



Let's do an example where we identify the type of conic represented by the equation.

Example

Identify each equation as a parabola, hyperbola, ellipse, or circle.

(a) $x^2 + y^2 - 2x - 4y = 0$

(b) $12x + 5y^2 + 28 = -2x^2 + 20y$

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

(d) $36y^2 - 9x^2 = 36$

Equation (a) has both an x^2 and y^2 term, so it's either a circle, an ellipse, or a hyperbola. Because both of those terms have the same sign (they're both positive) and they're on the same side of the equation, this is either an ellipse or a circle. Because the coefficients on those terms are equal (they're both 1), the equation represents a circle.

Equation (b) has both an x^2 and y^2 term, so it's either a circle, an ellipse, or a hyperbola. If we move everything but the constant to the same side of the equation,

$$12x + 5y^2 + 28 = -2x^2 + 20y$$

$$2x^2 + 12x + 5y^2 - 20y = -28$$

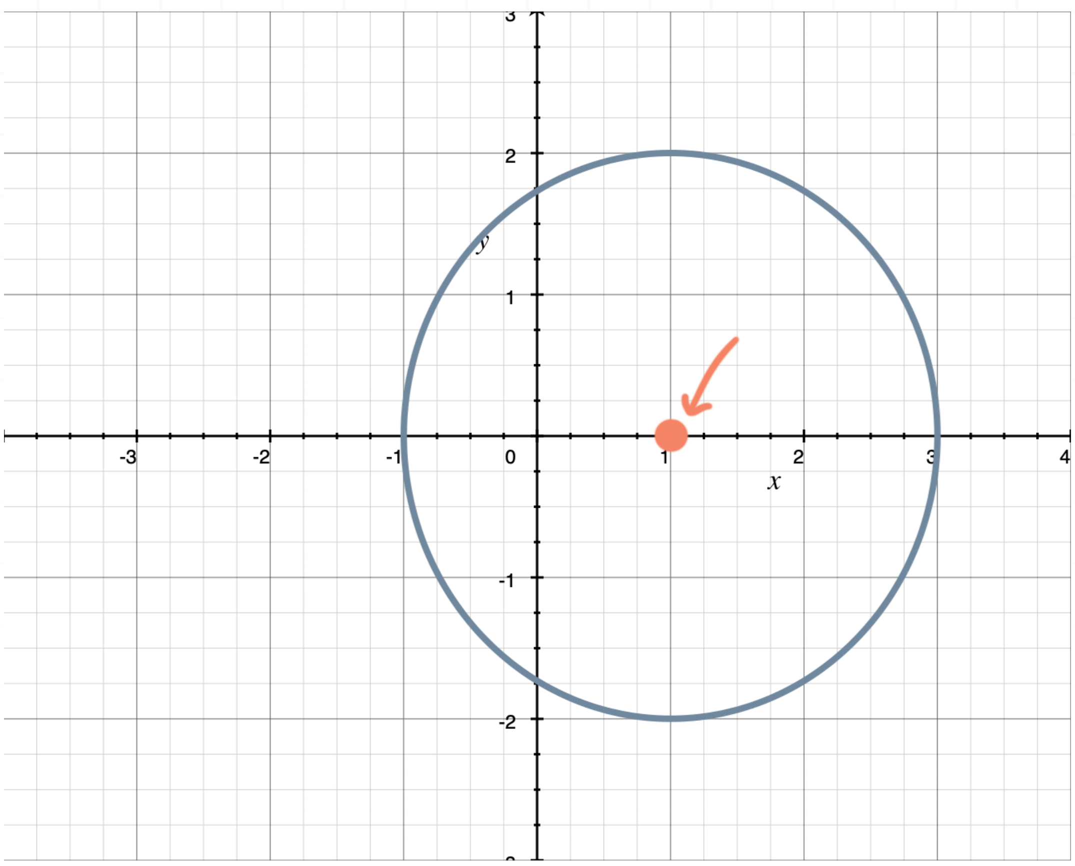
we see that both x^2 and y^2 have positive coefficients, so this is either an ellipse or a circle. Because the coefficients on those terms are unequal (they're 2 and 5), the equation represents an ellipse.

Equation (c) has an x^2 term and a y term. Because one variable is squared and the other has an exponent of 1, the equation represents a parabola.

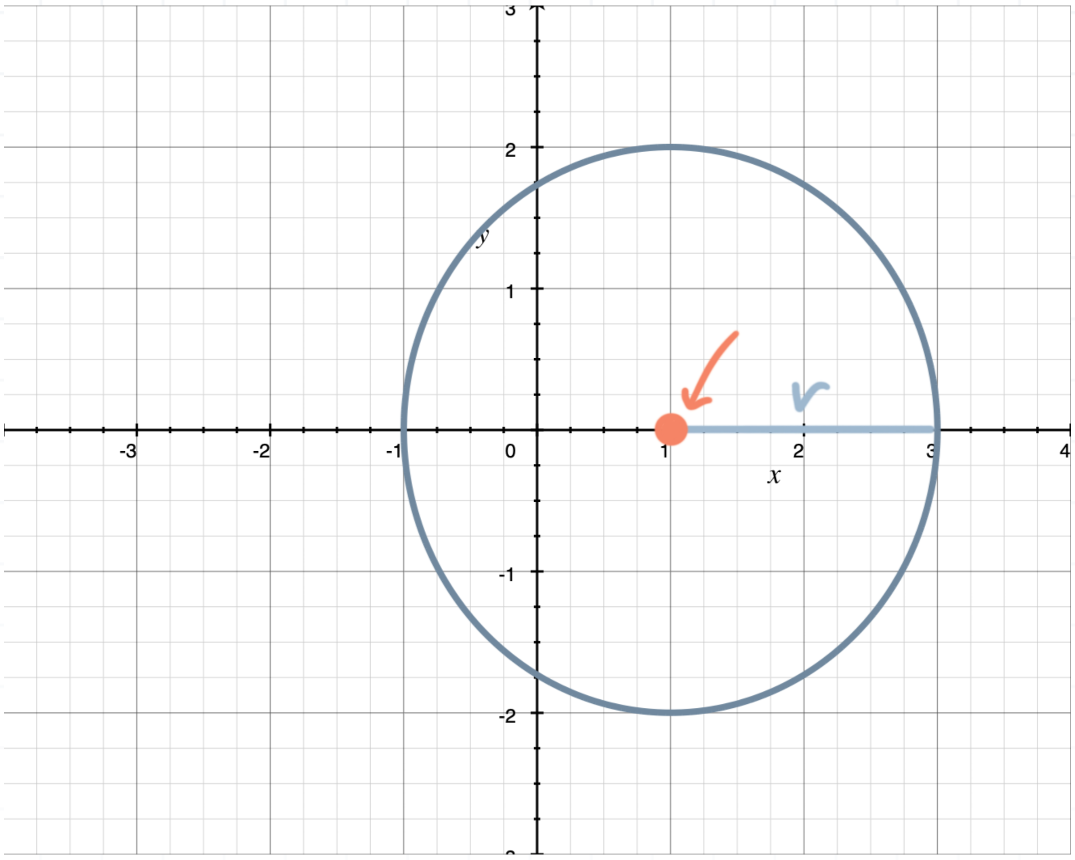
Equation (d) has both an x^2 and y^2 term, so it's either a circle, an ellipse, or a hyperbola. Because those terms have different signs (the y^2 term is positive and the x^2 term is negative) when they're on the same side of the equation, the equation represents a hyperbola.

Circles

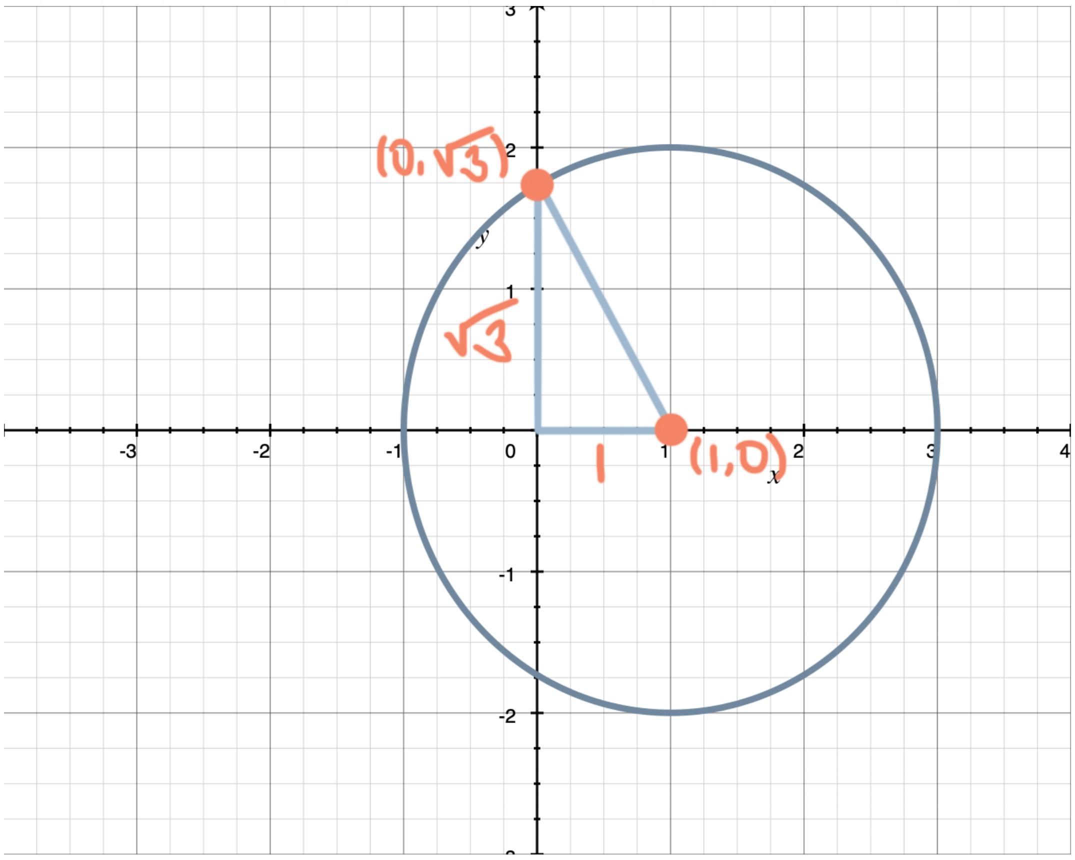
We all know what a circle looks like but, technically speaking, a **circle** is the set of points that are equidistant from one singular center point. Given the graph of a circle, we can visually find its **center** by locating the point in the middle.



The **radius** of a circle is the distance from the center to the edge.



So if we know the center point of the circle $(1,0)$, and a point on its edge like $(0,\sqrt{3})$, we can calculate the horizontal and vertical distance between the points, and use those as the legs of a right triangle.



Then we can find the radius of the circle using the Pythagorean Theorem.

$$a^2 + b^2 = r^2$$

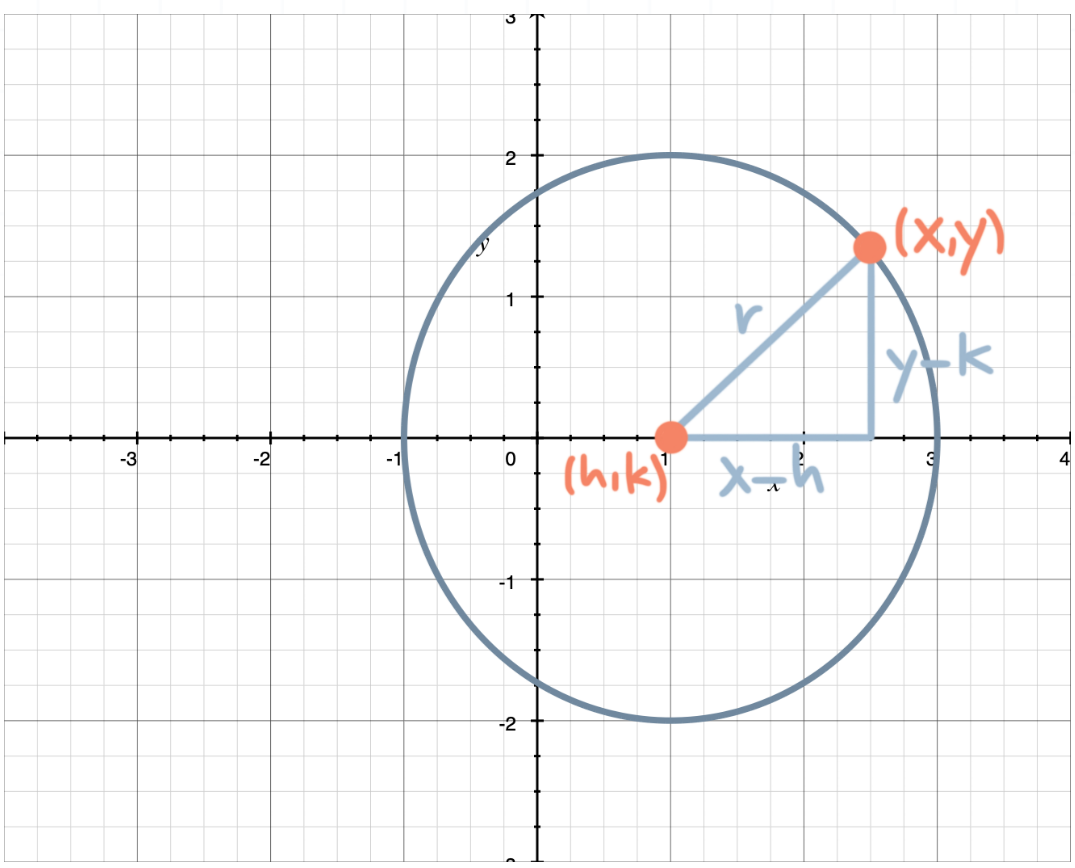
$$1^2 + (\sqrt{3})^2 = r^2$$

$$r^2 = 1 + 3$$

$$r^2 = 4$$

$$r = 2$$

If we call the center of the circle (h, k), and we choose a point on the circle (x, y), then we can generalize the formula for the equation of a circle.



The horizontal distance between the points is $x - h$, and the vertical distance between the points is $y - k$, so we can plug these into the Pythagorean Theorem to get the standard form for the **equation of a circle**.

$$a^2 + b^2 = c^2$$

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

Notice that if the circle is centered at the origin, then $(h,k)=(0,0)$ and the equation of the circle becomes

$$(x-0)^2+(y-0)^2=r^2$$

$$x^2+y^2=r^2$$

Let's do an example where we use the center of the circle and a radius to find its equation.

Example

If the center of a circle with radius 7 is $(-2,5)$, find the equation of the circle.

We have $r=7$, $h=-2$, and $k=5$. Plug the center and radius into the equation of the circle.

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

$$(x-(-2))^2+(y-5)^2=7^2$$

$$(x+2)^2+(y-5)^2=49$$

Let's do an example where we use the center of the circle and a point on the edge of the circle to find its equation.

Example

If the center of a circle is $(1,2)$ and a point on the circle is $(3,3)$, find the equation of the circle.

We need to start by finding the radius. The horizontal distance between the two points is $3 - 1 = 2$. The vertical distance between the points is $3 - 2 = 1$. Plugging these into the Pythagorean Theorem, we get

$$a^2 + b^2 = c^2$$

$$(3 - 1)^2 + (3 - 2)^2 = r^2$$

$$2^2 + 1^2 = r^2$$

$$r^2 = 4 + 1 = 5$$

$$r = \sqrt{5}$$

Now plug the center and radius into the equation of the circle.

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

$$(x - 1)^2 + (y - 2)^2 = (\sqrt{5})^2$$

$$(x - 1)^2 + (y - 2)^2 = 5$$

Keep in mind that we could also rewrite this equation by expanding the left side.

$$(x-1)^2+(y-2)^2=5$$

$$x^2-2x+1+y^2-4y+4=5$$

$$x^2+y^2-2x-4y+5=5$$

$$x^2+y^2-2x-4y=0$$

Sometimes we'll have this kind of expanded equation for the circle, and we'll need to complete the square in order to put the equation back into standard form, $(x-h)^2+(y-k)^2=r^2$.

We learned how to "complete the square" in Algebra when we learned how to solve quadratic equations, but let's do another example here to remind ourselves what that process looks like.

Example

Find the center and radius of the circle.

$$x^2+y^2+2x+6y+1=0$$

Group together the x and y terms.

$$(x^2+2x)+(y^2+6y)+1=0$$

To complete the square with respect to x, we'll take the coefficient of 2 on $2x$ and divide it by 2 to get $2/2=1$. We'll square this value to get $1^2=1$, and then we'll add this result to both sides of the equation.

$$(x^2+2x+1)+(y^2+6y)+1=0+1$$

To complete the square with respect to y, we'll take the coefficient of 6 on $6y$ and divide it by 2 to get $6/2 = 3$. We'll square this value to get $3^2 = 9$, and then we'll add this value to both sides of the equation.

$$(x^2 + 2x + 1) + (y^2 + 6y + 9) + 1 = 0 + 1 + 9$$

Simplify the constants.

$$(x^2 + 2x + 1) + (y^2 + 6y + 9) = 9$$

The steps we've taken to complete the square with respect to both x and y leaves us with two trinomials that we can factor as perfect squares.

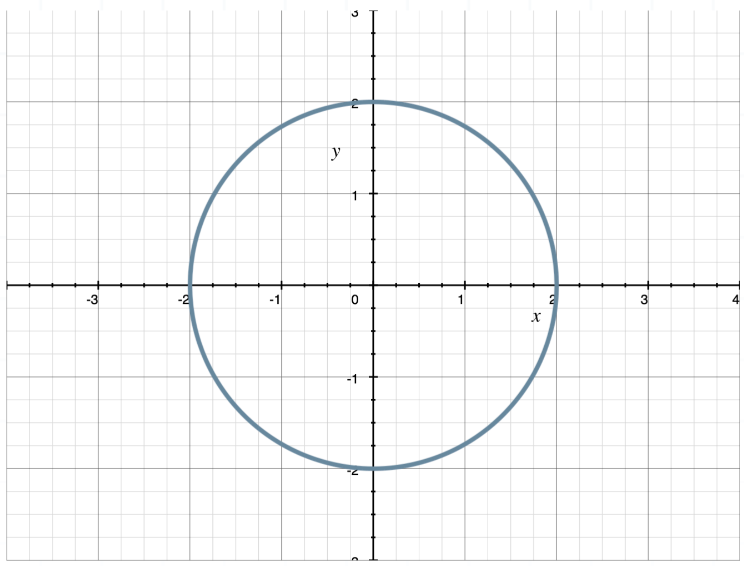
$$(x + 1)^2 + (y + 3)^2 = 9$$

Now that the equation is in standard form, we can see that the center of the circle is at $(h, k) = (-1, -3)$, and the radius is $r = \sqrt{9} = 3$.

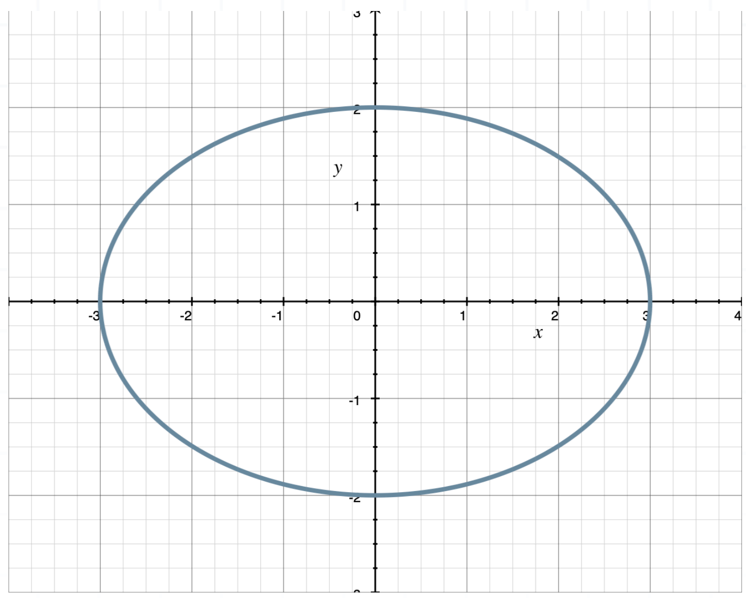
Ellipses

An ellipse is a lot like a circle, except that in an ellipse, the length of the radius can change. A circle is a special case of an ellipse in which the radius is constant everywhere.

*Circle with **constant radius**:*

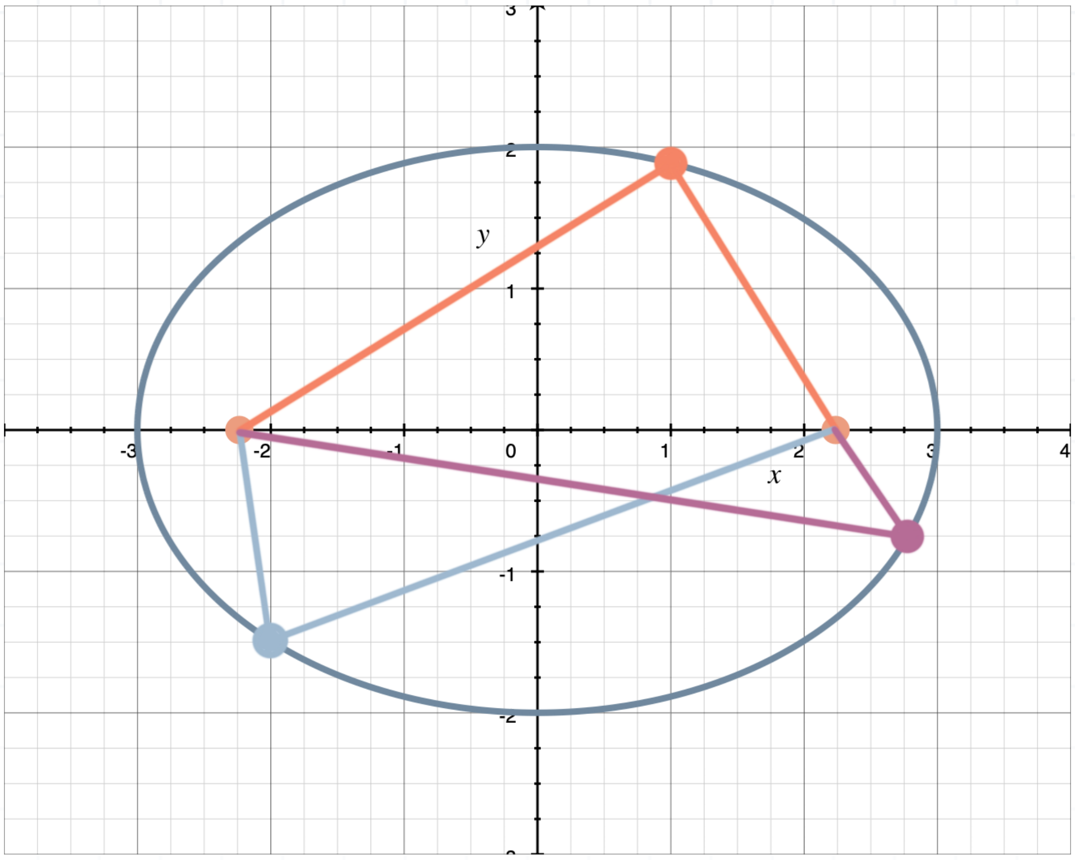


*Ellipse with **varying radius**:*



Put simply, an **ellipse** is the set of points that are equidistant from two other points. In other words, if we set two points in the plane, we can draw an ellipse around them by keeping the sum of the distances from those points constant.

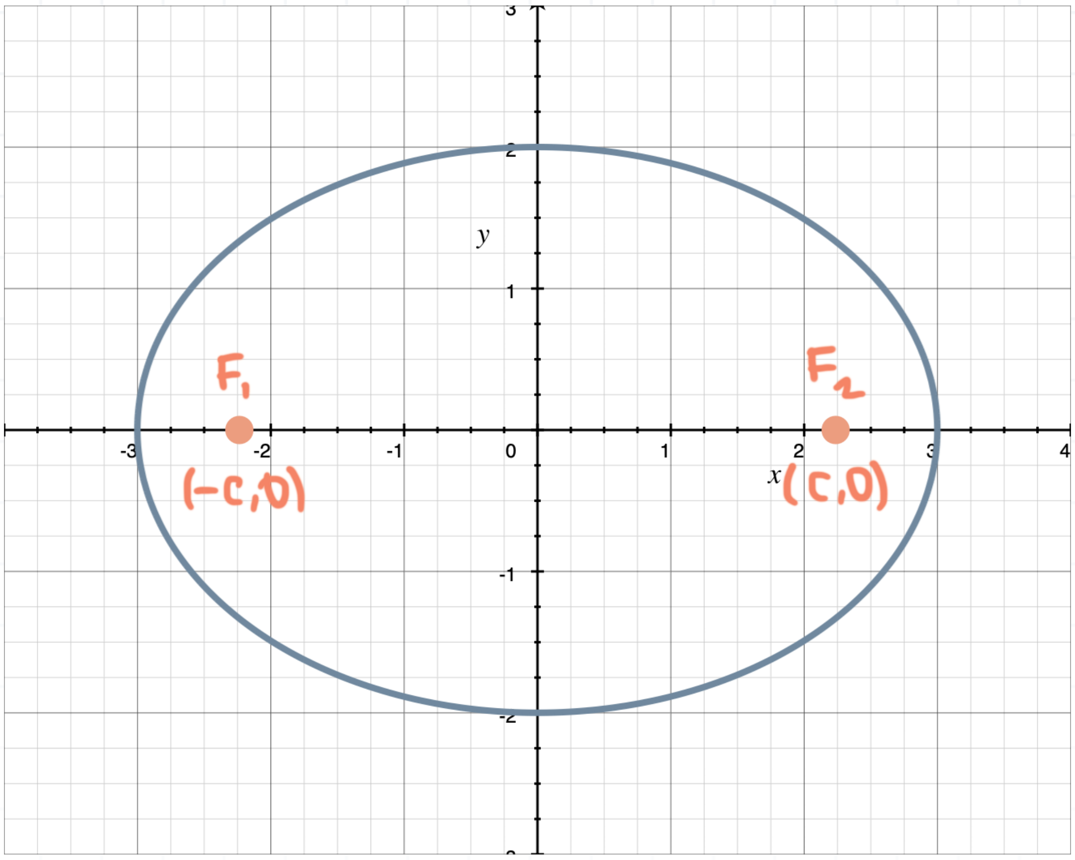
In the sketch, the sum of the two red distances is equivalent to the sum of the two light blue distances which is equivalent to the sum of the two purple distances, etc.



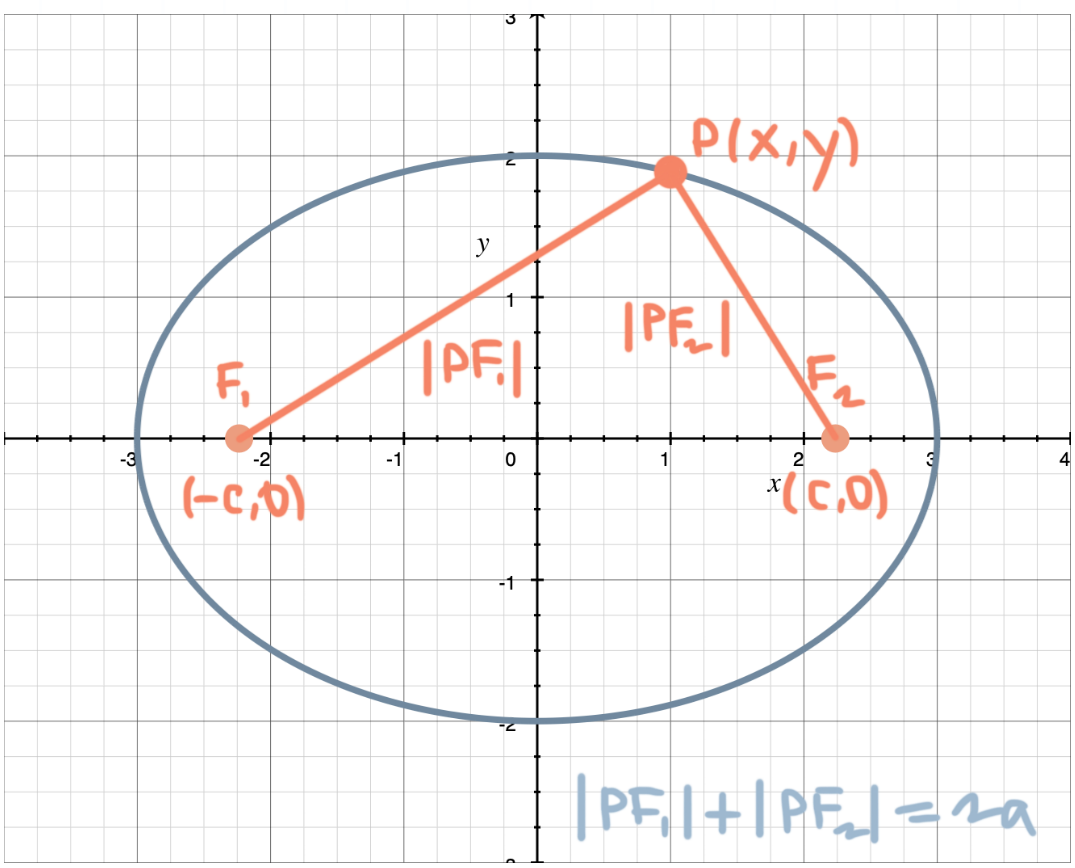
So we can sketch two lines connecting any point on the ellipse to the two special points, and the sum of the lengths of those two lines will always be the same.

The equation of an ellipse

To find the standard form for the equation of an ellipse, we can position the ellipse so that the special points that define it are $F_1(-c,0)$ and $F_2(c,0)$, symmetric across the vertical axis and equidistant from the origin.



We can pick any point on the ellipse $P(x, y)$, and then define the distances from P to the special points as $|PF_1|$ and $|PF_2|$. We know that $|PF_1| + |PF_2|$ is constant, so we'll say that this value is always equal to $2a$.



We choose $2a$ because describing the constant $|PF_1| + |PF_2|$ this way will make our math easier.

$$|PF_1| + |PF_2| = 2a$$

$$\sqrt{(x+c)^2 + y^2} + \sqrt{(x-c)^2 + y^2} = 2a$$

$$\sqrt{(x-c)^2 + y^2} = 2a - \sqrt{(x+c)^2 + y^2}$$

If we square both sides of the equation and then expand it, we get

$$(x-c)^2 + y^2 = 4a^2 - 4a\sqrt{(x+c)^2 + y^2} + (x+c)^2 + y^2$$

$$x^2 - 2cx + c^2 + y^2 = 4a^2 - 4a\sqrt{(x+c)^2 + y^2} + x^2 + 2cx + c^2 + y^2$$

$$4a\sqrt{(x+c)^2 + y^2} = 4a^2 + 4cx$$

$$a\sqrt{(x+c)^2 + y^2} = a^2 + cx$$

Square both sides and expand again.

$$a^2((x+c)^2 + y^2) = (a^2 + cx)^2$$

$$a^2(x^2 + 2cx + c^2 + y^2) = a^4 + 2a^2cx + c^2x^2$$

$$a^2x^2 + 2a^2cx + a^2c^2 + a^2y^2 = a^4 + 2a^2cx + c^2x^2$$

$$a^2x^2 + a^2c^2 + a^2y^2 = a^4 + c^2x^2$$

Rearrange the terms, then factor.

$$a^2x^2 - c^2x^2 + a^2y^2 = a^4 - a^2c^2$$

$$(a^2 - c^2)x^2 + a^2y^2 = a^2(a^2 - c^2)$$

If we make the substitution $b^2 = a^2 - c^2$, then we get

$$b^2x^2 + a^2y^2 = a^2b^2$$

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

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$$

Therefore, the standard form for **the equation of an ellipse** is

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$$ Centered at the origin

$$\frac{(x-h)^2}{a^2} + \frac{(y-k)^2}{b^2} = 1$$ Shifted off the origin

where a is the radius in the x-direction, b is the radius in the y-direction, and (h, k) is the center of the ellipse.

Sometimes we'll see the equation of the ellipse given with a^2 and b^2 in opposite positions to their places above, such that a^2 is paired with y^2 and b^2 with x^2, instead of a^2 paired with x^2 and b^2 paired with y^2.

$$\frac{x^2}{b^2} + \frac{y^2}{a^2} = 1$$ Centered at the origin

$$\frac{(x-h)^2}{b^2} + \frac{(y-k)^2}{a^2} = 1$$ Shifted off the origin

The constant a always represents the length of the major radius, while the constant b always represents the length of the minor radius. Which means that ellipse equations that pair a with x are wide ellipses, while ellipse equations that pair a with y are tall ellipses.

$$\frac{x^2}{a^2}+\frac{y^2}{b^2}=1$$ Wide ellipse centered at the origin

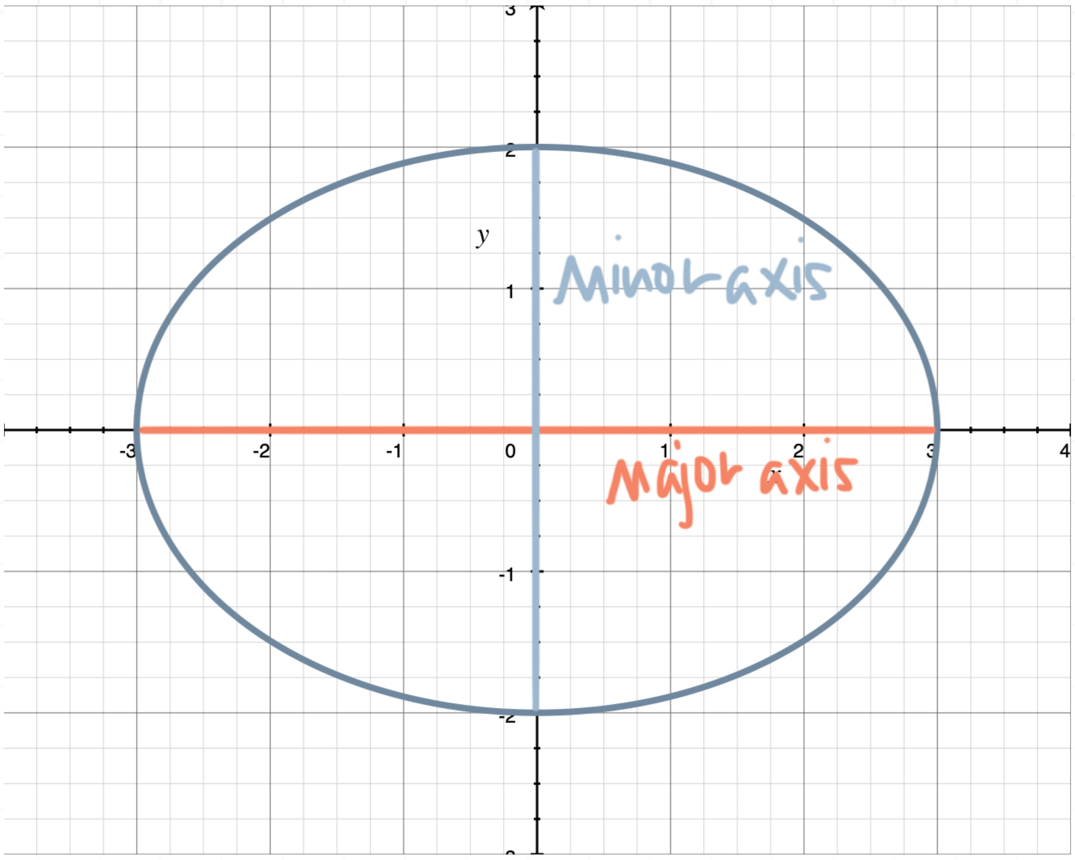
$$\frac{x^2}{b^2}+\frac{y^2}{a^2}=1$$ Tall ellipse centered at the origin

$$\frac{(x-h)^2}{a^2}+\frac{(y-k)^2}{b^2}=1$$ Wide ellipse shifted off the origin

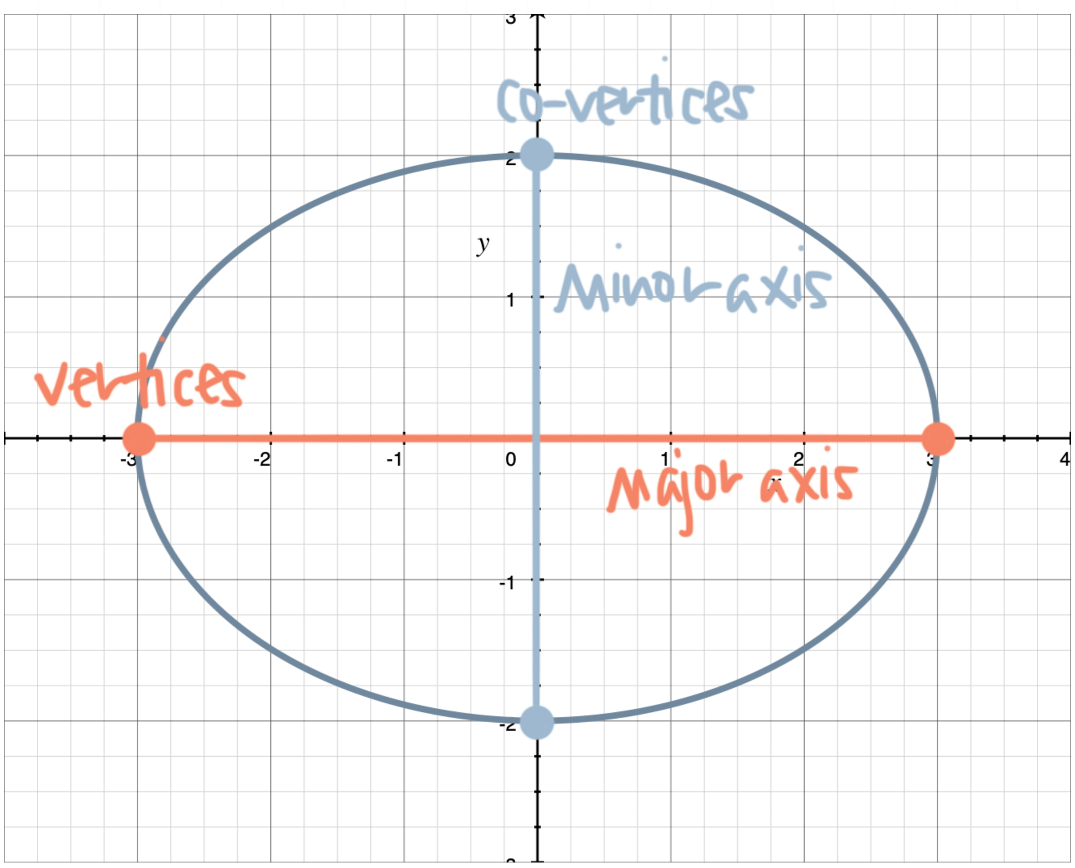
$$\frac{(x-h)^2}{b^2}+\frac{(y-k)^2}{a^2}=1$$ Tall ellipse shifted off the origin

Axes and vertices

An ellipse always has two axes that run through its center. The **major axis** is the longer axis, and the **minor axis** is the shorter axis. Sometimes half of the major axis is called the **semi-major axis** or major radius, and half of the minor axis is called the **semi-minor axis** or minor radius.



The points on the ellipse at the end of the major axis are the **vertices**, and the points on the ellipse at the end of the minor axis are the **co-vertices**.



Let’s do an example where we sketch the graph of the ellipse from its equation.

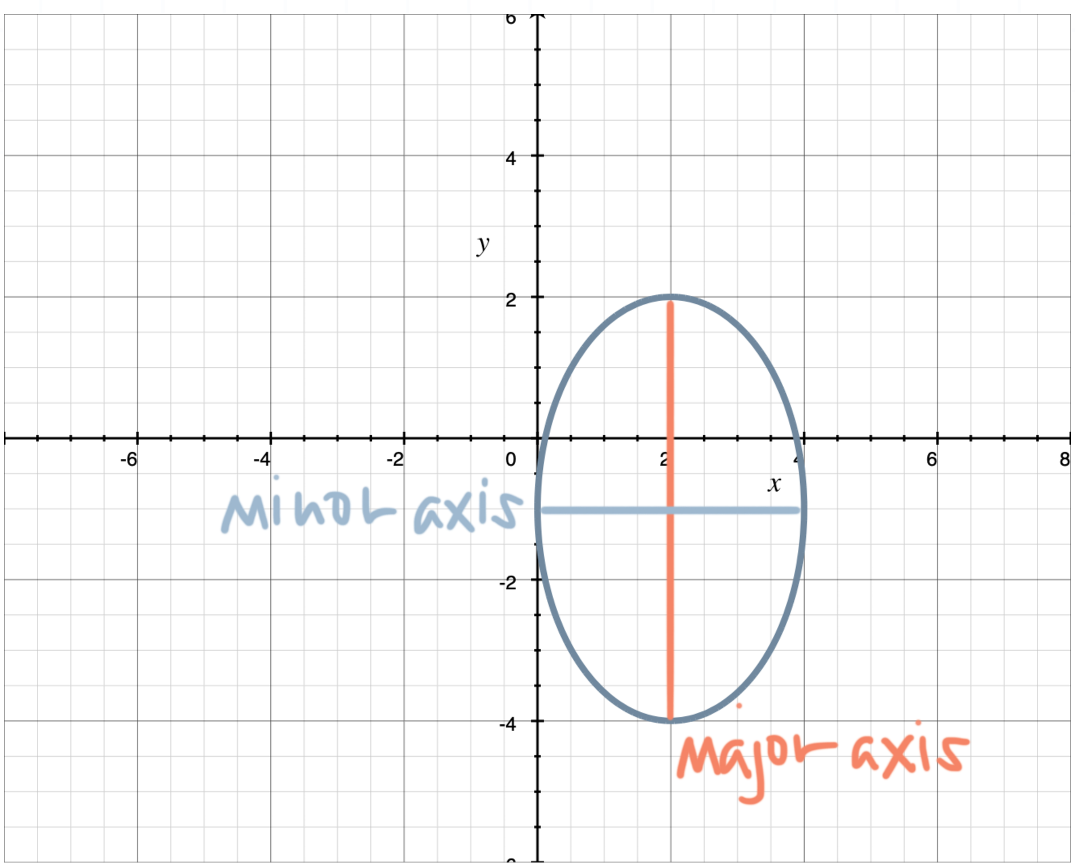
Example

Graph the ellipse by finding its center and major and minor radii.

$$\frac{(x-2)^2}{4}+\frac{(y+1)^2}{9}=1$$

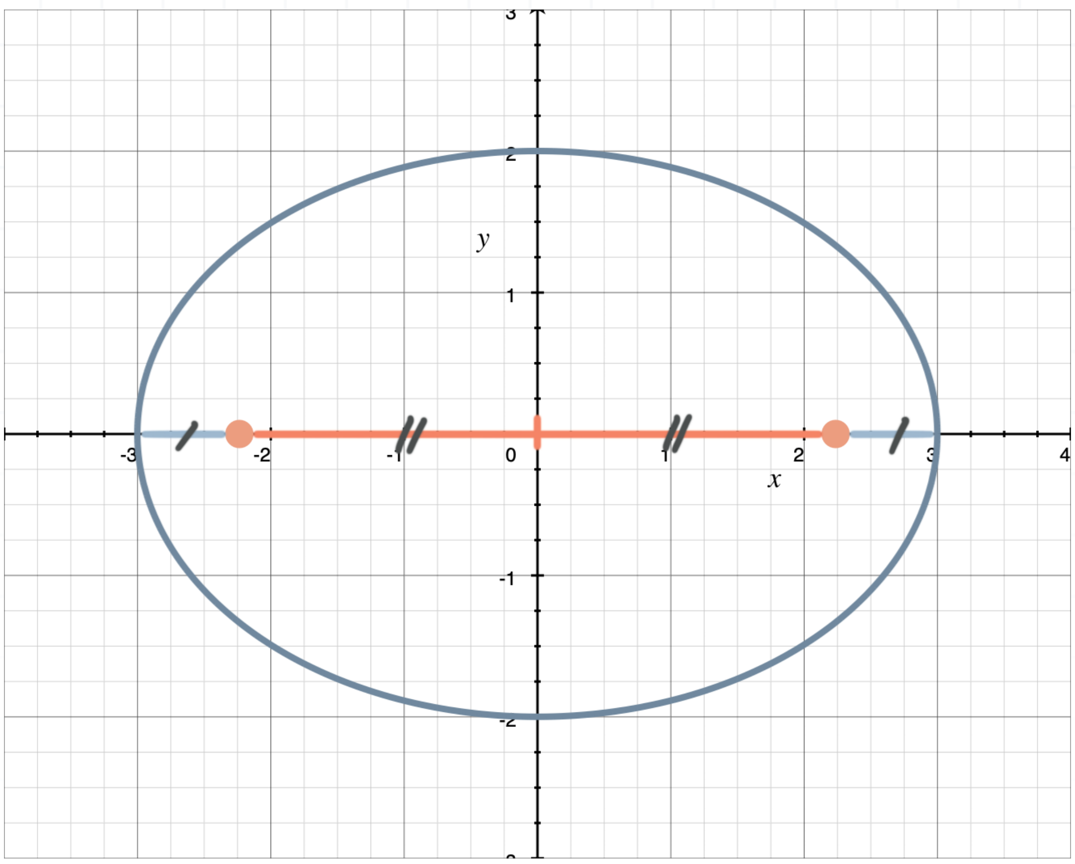
We need to find the center of the ellipse and the length of each axis. Based on the equation, the center is at $(2,-1)$.

The length of the horizontal radius is $\sqrt{4}=2$ and the length of the vertical radius is $\sqrt{9}=3$. So the graph of the ellipse is

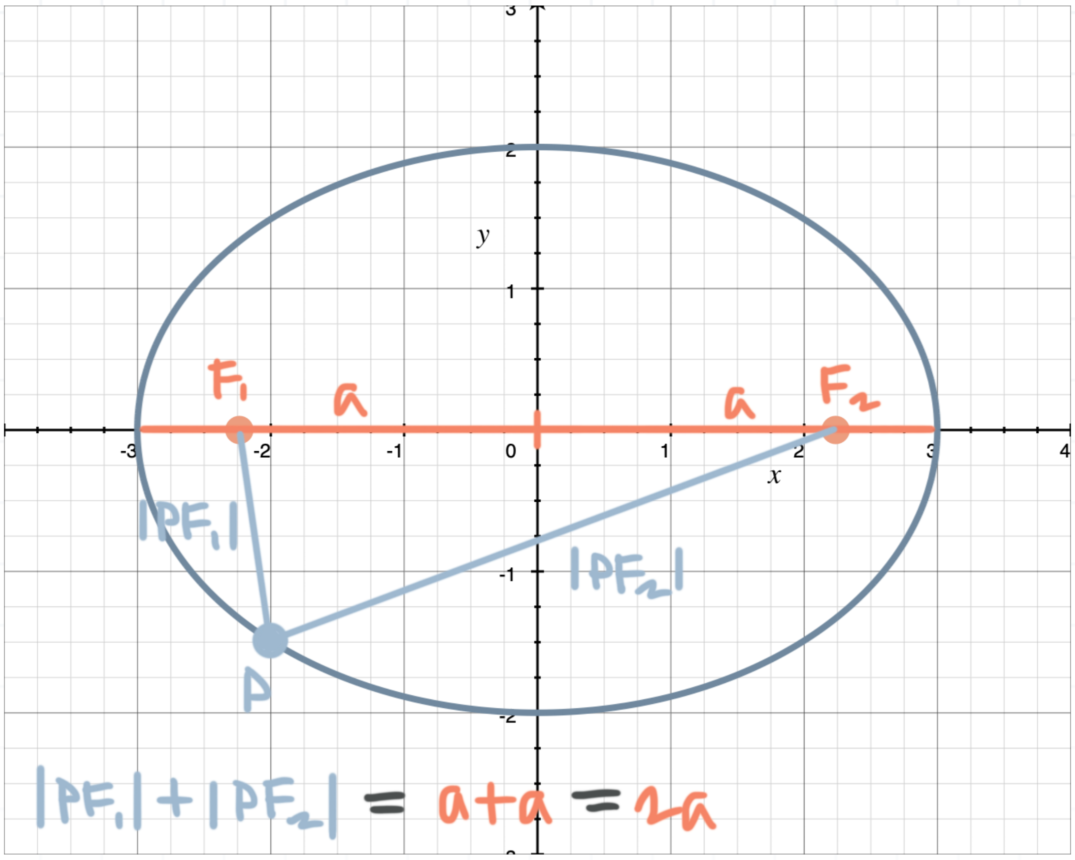


Foci and directrices

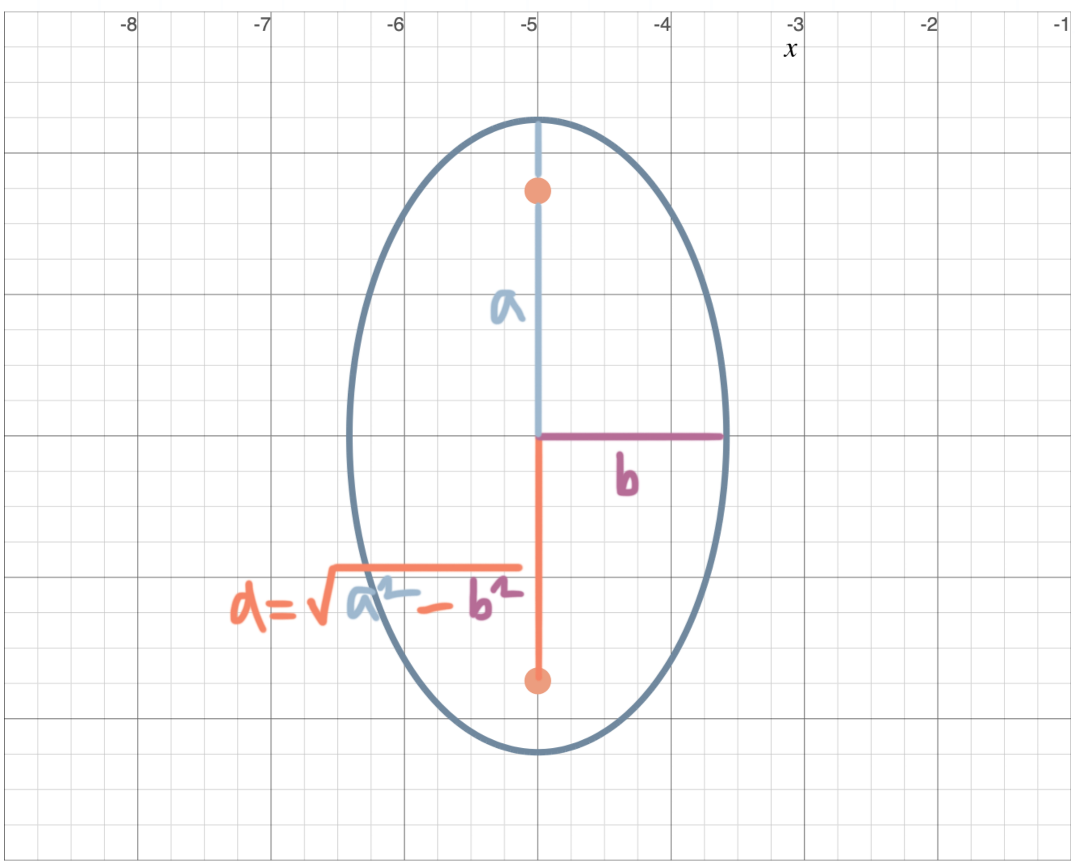
An ellipse also has two very important points, called the **foci** of the ellipse, that sit on the major axis, inside the ellipse. These are the two special points we talked about earlier that define the ellipse, and they're equidistant from the center of the ellipse and from the edge of the ellipse.



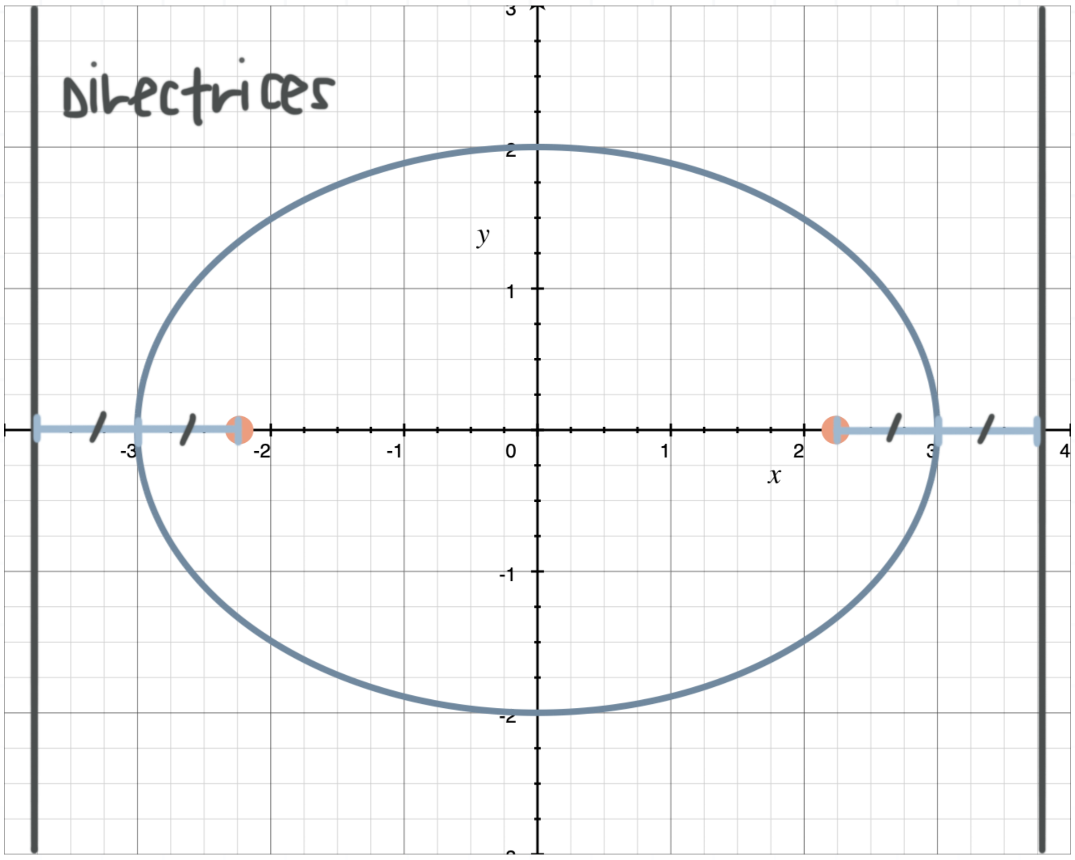
If we sum the distances from any point on the ellipse to both foci, the sum of those distances will always be equivalent to the length of the major axis (or double the length of the major radius), no matter which point we choose.



The **focal length**, which is the distance from the center of the ellipse to the focus, is always given by $d = \sqrt{a^2 - b^2}$, where a is the length of the major radius and b is the length of the minor radius.

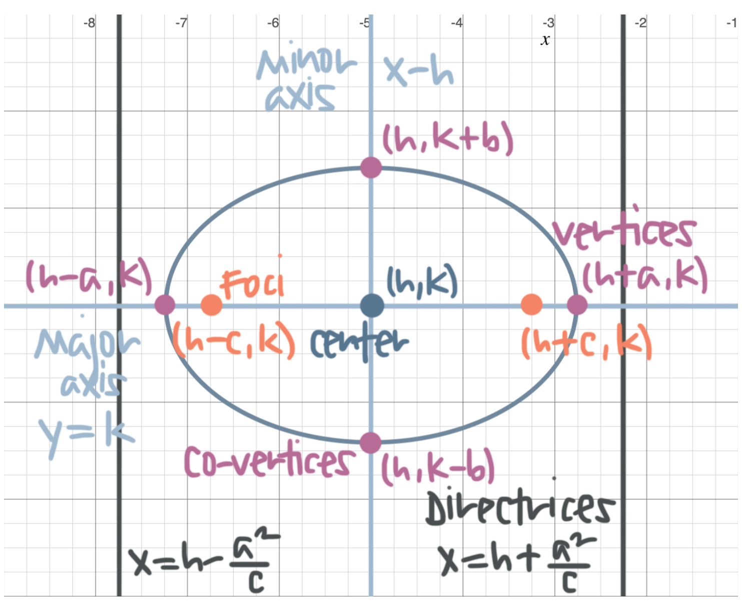


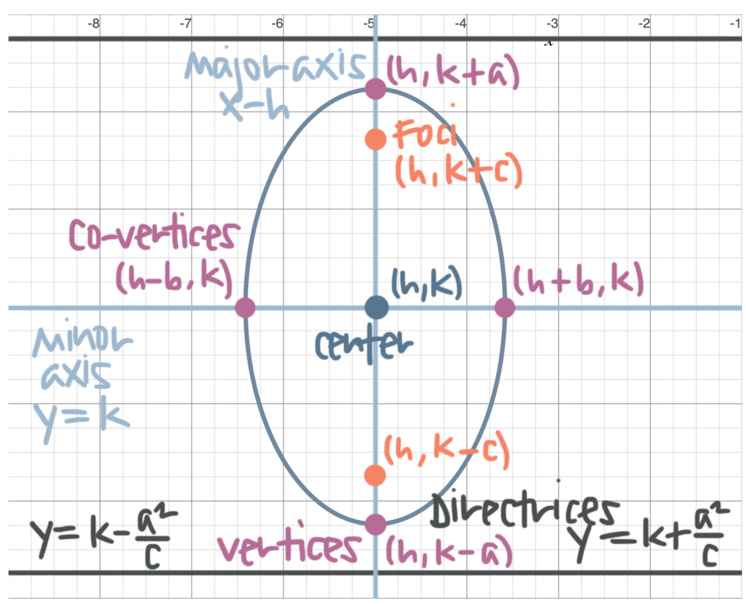
The **directrices** of the ellipse are two lines parallel to the minor axis (and perpendicular to the major axis) that sit outside the ellipse. The distance between the directrices and the vertices is the same as the distance between the vertices and the foci.



For ellipses centered away from the origin at (h, k), here's everything we know.

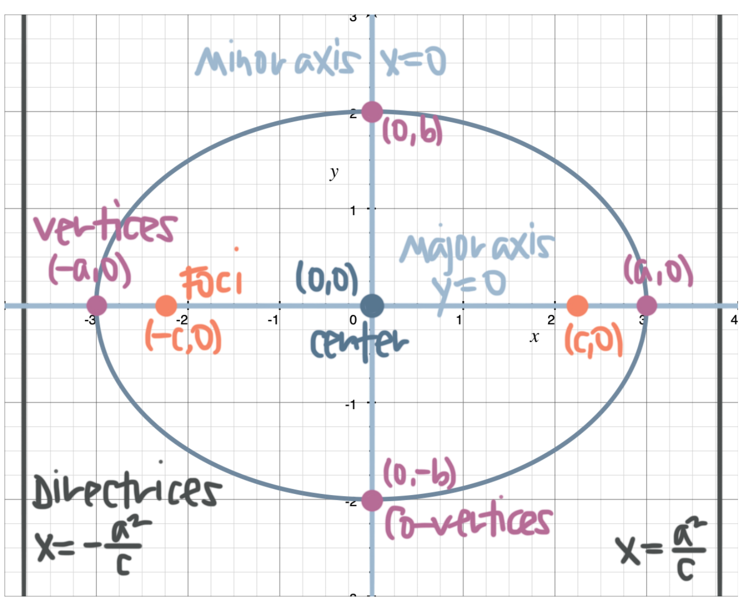
	Wide, $a \geq b > 0$	**Tall,** $a \geq b > 0$
Equation	$\dfrac{(x-h)^2}{a^2} + \dfrac{(y-k)^2}{b^2} = 1$	$\dfrac{(x-h)^2}{b^2} + \dfrac{(y-k)^2}{a^2} = 1$
Center	(h, k)	(h, k)
Vertices	$(h \pm a, k)$	$(h, k \pm a)$
Co-vertices	$(h, k \pm b)$	$(h \pm b, k)$
Major axis	$y = k$, length $2a$	$x = h$, length $2a$
Minor axis	$x = h$, length $2b$	$y = k$, length $2b$
Foci	$(h \pm c, k)$ with $c^2 = a^2 - b^2$	$(h, k \pm c)$ with $c^2 = a^2 - b^2$
Directrices	$x = h \pm \dfrac{a^2}{c}$	$y = k \pm \dfrac{a^2}{c}$

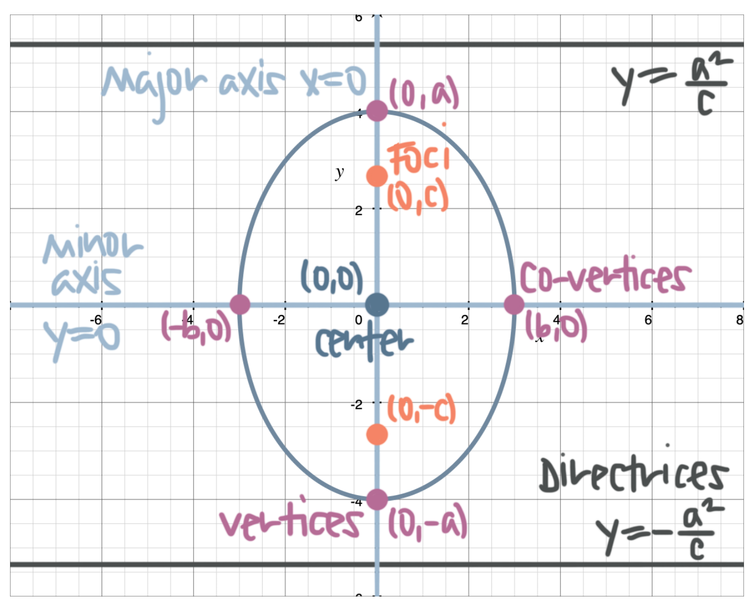




And for ellipses centered at the origin $(0,0)$, here's how that table simplifies.

	Wide, $a \geq b > 0$	**Tall,** $a \geq b > 0$
Equation	$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$	$\frac{x^2}{b^2} + \frac{y^2}{a^2} = 1$
Center	$(0,0)$	$(0,0)$
Vertices	$(\pm a,0)$	$(0, \pm a)$
Co-vertices	$(0, \pm b)$	$(\pm b,0)$
Major axis	$y = 0$, length $2a$	$x = 0$, length $2a$
Minor axis	$x = 0$, length $2b$	$y = 0$, length $2b$
Foci	$(\pm c,0)$ with $c^2 = a^2 - b^2$	$(0, \pm c)$ with $c^2 = a^2 - b^2$
Directrices	$x = \pm \frac{a^2}{c}$	$y = \pm \frac{a^2}{c}$





Let's do an example where we find the coordinates of the foci.

Example

Find the coordinates of each focus of the ellipse.

$$\frac{(x-2)^2}{4}+\frac{(y+1)^2}{9}=1$$

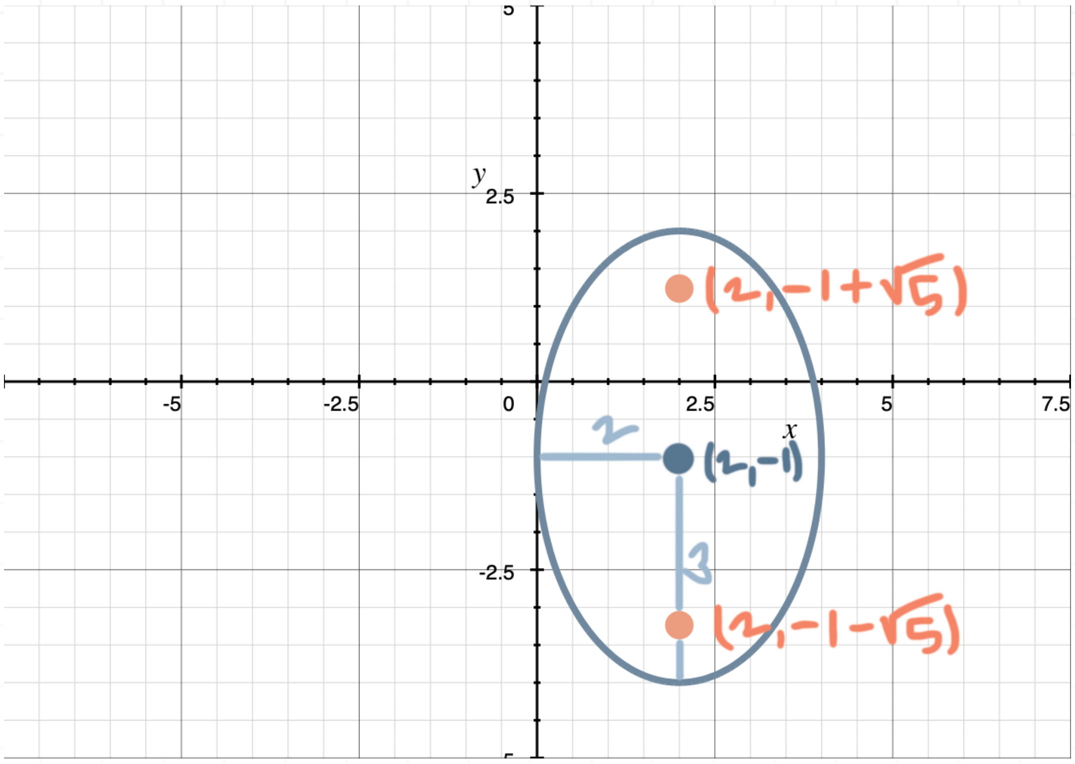
The center of the ellipse is at $(2,-1)$. The length of the major radius is $\sqrt{9}=3$, and the length of the minor radius is $\sqrt{4}=2$. Therefore, the focal length is

$$c=\sqrt{b^2-a^2}=\sqrt{3^2-2^2}=\sqrt{9-4}=\sqrt{5}$$

The major axis is the vertical axis, which means we need to add and subtract $\sqrt{5}$ from from the y-value of the center in order to find the foci. Therefore, the coordinates of the foci are

$(2,-1+\sqrt{5})$ and $(2,-1-\sqrt{5})$

If we sketch the ellipse and include the foci, we can see how they sit along the major axis inside the ellipse.



We won't always be given the equation of the ellipse in standard form. So let's walk through an example where we're required to put the ellipse in standard form before we sketch it.

Example

Sketch the graph of the ellipse.

$$12x + 5y^2 + 28 = -2x^2 + 20y$$

To put the equation of the ellipse in standard form, start by collecting all the non-constant terms onto one side of the equation. We can put the constant term alone on the other side.

$$12x + 5y^2 + 28 = -2x^2 + 20y$$

$$2x^2 + 12x + 5y^2 - 20y = -28$$

Because the x^2 and y^2 terms have the same sign when they're on the same side of the equation (they're both positive), we know the equation has to represent an ellipse.

We need to complete the square with respect to both variables. We'll start with x. Remember that the coefficient on x^2 must be 1 before we can complete the square.

$$x^2 + 6x + \frac{5}{2}y^2 - 10y = -14$$

$$x^2 + 6x + 9 + \frac{5}{2}y^2 - 10y = -14 + 9$$

$$(x+3)^2 + \frac{5}{2}y^2 - 10y = -5$$

Now we'll complete the square with respect to y. We'll start by making the coefficient on y^2 equal to 1.

$$\frac{2}{5}(x+3)^2 + y^2 - 4y = -2$$

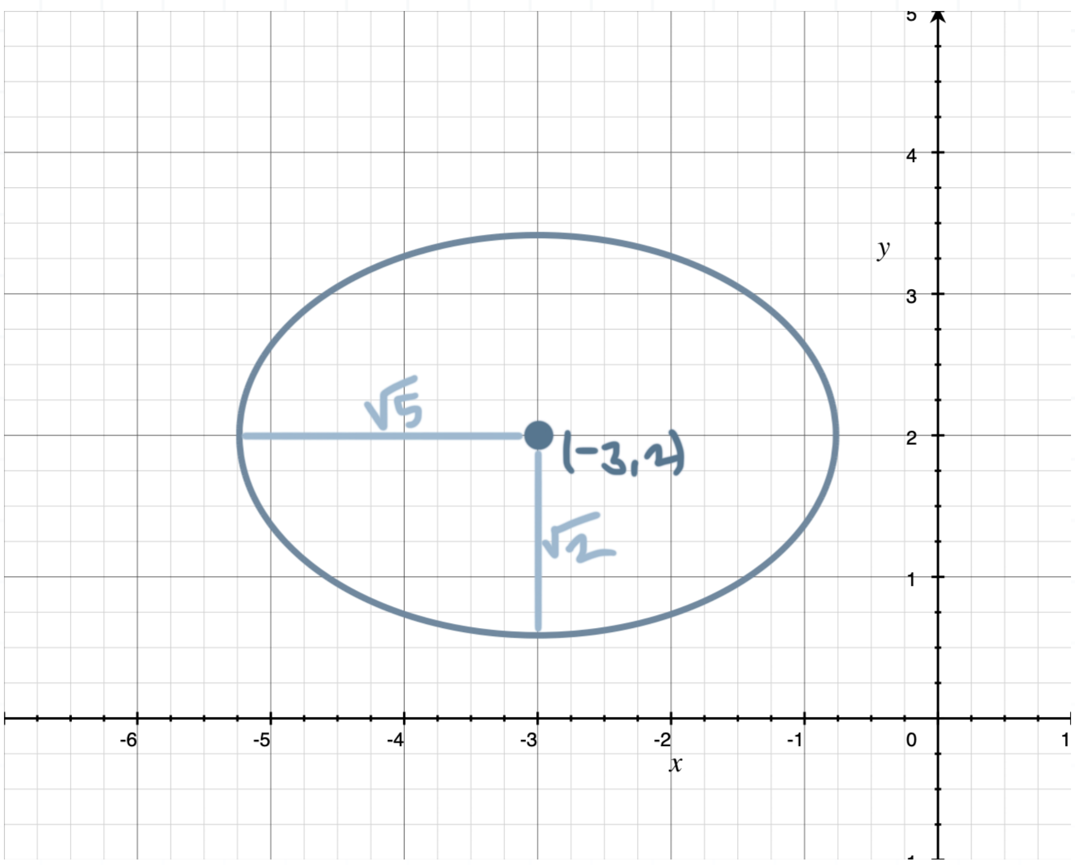
$$\frac{2}{5}(x+3)^2 + y^2 - 4y + 4 = -2 + 4$$

$$\frac{2}{5}(x+3)^2 + (y-2)^2 = 2$$

Last, we'll divide through by 2 to put the ellipse in standard form.

$$\frac{(x+3)^2}{5} + \frac{(y-2)^2}{2} = 1$$

With the ellipse in standard form, we can see that the center is at $(-3,2)$. The length of the major radius is $\sqrt{5}$, and the length of the minor radius is $\sqrt{2}$. The ellipse is wider than it is tall, and the sketch of its graph is

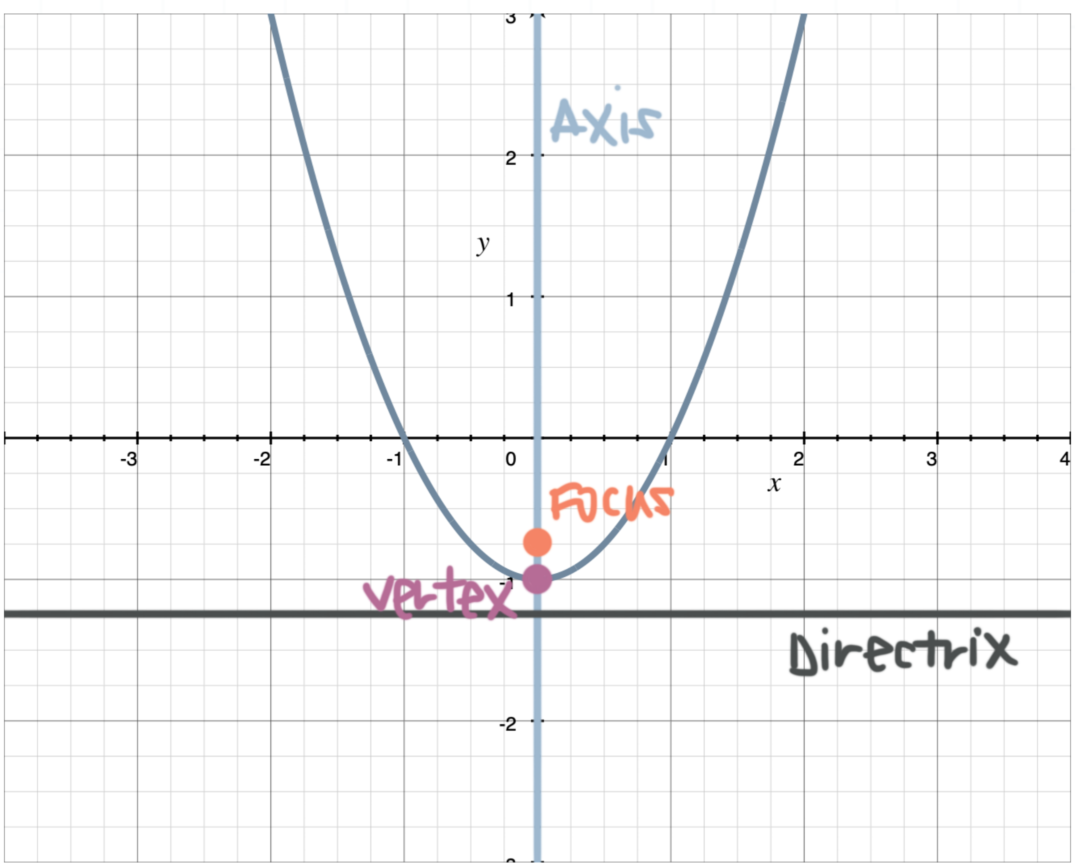


Parabolas

A **parabola** is the set of all points which are equidistant from its own focus and its directrix. The **focus** will be a point inside the parabola's "bowl", and the **directrix** will be a line outside the parabola that's perpendicular to the parabola's axis.

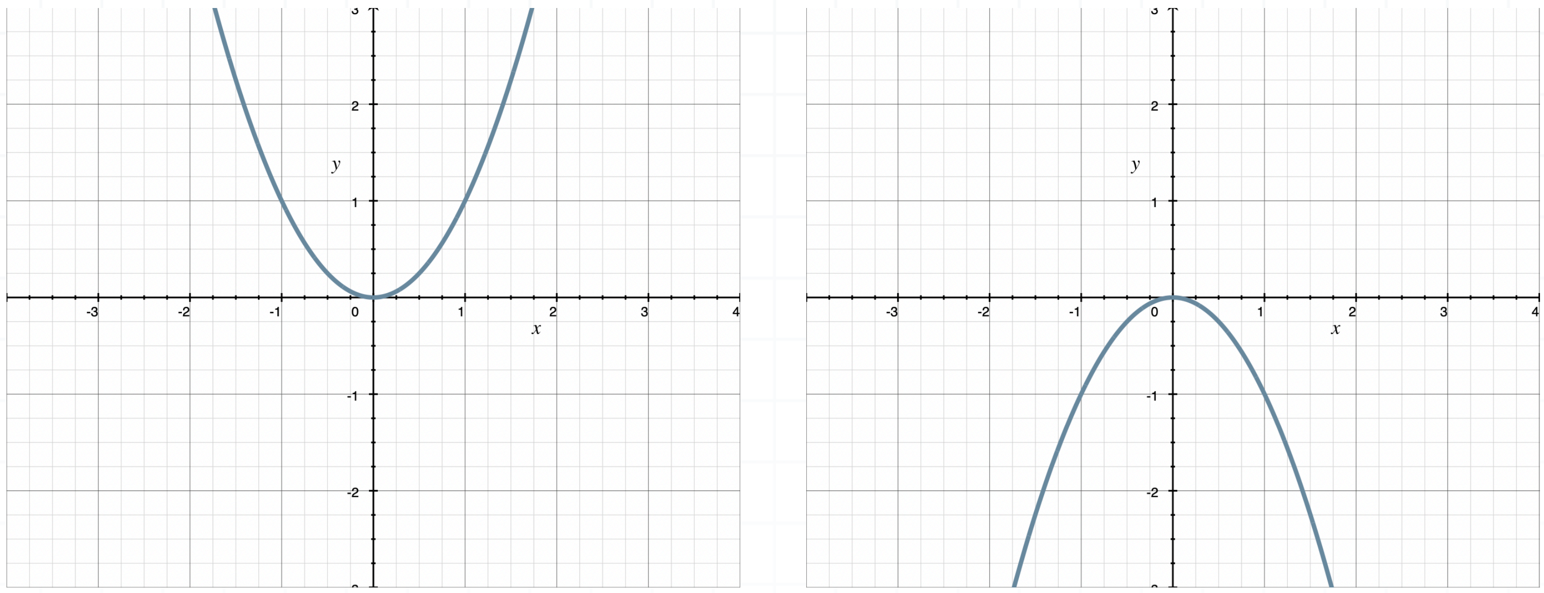
The **axis** is the line of symmetry of the parabola, which runs through the parabola's **vertex**, the point at which the parabola intersects its axis.

Because the focus and directrix are always equidistant from the parabola, the parabola's vertex is the midpoint between the focus and the directrix.

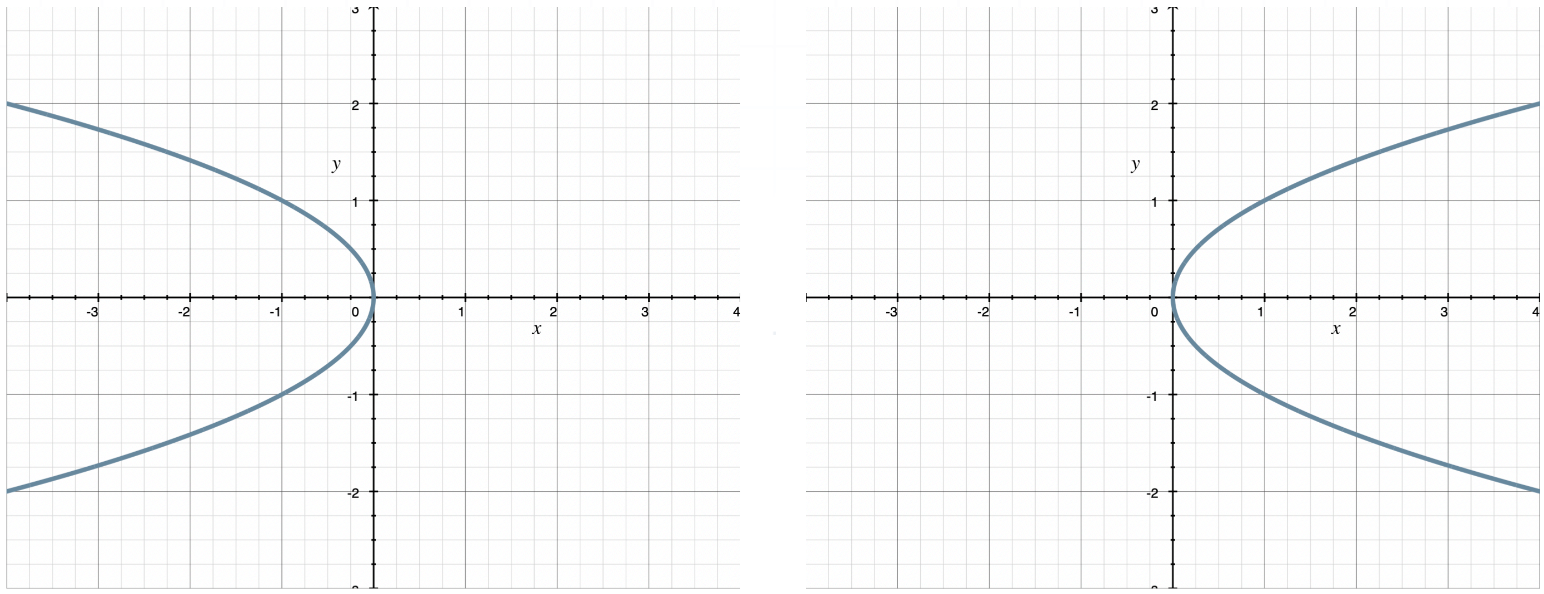


Directionality and position

When we talked earlier about ellipses, we learned that they could be taller than they are wide, or wider than they are tall. Parabolas are similar in the sense that they can open up or down,

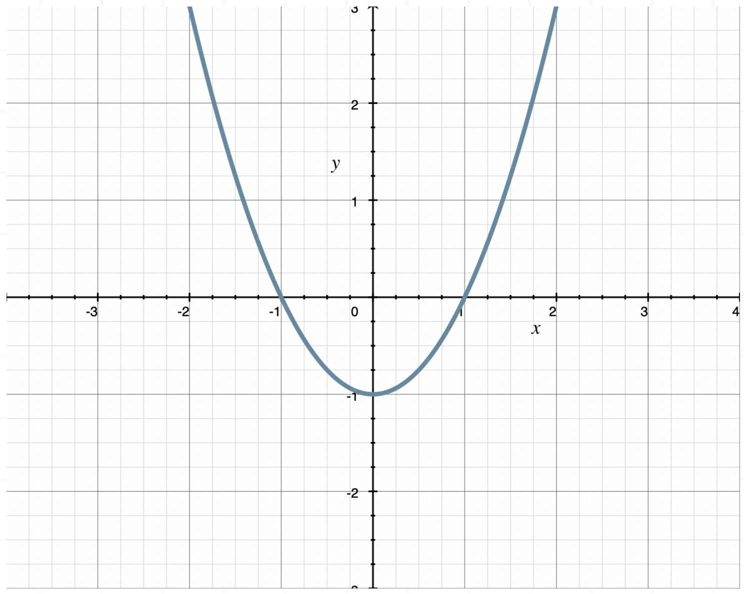


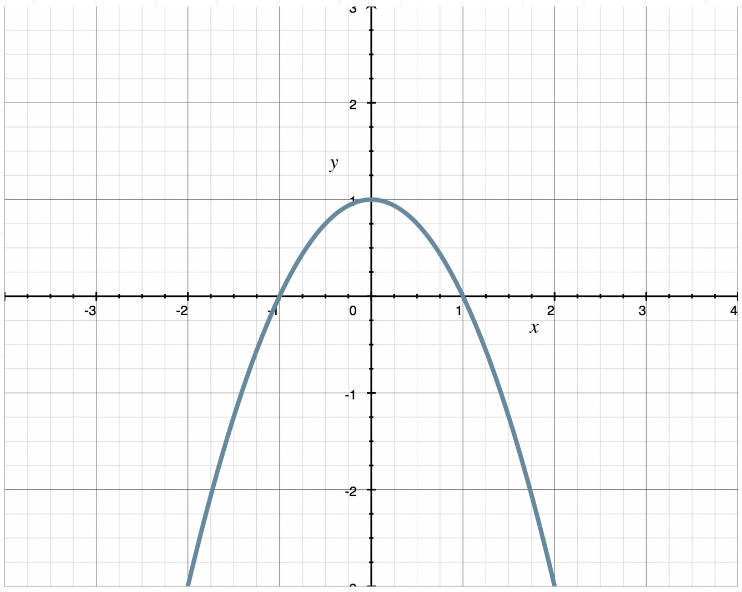
but they can also open to the left or right.



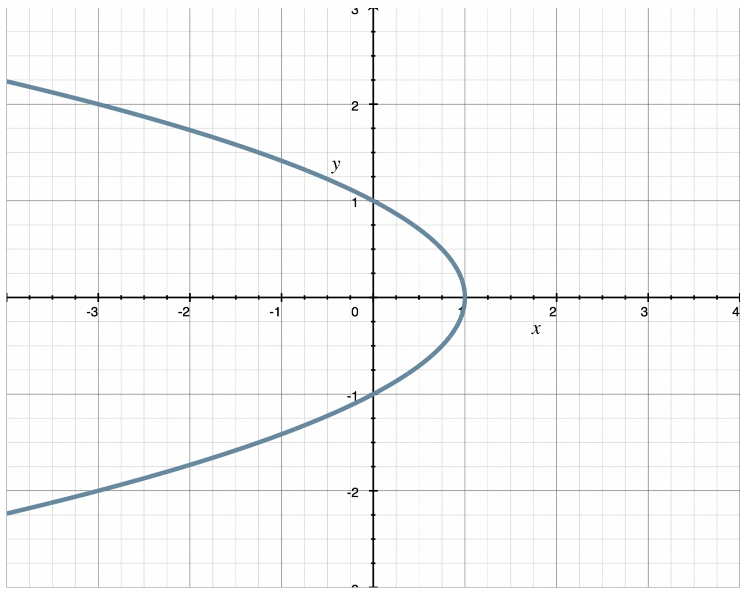
The vertex of the parabola is the minimum point if the parabola opens up, it’s the maximum point if the parabola opens down, it’s the right-most point if the parabola opens to the left, and it’s the left-most point if the parabola opens to the right.

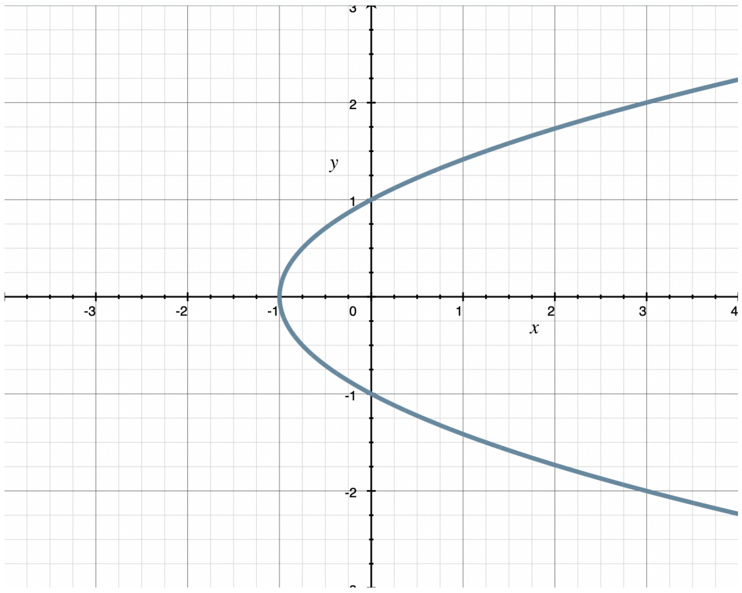
And just like ellipses, which could be centered at the origin or shifted off the origin, parabolas can have their vertex at the origin, like each of the directional parabolas above, or the vertex can be shifted off the origin, like these up/down parabolas,





and these left/right parabolas.





Three forms of the parabola's equation

The equation of a parabola is usually given in one of three forms.

Vertex form	$y = a(x - h)^2 + k$
Conics form	$4p(y - k) = (x - h)^2$

Standard form $y = ax^2 + bx + c$

Sometimes conics form is also called standard form, but we'll call it "conics form" to distinguish it from the other "standard form."

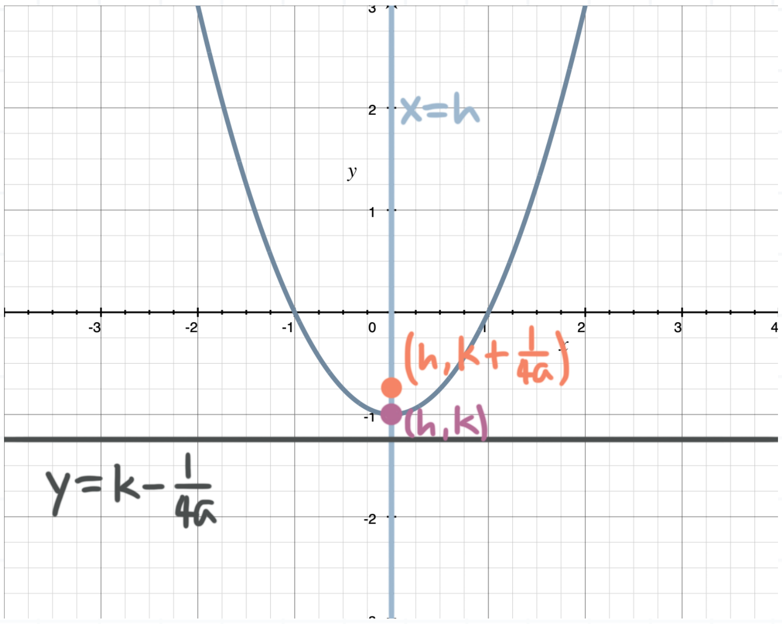
These forms are all useful in different contexts. Standard form is what we usually see in Algebra when we're factoring quadratics. Vertex form allows us to quickly identify the vertex as (h, k), while conics form is especially useful for identifying p as the distance between the vertex and the focus (and therefore also the distance between the vertex and the directrix).

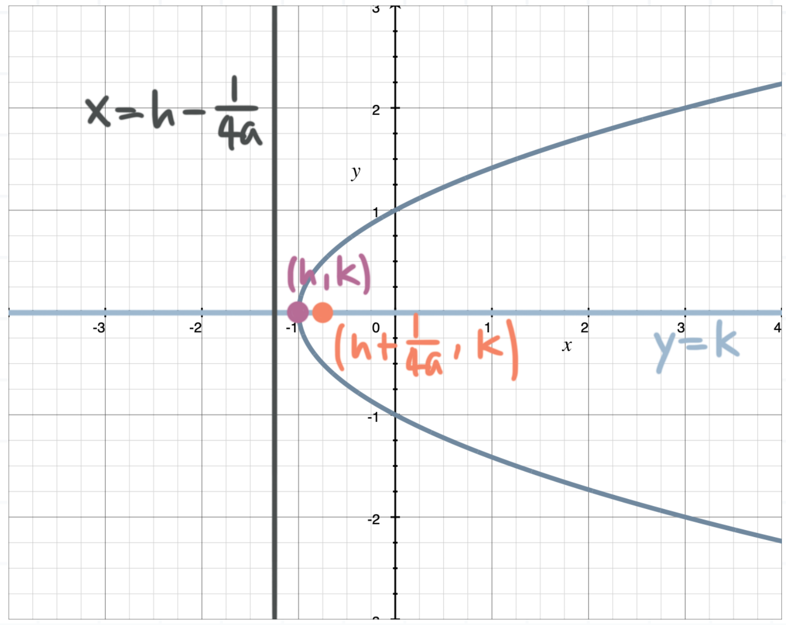
Here are the formulas we'll use for each part of the parabola when the equation is given in vertex form.

	Equation	**Vertex**	**Axis**	**Focus**	**Directrix**
Up/down shifted	$y = a(x-h)^2 + k$	(h, k)	$x = h$	$\left(h, k + \frac{1}{4a}\right)$	$y = k - \frac{1}{4a}$
Left/right shifted	$x = a(y-k)^2 + h$	(h, k)	$y = k$	$\left(h + \frac{1}{4a}, k\right)$	$x = h - \frac{1}{4a}$
Up/down origin	$y = ax^2$	$(0,0)$	$x = 0$	$\left(0, \frac{1}{4a}\right)$	$y = -\frac{1}{4a}$
Left/right origin	$x = ay^2$	$(0,0)$	$y = 0$	$\left(\frac{1}{4a}, 0\right)$	$x = -\frac{1}{4a}$

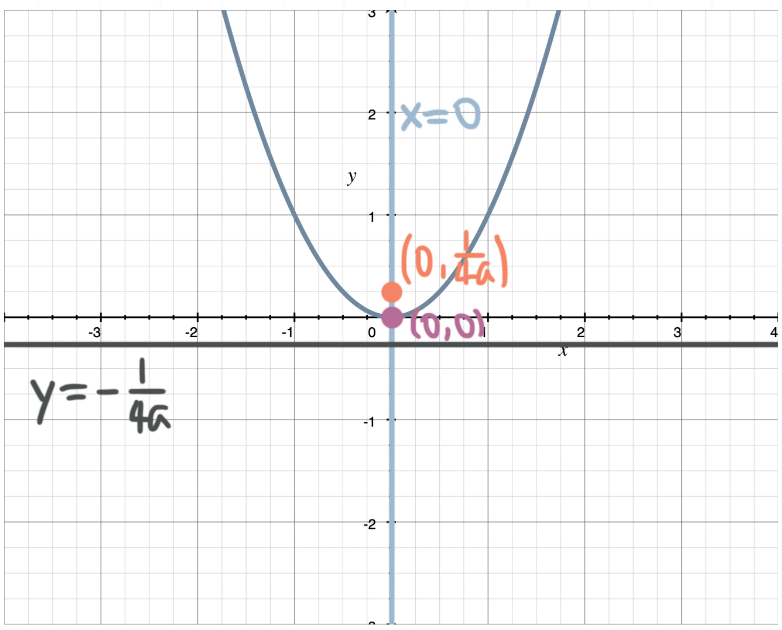
Here are sketches of each of these parabolas:

Up/down shifted: *Left/right shifted:*

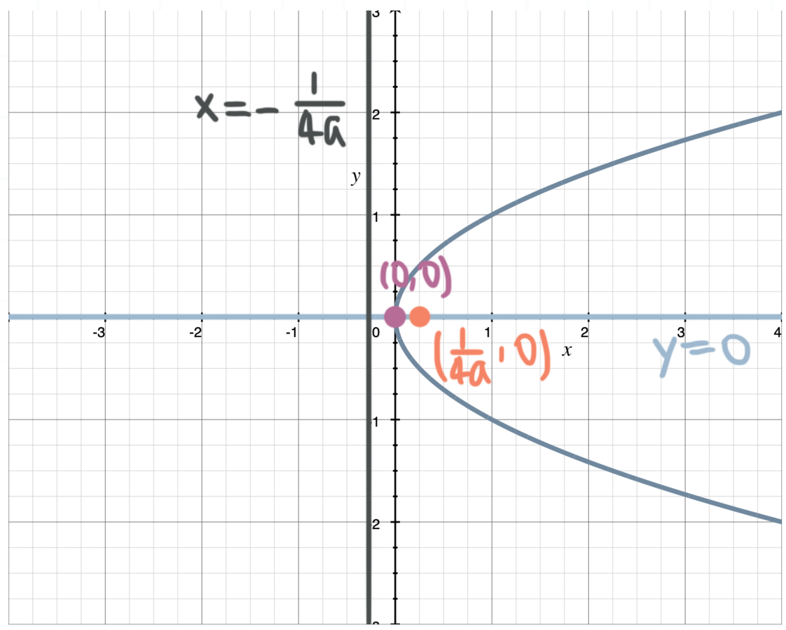




Up/down origin:



Left/right origin:



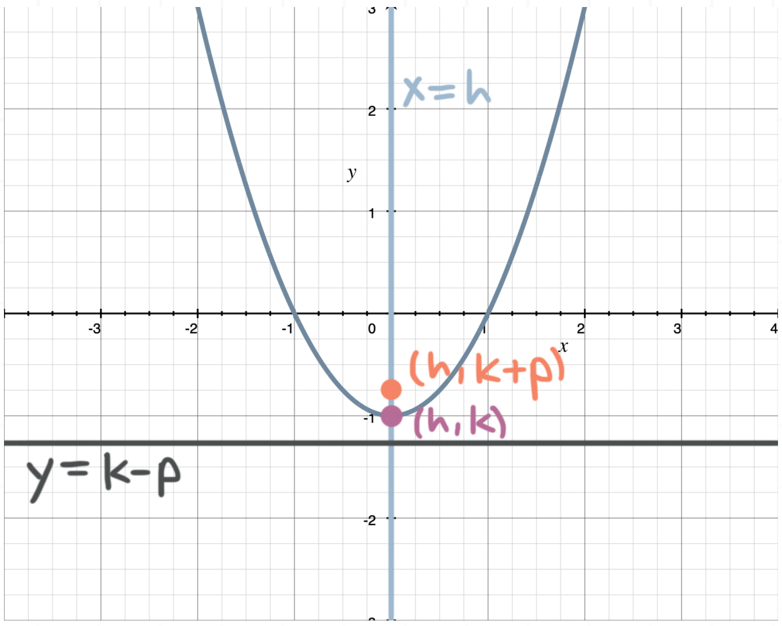
And here's everything we'll need when the equation of the parabola is in conics form.

	Equation	**Vertex**	**Axis**	**Focus**	**Directrix**
Up/down shifted	$4p(y-k)=(x-h)^2$	(h,k)	$x=h$	$(h,k+p)$	$y=k-p$
Left/right shifted	$4p(x-h)=(y-k)^2$	(h,k)	$y=k$	$(h+p,k)$	$x=h-p$
Up/down origin	$4py=x^2$	$(0,0)$	$x=0$	$(0,p)$	$y=-p$

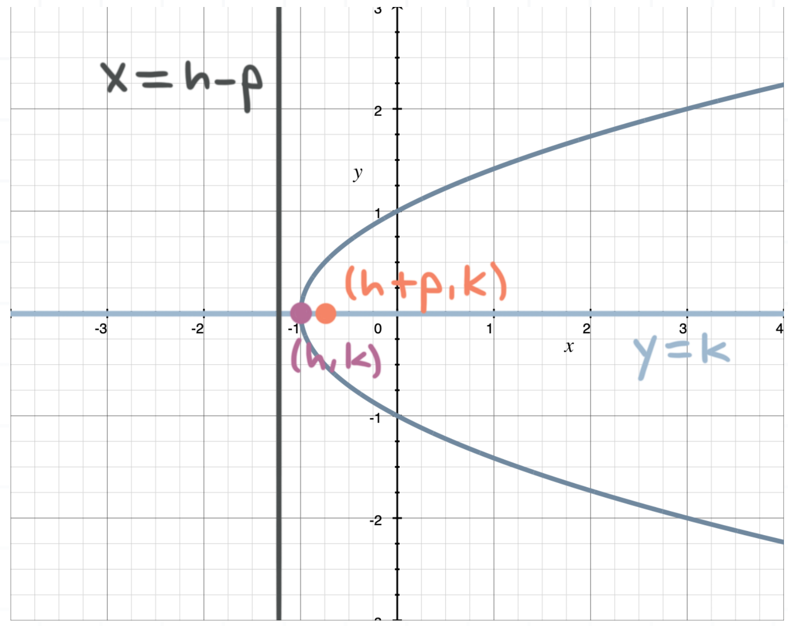
Left/right origin	$4px = y^2$	$(0,0)$	$y = 0$	$(p,0)$	$x = -p$

Here are sketches of each of these parabolas:

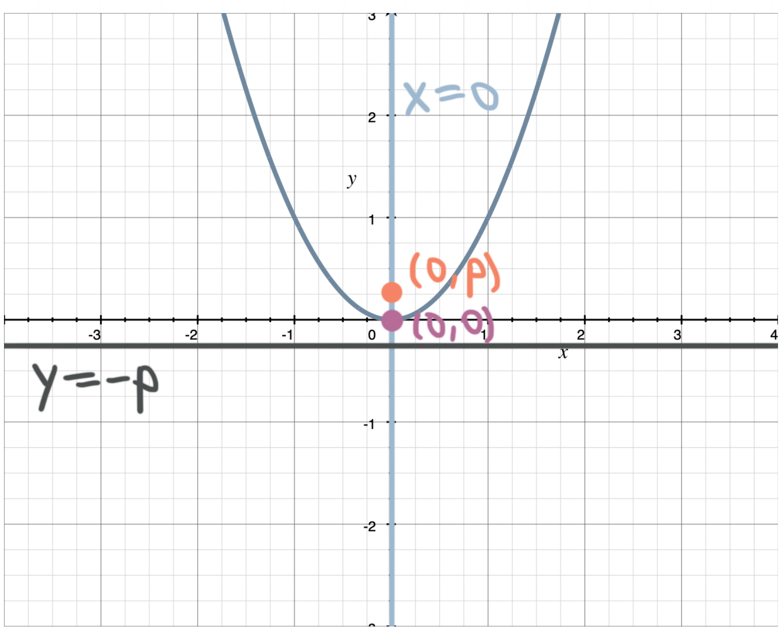
Up/down shifted:



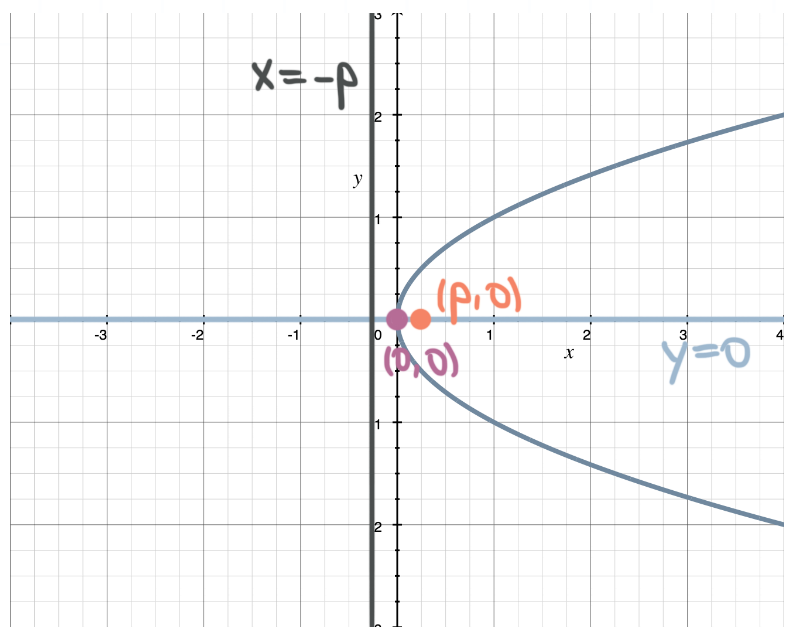
Left/right shifted:



Up/down origin:



Left/right origin:



If the equation is given in standard form, we can complete the square to put the equation in vertex form instead. For example, given the standard form equation $y = -x^2 + 2x + 5$, its vertex form is

$$-y = x^2 - 2x - 5$$

$$-y + 1 = x^2 - 2x + 1 - 5$$

$$-y + 1 = (x - 1)^2 - 5$$

$$-y = (x - 1)^2 - 6$$

$$y = -(x - 1)^2 + 6$$

With all these formulas, we can figure out almost anything we need to about the parabola. For instance, let's do an example where we find the equation of the parabola from only its focus and directrix.

Example

Find the equation of the parabola with focus $(1,3)$ and directrix $y = 1$.

From the focus $(1,3)$, we know the axis of symmetry is $x = 1$. Because the y-coordinate of the focus is $y = 3$, the directrix $y = 1$ is below the focus, which means this parabola opens up with a vertex shifted off the origin.

The formulas for the focus and directrix are simpler in conics form than they are in vertex form for an up/down shifted parabola, so we'll identify the focus as $(h, k + p)$ and the directrix as $y = k - p$. Matching these up to our focus $(1,3)$ and directrix $y = 1$, we get $h = 1$ and

$$k - p = 1$$

$$k + p = 3$$

Adding the equations together, we get $2k = 4$, or $k = 2$, which means $p = 1$. Plugging all these values into the conics form for the equation of an up/down shifted parabola, we get

$$4p(y - k) = (x - h)^2$$

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

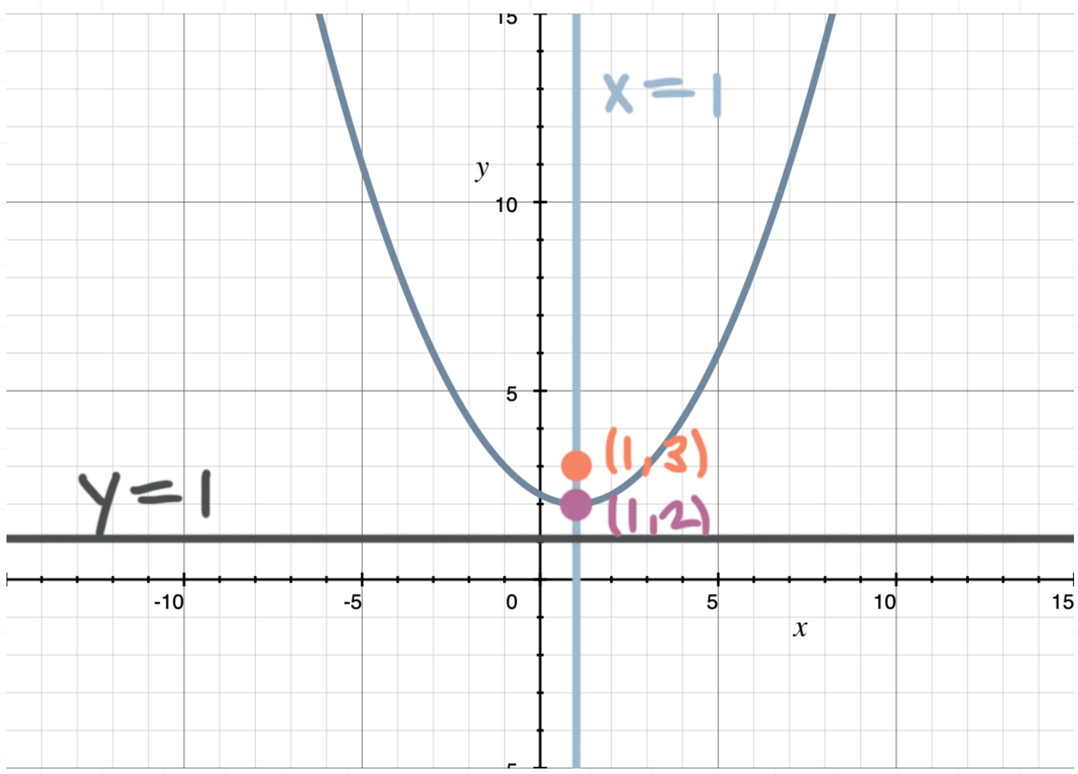
$$y - 2 = \frac{1}{4}(x - 1)^2$$

$$y = \frac{1}{4}(x - 1)^2 + 2$$

Now that we have the equation of the parabola in vertex form, we can define every piece of the parabola.

Equation	$y = a(x - h)^2 + k$	$y = \frac{1}{4}(x - 1)^2 + 2$
Vertex	(h, k)	$(1,2)$
Axis	$x = h$	$x = 1$
Focus	$\left(h, k + \frac{1}{4a}\right)$	$(1,3)$
Directrix	$y = k - \frac{1}{4a}$	$y = 1$

If we graph the parabola, we can see a sketch of all these pieces together.



Let's do another example.

Example

Find each piece of the parabola from its equation.

$$y = x^2 - 1$$

Rewrite the equation in vertex form.

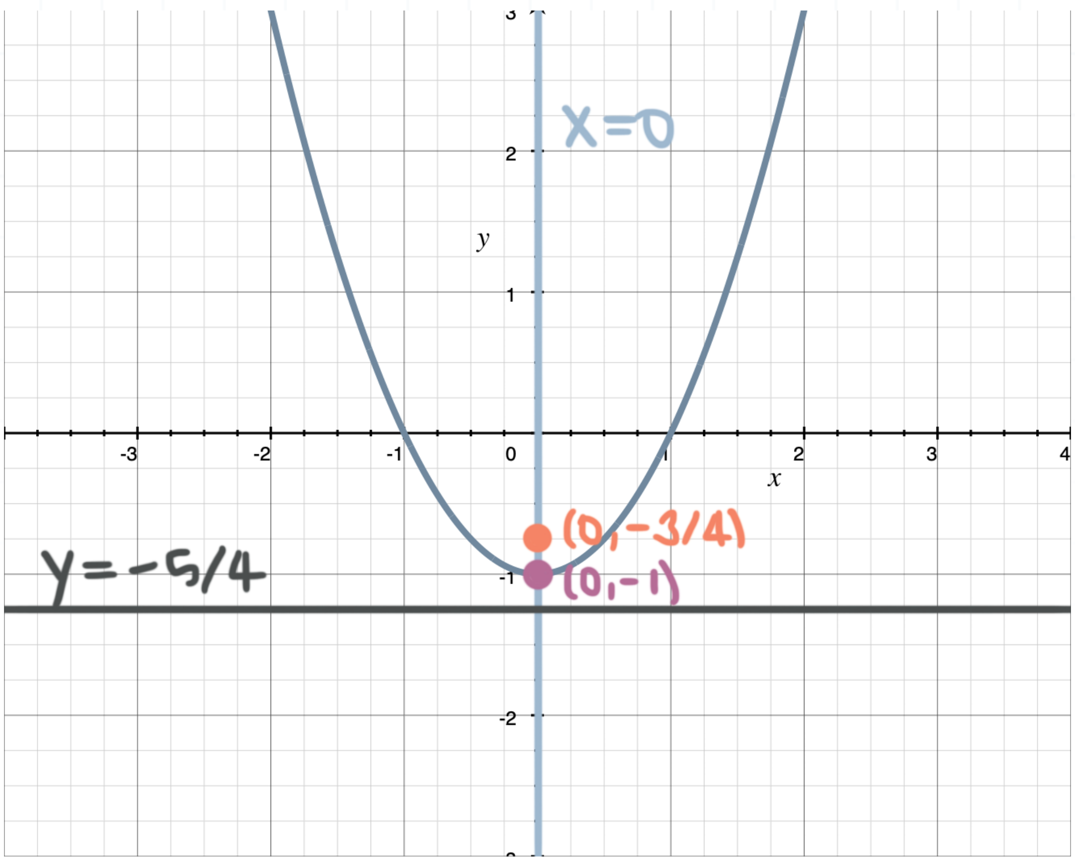
$$y = 1(x - 0)^2 - 1$$

From vertex form, we can pull all the components of the parabola.

Equation	$y = a(x - h)^2 + k$	$y = 1(x - 0)^2 - 1$
Vertex	(h, k)	$(0, -1)$

Axis	$x = h$	$x = 0$
Focus	$\left(h, k + \dfrac{1}{4a}\right)$	$\left(0, -\dfrac{3}{4}\right)$
Directrix	$y = k - \dfrac{1}{4a}$	$y = -\dfrac{5}{4}$

If we graph the parabola, we can see a sketch of all these pieces together.



Hyperbolas

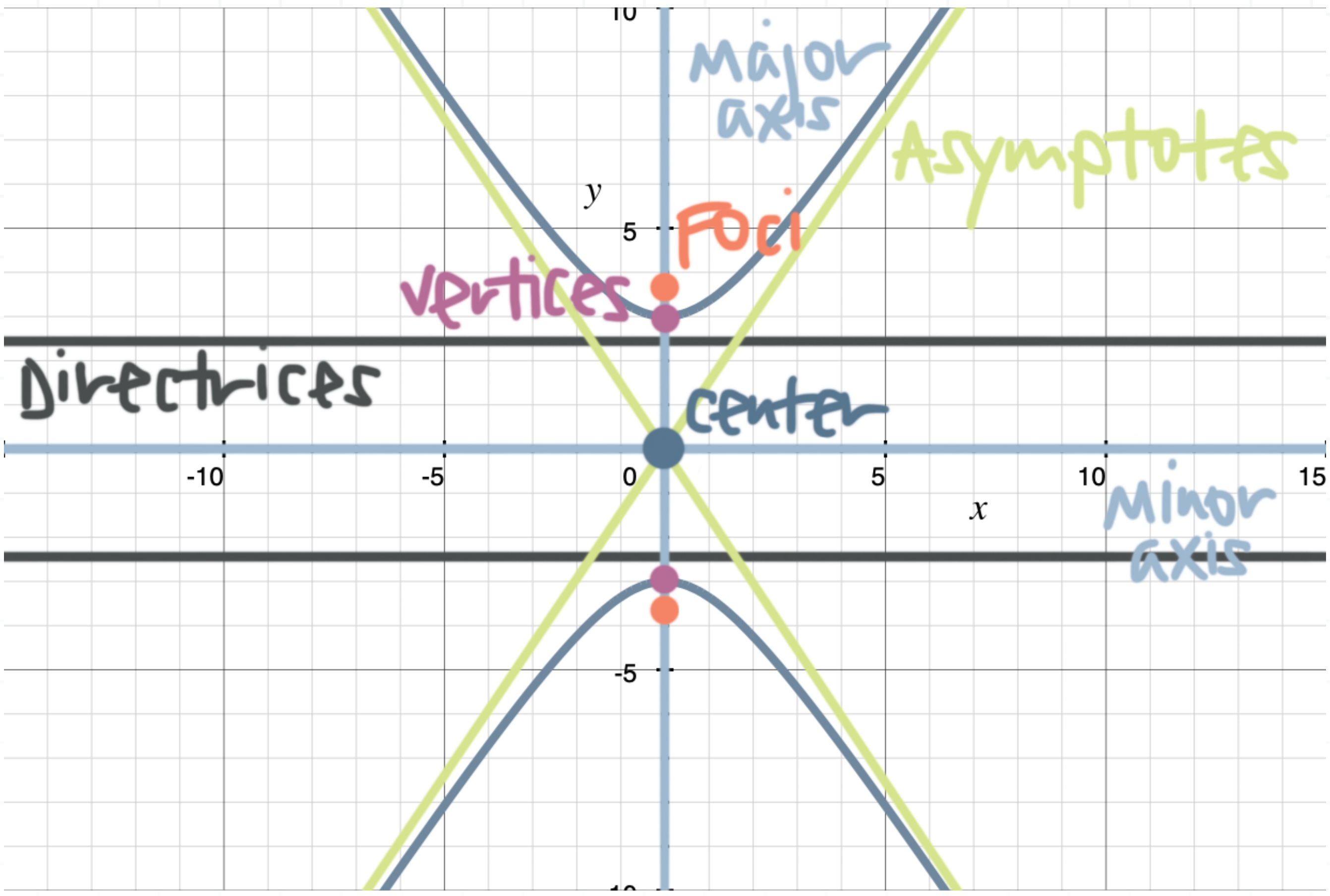
A **hyperbola** is a lot like an ellipse, except that instead of the two halves opening toward each other, the two halves open away from each other.

For that reason, similarly to the way that an ellipse was the set of points for which the sum of distances from two foci was constant, the hyperbola is the set of points for which the difference of distances from its two foci is constant.

Geometry of the hyperbola

The **foci** will be two points "inside" the hyperbola, and the **directrices** will be lines "outside" the hyperbola. The two **vertices** are the two points at which the hyperbola intersects its **major axis**, which is the axis of symmetry intersecting the two curves of the hyperbola's graph. The hyperbola's **minor axis** sits perpendicular to the major axis, and never intersects the curves of the hyperbola, instead running between them.

The **asymptotes** of the hyperbola define its boundaries, and intersect each other at the hyperbola's **center**.



And just like with ellipses, given that a is the distance between the center and the vertices, these are the standard forms for the equation of a hyperbola.

$$\frac{(x-h)^2}{a^2} - \frac{(y-k)^2}{b^2} = 1$$ Opens left and right, shifted center

$$\frac{(y-k)^2}{a^2} - \frac{(x-h)^2}{b^2} = 1$$ Opens up and down, shifted center

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1$$ Opens left and right, origin center

$$\frac{y^2}{a^2} - \frac{x^2}{b^2} = 1$$ Opens up and down, origin center

For each of these particular hyperbolas, we can define all of the pieces of the hyperbola.

	Left/right shifted	Left/right origin
Equation	$\frac{(x-h)^2}{a^2}-\frac{(y-k)^2}{b^2}=1$	$\frac{x^2}{a^2}-\frac{y^2}{b^2}=1$
Center	(h,k)	$(0,0)$
Major, minor axis	$y=k$, $x=h$	$y=0$, $x=0$
Vertices	$(h\pm a,k)$	$(\pm a,0)$
Foci, $c=\sqrt{a^2+b^2}$	$(h\pm c,k)$	$(\pm c,0)$
Asymptotes	$y=\pm\frac{b}{a}(x-h)+k$	$y=\pm\frac{b}{a}x$
Directrices	$x=h\pm\frac{a^2}{c}$	$x=\pm\frac{a^2}{c}$

	Up/down shifted	Up/down origin
Equation	$\frac{(y-k)^2}{a^2}-\frac{(x-h)^2}{b^2}=1$	$\frac{y^2}{a^2}-\frac{x^2}{b^2}=1$
Center	(h,k)	$(0,0)$
Major, minor axis	$x=h$, $y=k$	$x=0$, $y=0$
Vertices	$(h,k\pm a)$	$(0,\pm a)$
Foci, $c=\sqrt{a^2+b^2}$	$(h,k\pm c)$	$(0,\pm c)$
Asymptotes	$y=\pm\frac{a}{b}(x-h)+k$	$y=\pm\frac{a}{b}x$

Directrices	$y = k \pm \frac{a^2}{c}$	$y = \pm \frac{a^2}{c}$

Sketching the hyperbola

To sketch the graph of the hyperbola, we should start with the asymptotes, making sure they intersect at the center, which will help us define the boundaries of the curves of the hyperbola.

We can sketch in the major and minor axes. When the x^2 term is positive the hyperbola opens left and right, and when the y^2 term is positive the hyperbola opens up and down. So if we find the vertices next, we can sketch the two halves of the hyperbola, making sure they intersect the vertices and obey the asymptotes.

Lastly, we can find the foci and directrices and add those to our sketch. Let's do an example where we find and sketch all the pieces of a hyperbola.

Example

Sketch the graph of the hyperbola.

$$\frac{y^2}{9} - \frac{x^2}{4} = 1$$

Because the y^2 term is the positive term, we know the hyperbola opens up and down, that $a=\sqrt{9}=3$, and that $b=\sqrt{4}=2$. We can also see that the center is at $(h,k)=(0,0)$. So the asymptotes are

$$y=\pm\frac{a}{b}x=\pm\frac{3}{2}x$$

The major axis must be $x=0$, while the minor axis is $y=0$, and the vertices are at

$$(0,\pm a)=(0,\pm 3)$$

We can find the focal length,

$$c=\sqrt{a^2+b^2}$$

$$c=\sqrt{9+4}=\sqrt{13}$$

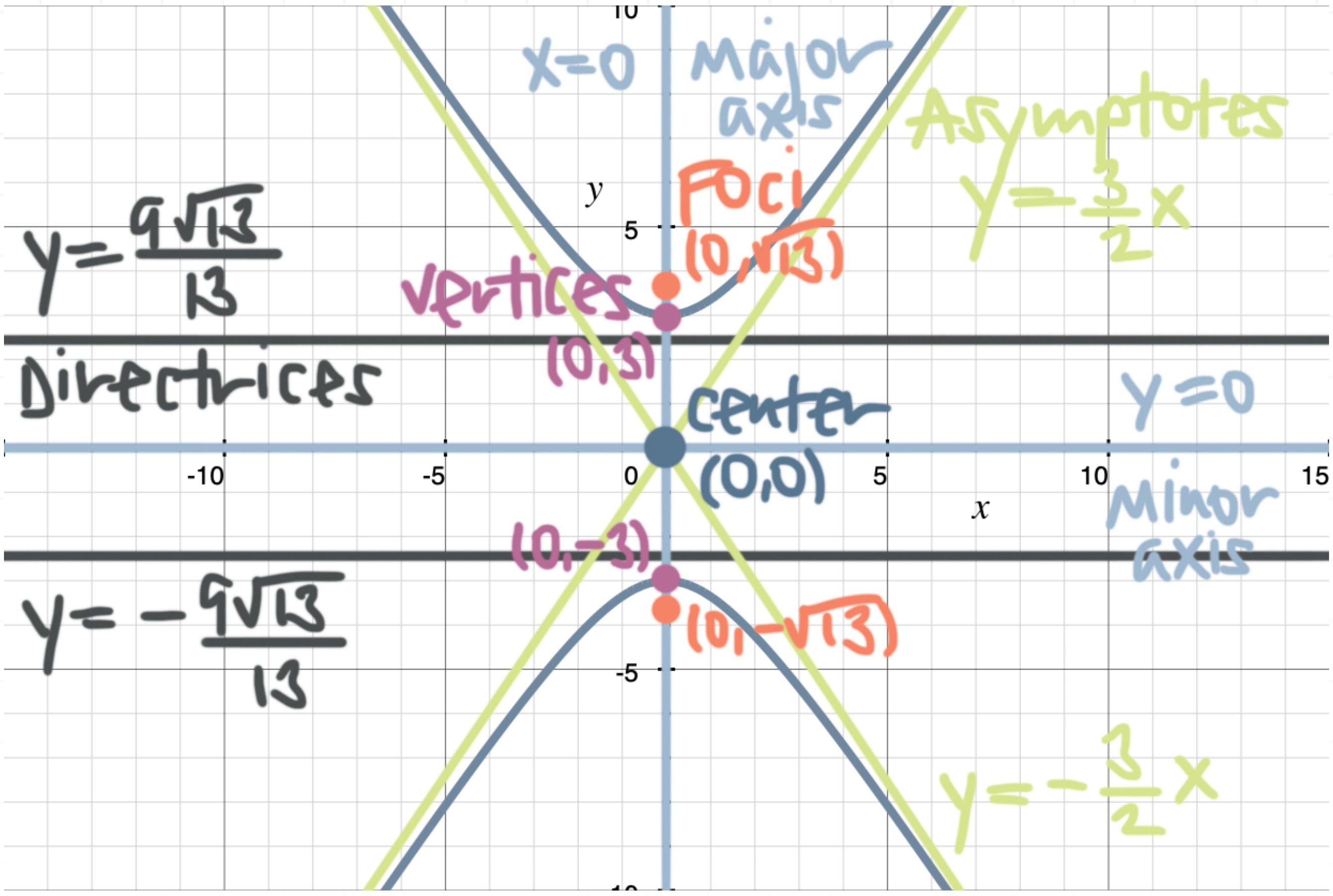
and then using that to find the foci.

$$(0,\pm c)=(0,\pm\sqrt{13})$$

Finally, the directrices are at

$$y=\pm\frac{a^2}{c}=\pm\frac{9}{\sqrt{13}}=\pm\frac{9\sqrt{13}}{13}$$

If we sketch the graph of the hyperbola and include all of these pieces of its geometry, we get


x=0
Major axis
Asymptotes
y=3/2 x
Foci
(0,√13)
y=9√13/13
vertices
(0,3)
Directrices
center
(0,0)
y=0
Minor axis
(0,−3)
(0,−√13)
y=−9√13/13
y=−3/2 x
y
x
-10
-5
0
5
10
15
5
-5

Rotating axes

We're used to seeing equations of conic sections like

$$x^2 + y^2 - 2x - 4y = 0$$

$$12x + 5y^2 + 28 = -2x^2 + 20y$$

$$4y = x^2 - 2x + 1 + 8$$

$$36y^2 - 9x^2 = 36$$

What these all have in common is that they only include some combination of x^2, y^2, x, y, and constant terms. But sometimes we'll encounter an equation of a conic section that includes a mixed xy term, like

$$64x^2 + 96xy + 36y^2 - 15x + 20y - 25 = 0$$

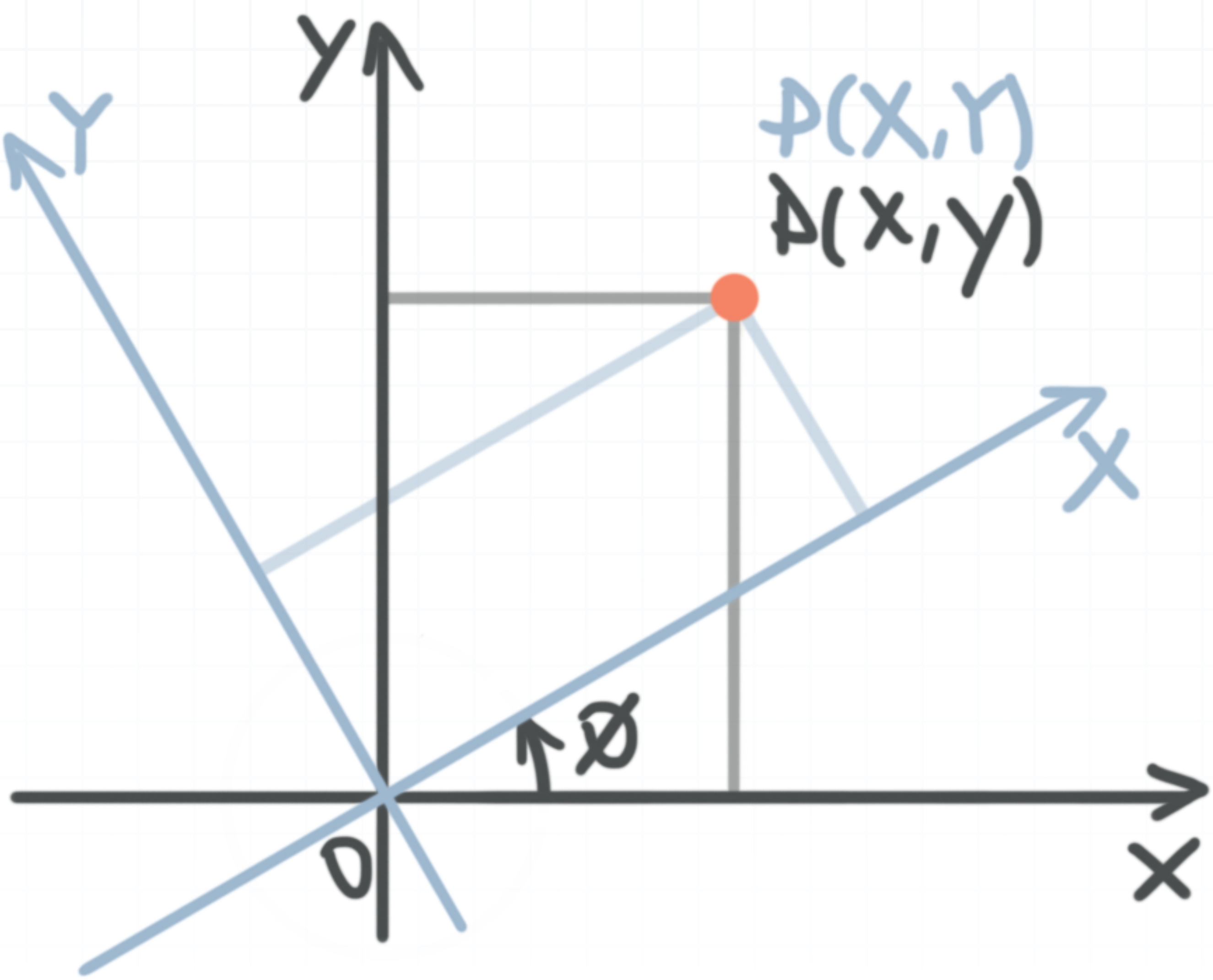
Notice the mixed $96xy$ term that's included in the equation of this parabola. We don't know yet how that $96xy$ term changes the parabola, but we do know that, if we could somehow get rid of that mixed term, that we would definitely have the equation of a parabola. In fact, that's exactly what we'll do.

Eliminating the mixed term

To remove the mixed term from the equation of a conic section in the form,

$$Ax^2 + Bxy + Cy^2 + Dx + Ey + F = 0$$

we'll rotate the regular x- and y-axes to get a new set of axes, which we'll call the X- and Y-axes.



The counterclockwise angle of rotation around the origin, ϕ, will satisfy

$$\cot(2\phi) = \frac{A - C}{B}$$

and the relationship between x and y and X and Y is given by

$$x = X\cos\phi - Y\sin\phi \qquad X = x\cos\phi + y\sin\phi$$

$$y = X\sin\phi + Y\cos\phi \qquad Y = -x\sin\phi + y\cos\phi$$

Let's do an example where we identify the form of a rotated conic, as well as its angle of rotation.

Example

Identify the conic section represented by $xy = x + y$, and determine its angle of rotation.

If we graph the equation $xy = x + y$ on the regular xy-coordinate axes, we can see that the equation represents a rotated hyperbola. But even if we couldn't sketch the graph of $xy = x + y$, we could still use rotation of axes to identify it as a hyperbola.

Start by rewriting the equation in the generic form $Ax^2 + Bxy + Cy^2 + Dx + Ey + F = 0$.

$$xy = x + y$$

$$xy - x - y = 0$$

$$0x^2 + 1xy + 0y^2 - 1x - 1y + 0 = 0$$

In this form, we can identify $A = 0$, $B = 1$, and $C = 0$, which means the angle of rotation is given by

$$\cot(2\phi) = \frac{A - C}{B}$$

$$\cot(2\phi) = \frac{0 - 0}{1} = 0$$

$$\frac{\cos(2\phi)}{\sin(2\phi)} = 0$$

This equation can only be true when

$$\cos(2\phi) = 0$$

$$2\phi = \frac{\pi}{2}, \frac{3\pi}{2}$$

$$\phi = \frac{\pi}{4}, \frac{3\pi}{4}$$

We'll choose rotation through $\phi = \pi/4 = 45^\circ$, and plug this angle into our formulas for the rotation of axes,

$$x = X\cos\phi - Y\sin\phi$$

$$y = X\sin\phi + Y\cos\phi$$

to get

$$x = X\cos\left(\frac{\pi}{4}\right) - Y\sin\left(\frac{\pi}{4}\right)$$

$$x = \frac{\sqrt{2}}{2}X - \frac{\sqrt{2}}{2}Y$$

and

$$y = X\sin\left(\frac{\pi}{4}\right) + Y\cos\left(\frac{\pi}{4}\right)$$

$$y = \frac{\sqrt{2}}{2}X + \frac{\sqrt{2}}{2}Y$$

Then we'll substitute these values of x and y into the original conic section equation.

$$xy = x + y$$

$$\left(\frac{\sqrt{2}}{2}X - \frac{\sqrt{2}}{2}Y\right)\left(\frac{\sqrt{2}}{2}X + \frac{\sqrt{2}}{2}Y\right) = \frac{\sqrt{2}}{2}X - \frac{\sqrt{2}}{2}Y + \frac{\sqrt{2}}{2}X + \frac{\sqrt{2}}{2}Y$$

$$\frac{1}{2}X^2 + \frac{1}{2}XY - \frac{1}{2}XY - \frac{1}{2}Y^2 = \sqrt{2}X$$

$$\frac{1}{2}X^2 - \sqrt{2}X - \frac{1}{2}Y^2 = 0$$

$$X^2 - 2\sqrt{2}X - Y^2 = 0$$

Complete the square on X.

$$X^2 - 2\sqrt{2}X + 2 - Y^2 = 0 + 2$$

$$(X - \sqrt{2})^2 - Y^2 = 2$$

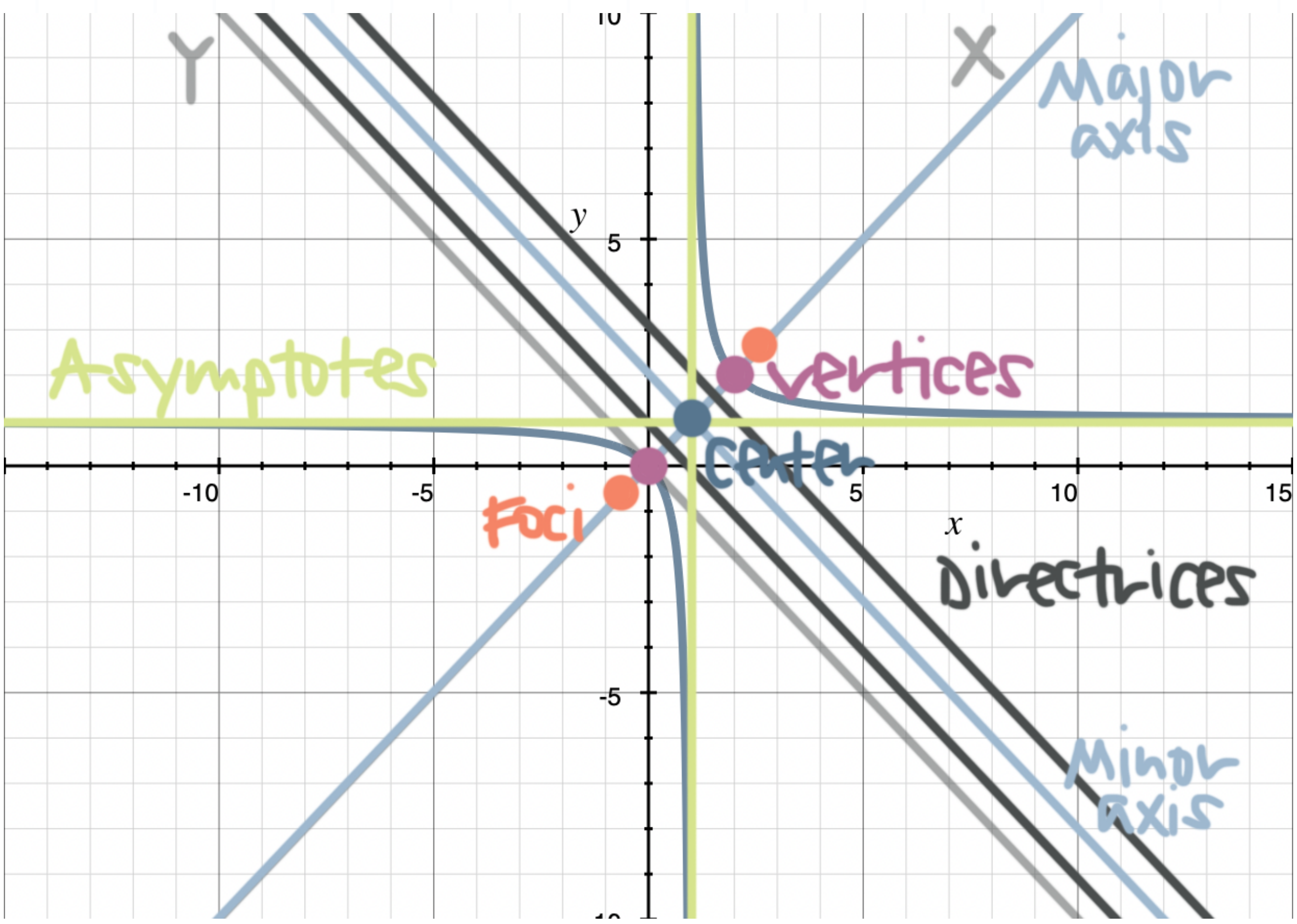
$$\frac{(X - \sqrt{2})^2}{2} - \frac{Y^2}{2} = 1$$

This equation in X and Y represents a shifted hyperbola that opens left and right in the XY-coordinate system. Which means we can identify all of its component parts using the table of formulas we covered earlier for left-right shifted hyperbolas.

Equation	$\frac{(X - h)^2}{a^2} - \frac{(Y - k)^2}{b^2} = 1$	$\frac{(X - \sqrt{2})^2}{2} - \frac{Y^2}{2} = 1$

Center	$(X, Y) = (h, k)$	$(X, Y) = (\sqrt{2}, 0)$
Major, minor axis	$Y = k,\ X = h$	$Y = 0,\ X = \sqrt{2}$
Vertices	$(X, Y) = (h \pm a, k)$	$(X, Y) = (0,0),\ (2\sqrt{2}, 0)$
Foci, $c = \sqrt{a^2 + b^2}$	$(X, Y) = (h \pm c, k)$	$(X, Y) = (\sqrt{2} \pm 2, 0)$
Asymptotes	$Y = \pm \frac{b}{a} X$	$Y = \pm X$
Directrices	$X = h \pm \frac{a^2}{c}$	$X = \sqrt{2} \pm 1$

Using this information, we can sketch the hyperbola.



We can see the original x- and y- axes, the X- and Y-axes that are rotated $\phi = \pi/4 = 45^\circ$ from the original x- and y-axes, as well as all the hyperbola's component pieces.

Half-angle identities

We'll very often need to use the half-angle identities from trigonometry in order to find the values of sine and cosine at the angle of rotation ϕ.

$$\cos\phi = \sqrt{\frac{1+\cos(2\phi)}{2}} \qquad \sin\phi = \sqrt{\frac{1-\cos(2\phi)}{2}}$$

Let's work through an example where we use these half angle formulas.

Example

Identify the conic section and determine its angle of rotation.

$$9x^2 - 24xy + 16y^2 = 100(x - y - 1)$$

Rewrite the equation in standard form,

$$9x^2 - 24xy + 16y^2 = 100x - 100y - 100$$

$$9x^2 - 24xy + 16y^2 - 100x + 100y + 100 = 0$$

then identify $A = 9$, $B = -24$, and $C = 16$ and find the angle of rotation.

$$\cot(2\phi) = \frac{A - C}{B}$$

$$\cot(2\phi) = \frac{9 - 16}{-24}$$

$$\cot(2\phi) = \frac{7}{24}$$

Remember that cotangent is equivalent to adjacent/opposite, so the length of the adjacent leg is 7 and the length of the opposite leg is 24, and the length of the hypotenuse of the right triangle must be

$$a^2 + b^2 = c^2$$

$$7^2 + 24^2 = c^2$$

$$49 + 576 = c^2$$

$$625 = c^2$$

$$c = 25$$

We need the value of cosine for the half-angle formulas, and cosine is equivalent to adjacent/hypotenuse, so $\cos(2\phi) = 7/25$. The half-angle formulas give

$$\cos\phi = \sqrt{\frac{1 + \cos(2\phi)}{2}} = \sqrt{\frac{1 + \frac{7}{25}}{2}} = \sqrt{\frac{\frac{32}{25}}{2}} = \sqrt{\frac{16}{25}} = \frac{4}{5}$$

$$\sin\phi = \sqrt{\frac{1 - \cos(2\phi)}{2}} = \sqrt{\frac{1 - \frac{7}{25}}{2}} = \sqrt{\frac{\frac{18}{25}}{2}} = \sqrt{\frac{9}{25}} = \frac{3}{5}$$

Then the rotation of axes formulas give

$$x = X\cos\phi - Y\sin\phi$$

$$x = \frac{4}{5}X - \frac{3}{5}Y$$

and

$$y = X\sin\phi + Y\cos\phi$$

$$y = \frac{3}{5}X + \frac{4}{5}Y$$

Substitute these new values for x and y back into the original equation.

$$9x^2 - 24xy + 16y^2 - 100x + 100y + 100 = 0$$

$$9\left(\frac{4}{5}X - \frac{3}{5}Y\right)^2 - 24\left(\frac{4}{5}X - \frac{3}{5}Y\right)\left(\frac{3}{5}X + \frac{4}{5}Y\right) + 16\left(\frac{3}{5}X + \frac{4}{5}Y\right)^2$$

$$-100\left(\frac{4}{5}X - \frac{3}{5}Y\right) + 100\left(\frac{3}{5}X + \frac{4}{5}Y\right) + 100 = 0$$

$$\frac{144}{25}X^2 - \frac{288}{25}X^2 + \frac{144}{25}X^2 - \frac{216}{25}XY - \frac{168}{25}XY + \frac{384}{25}XY$$

$$+\frac{81}{25}Y^2 + \frac{288}{25}Y^2 + \frac{256}{25}Y^2$$

$$-\frac{400}{5}X + \frac{300}{5}X + \frac{300}{5}Y + \frac{400}{5}Y + 100 = 0$$

$$25Y^2 - 20X + 140Y + 100 = 0$$

$$5Y^2 - 4X + 28Y + 20 = 0$$

Rewrite the equation, then complete the square on Y.

$$Y^2 + \frac{28}{5}Y - \frac{4}{5}X + 4 = 0$$

$$Y^2 + \frac{28}{5}Y + \frac{784}{100} - \frac{4}{5}X + 4 = 0 + \frac{784}{100}$$

$$\left(Y + \frac{28}{10}\right)^2 - \frac{4}{5}X + 4 = \frac{784}{100}$$

$$\left(Y + \frac{14}{5}\right)^2 = \frac{4}{5}\left(X + \frac{24}{5}\right)$$

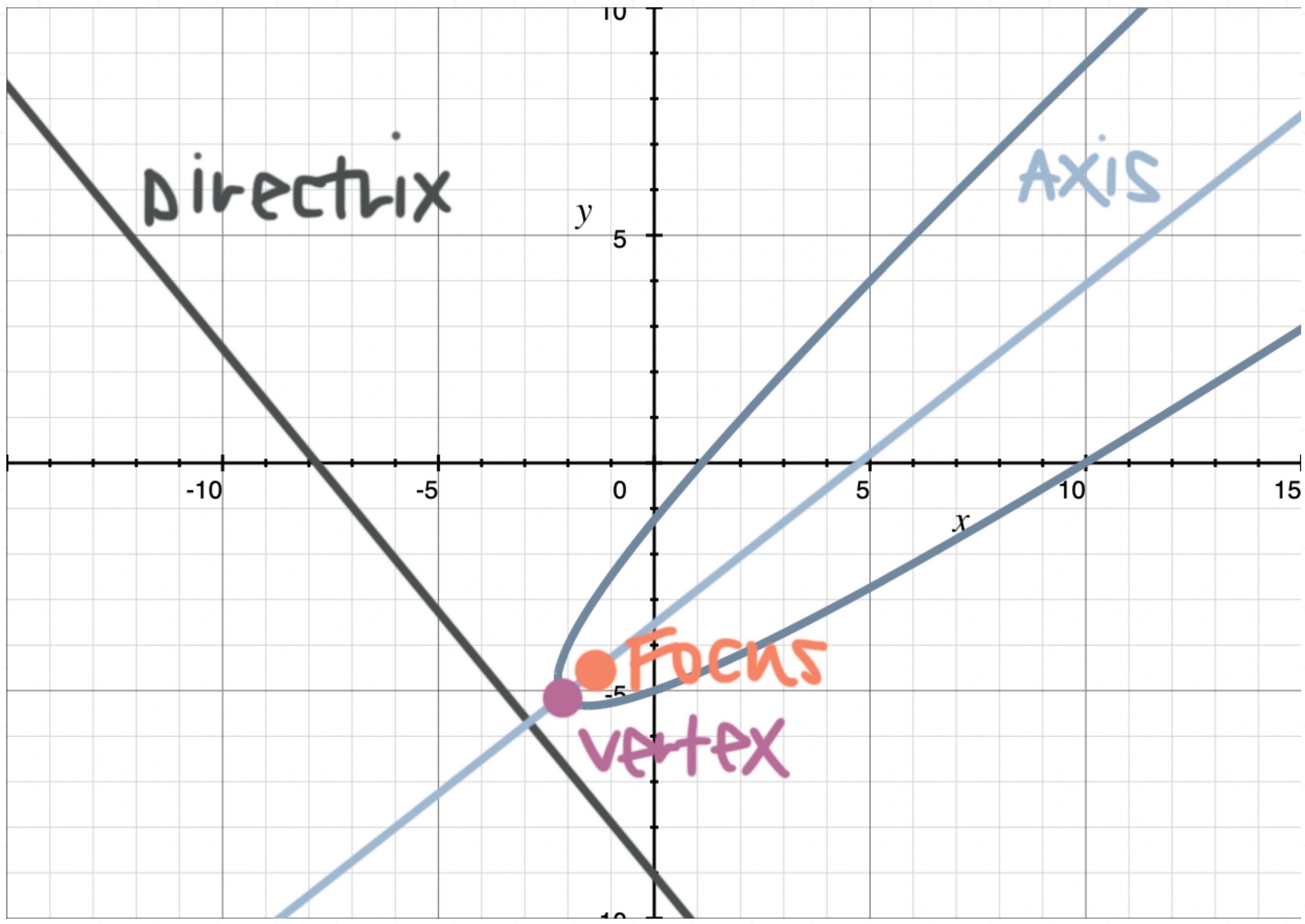
$$\left(Y + \frac{14}{5}\right)^2 = 4\left(\frac{1}{5}\right)\left(X + \frac{24}{5}\right)$$

If we match this equation to the conics form of a shifted parabola that opens to the left or right, $4p(x - h) = (y - k)^2$, then we can say that the parabola has its vertex at $(X, Y) = (-24/5, -14/5)$ and its axis at $Y = -14/5$.

Knowing that $\cos\phi = 4/5$, we can say that the angle of rotation is

$$\phi = \cos^{-1}\left(\frac{4}{5}\right) \approx 37^\circ$$

Using this information, we can sketch the parabola.



We can see the original x- and y- axes, the X- and Y-axes that are rotated $\phi \approx 37^\circ$ from the original x- and y-axes, as well as all the parabola's component pieces.

Identifying conics with the discriminant

We actually have a way, other than eliminating the mixed xy term, to identify the conic shape represented by the general equation

$$Ax^2 + Bxy + Cy^2 + Dx + Ey + F = 0$$

With a conic equation in this form, we can identify $B^2 - 4AC$, which is the **discriminant** of the equation. Based on the value of the discriminant, we know

- the conic section is a parabola when $B^2 - 4AC = 0$

- the conic section is an ellipse when $B^2 - 4AC < 0$

- the conic section is a hyperbola when $B^2 - 4AC > 0$

For instance, using the equation $xy = x + y$ we were looking at in the previous example, we can rewrite it in standard form,

$$0x^2 + 1xy + 0y^2 - 1x - 1y + 0 = 0$$

and then calculate $B^2 - 4AC$,

$$B^2 - 4AC = 1^2 - 4(0)(0)$$

$$B^2 - 4AC = 1$$

Because the discriminant is $1 > 0$, we can quickly see that the conic section is a hyperbola. Let's do one more example with the discriminant.

Example

Use the discriminant to determine the shape of the conic.

$$9x^2 - 24xy + 16y^2 = 100(x - y - 1)$$

Rewrite the equation so that it matches the standard form $Ax^2 + Bxy + Cy^2 + Dx + Ey + F = 0$.

$$9x^2 - 24xy + 16y^2 = 100x - 100y - 100$$

$$9x^2 - 24xy + 16y^2 - 100x + 100y + 100 = 0$$

From standard form, we can identify $A = 9$, $B = -24$, and $C = 16$, so the value of the discriminant is

$$B^2 - 4AC = (-24)^2 - 4(9)(16)$$

$$B^2 - 4AC = 576 - 576$$

$$B^2 - 4AC = 0$$

Because the discriminant is 0, we know the conic is a parabola.

Polar equations of conics

From studying parabolas, ellipses, and hyperbolas, we know that a parabola is partly defined using its focus, while ellipses and hyperbolas are defined by their two foci.

Eccentricity

But we can actually define all three of these conic sections in terms of their foci. If we say that F is the focus, l is the directrix, and e is the **eccentricity**, a positive constant, then the distance of a point on the conic to the focus, divided by the distance of that same point to the directrix, is always equal to the eccentricity.

$$\frac{d(P,F)}{d(P,l)} = e$$

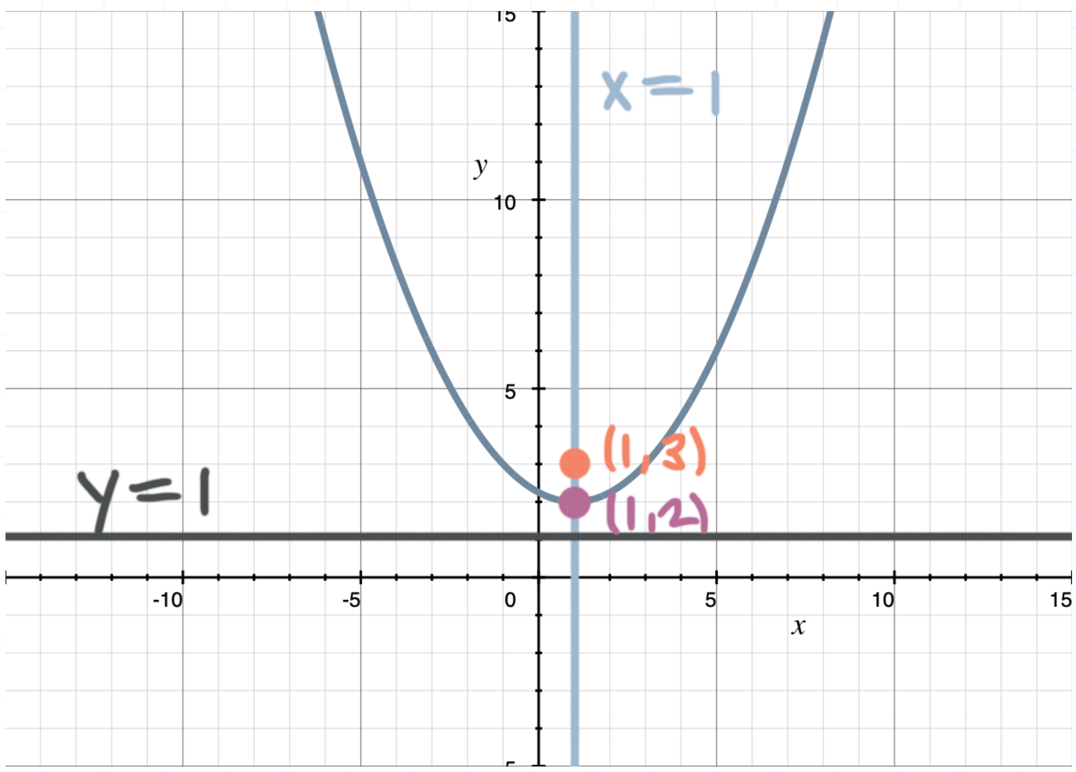
This is true for all points on all conic sections. The conic is

- a parabola when $e = 1$
- an ellipse when $e < 1$
- a hyperbola when $e > 1$

In other words, if we know one point on the conic (like a vertex), we can find the distance from that point to the focus, and from that point to the directrix, and then calculate the eccentricity as the quotient of those distances.

Example

Find the eccentricity of the conic section.



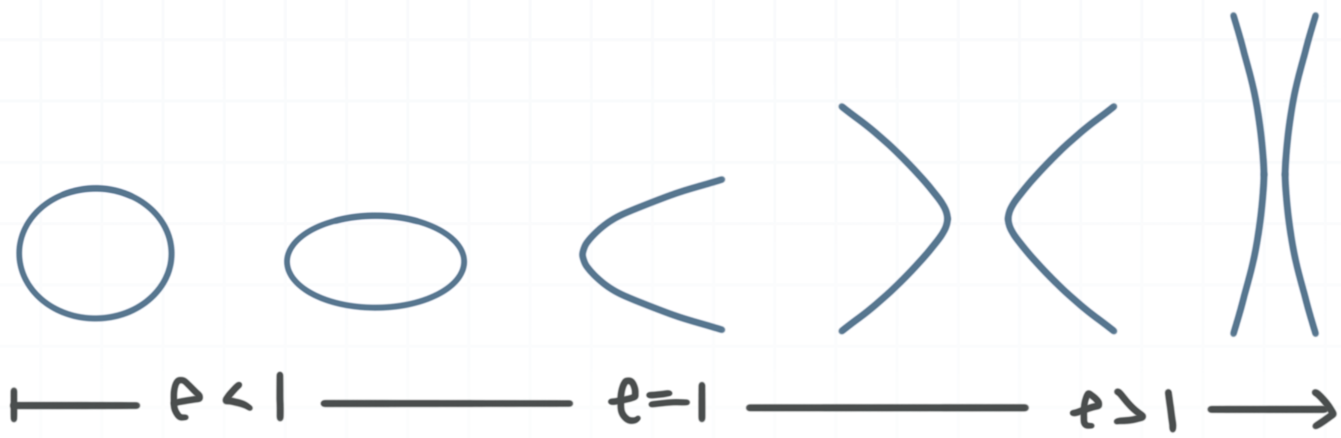
First, let's say that we know immediately that the eccentricity is $e = 1$, because every parabola, by definition, has this eccentricity.

But we can also prove this by calculating the distance between the parabola's vertex at $(1,2)$ and its focus at $(1,3)$ and directrix at $y = 1$. The distance from the vertex at $(1,2)$ to the focus at $(1,3)$ is 1. And the distance from the vertex at $(1,2)$ to the directrix at $y = 1$ is 1.

So the eccentricity of the parabola is

$$e = \frac{d(P, F)}{d(P, l)} = \frac{1}{1} = 1$$

Think about eccentricity as the amount of stretch in the conic section. When eccentricity is positive but very close to $e = 0$, the section is an ellipse, but with very little stretch, such that the ellipse is almost a perfect circle. As eccentricity increases, the ellipse becomes more and more stretched, and we see the most extreme stretch in the ellipse for eccentricity values that are just less than $e = 1$.



As soon as the eccentricity reaches $e = 1$ exactly, the conic section is a parabola. But it's only a parabola for an instance, because as soon as eccentricity becomes greater than 1, the conic section becomes a hyperbola. As the eccentricity becomes larger and larger above $e = 1$, the hyperbola becomes more and more stretched and skewed.

Polar form

Now that we understand how the eccentricity affects the shape of the conic section, we can write conic sections in polar coordinates, assuming we know the value of the eccentricity and the equation of a directrix. When we use these polar equations, we know that one focus is positioned exactly at the origin.

$$r = \frac{ed}{1 + e\cos\theta} \qquad r = \frac{ed}{1 - e\cos\theta} \qquad r = \frac{ed}{1 + e\sin\theta} \qquad r = \frac{ed}{1 - e\sin\theta}$$

$$x = d \qquad x = -d \qquad y = d \qquad y = -d$$

For instance, given a directrix at $x = -2$ and an eccentricity of $1/2$, the polar equation of the conic section is

$$r = \frac{ed}{1 - e\cos\theta}$$

$$r = \frac{\frac{1}{2}(2)}{1 - \frac{1}{2}\cos\theta}$$

$$r = \frac{1}{1 - \frac{1}{2}\cos\theta}$$

Multiplying by $2/2$ to clear the fraction, we get

$$r = \frac{2}{2 - \cos\theta}$$

Rotating conics in polar coordinates

Rotation of conics is easier in polar coordinates than it is in rectangular coordinates, which is one of the reasons why it's helpful to sometimes convert conic section equations from rectangular to polar coordinates.

If we want to rotate any of the four polar equations

$$r = \frac{ed}{1 \pm e\cos\theta}$$

$$r = \frac{ed}{1 \pm e \sin \theta}$$

by some angle α, all we have to do is subtract α from θ, and we'll have the equation of the rotated conic section.

$$r = \frac{ed}{1 \pm e \cos(\theta - \alpha)}$$

$$r = \frac{ed}{1 \pm e \sin(\theta - \alpha)}$$

Remember that conic sections in this form have one focus at the origin. When we rotate by α, the conic section will rotate around that focus, as opposed to rotating around the other focus or around some other point on the conic section.

Example

Find the equation of the conic section that has eccentricity $e = 3/4$, directrix $y = 3$, and is rotated by $\alpha = \pi/4$.

Given the directrix $y = 3$, The polar equation will be

$$r = \frac{ed}{1 + e \sin(\theta - \alpha)}$$

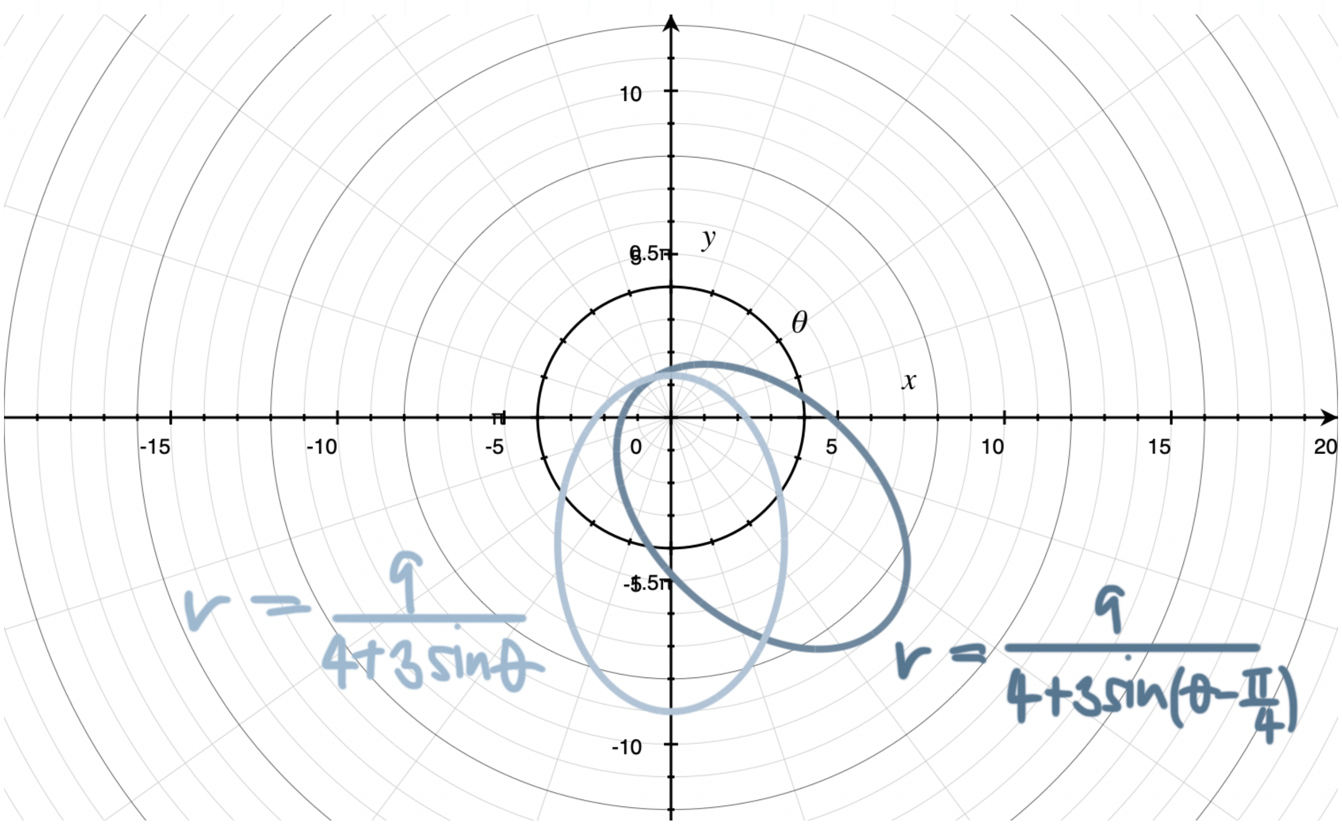
$$r = \frac{\frac{3}{4}(3)}{1 + \frac{3}{4} \sin\left(\theta - \frac{\pi}{4}\right)}$$

$$r = \frac{\frac{9}{4}}{1 + \frac{3}{4}\sin\left(\theta - \frac{\pi}{4}\right)}$$

Multiplying by $4/4$ to clear the fractions, we get

$$r = \frac{9}{4 + 3\sin\left(\theta - \frac{\pi}{4}\right)}$$

If we sketch the graph of this equation without the rotation, and the graph of the equation with the rotation, we see the rotation around the focus at the origin.



The graph in light blue is the ellipse without rotation, and the graph in dark blue is the ellipse after the rotation of $\alpha = \pi/4$.

Parametric curves and eliminating the parameter

At this point, we've learned how to describe curves in both rectangular and polar coordinates. In rectangular coordinates we usually defined the curve for y in terms of x, and in polar coordinates we usually defined the curve for r in terms of θ.

Introducing the parameter

But this isn't the only way to define a curve in the plane. We can also imagine how a curve in the rectangular plane has x and y values that change over time. Put another way, as we trace along a curve in the rectangular coordinate system, the value of y changes as we trace, and so does the value of x. So we could describe the changes in both y and x in terms of time t, which we'll call the **parameter**.

Therefore, $x = f(t)$ and $y = g(t)$ are a set of **parametric equations** in terms of the parameter t that describe the path of the plane curve.

Example

Sketch the curve defined by the parametric equations.

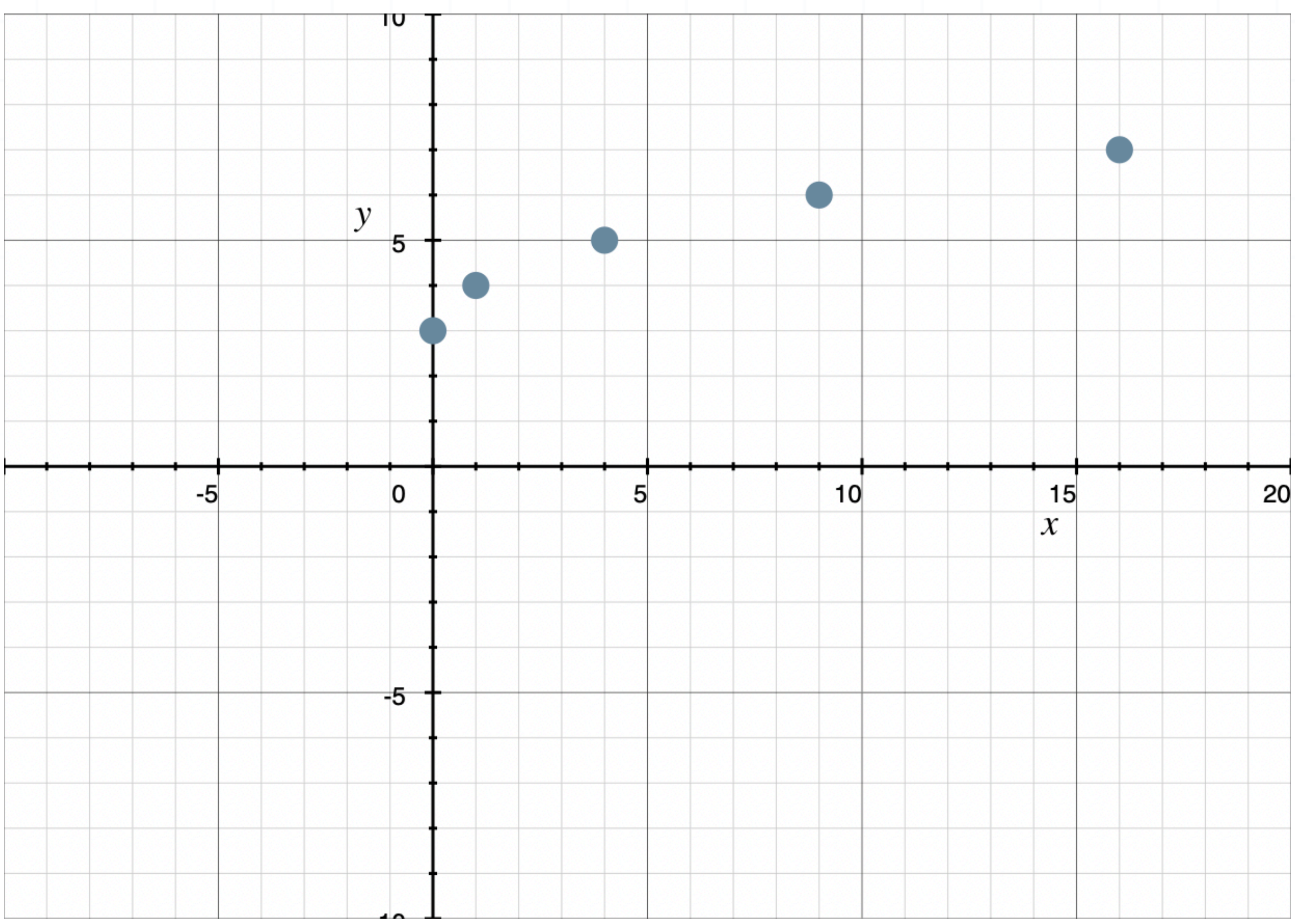
$$x = t^2$$

$$y = t + 3$$

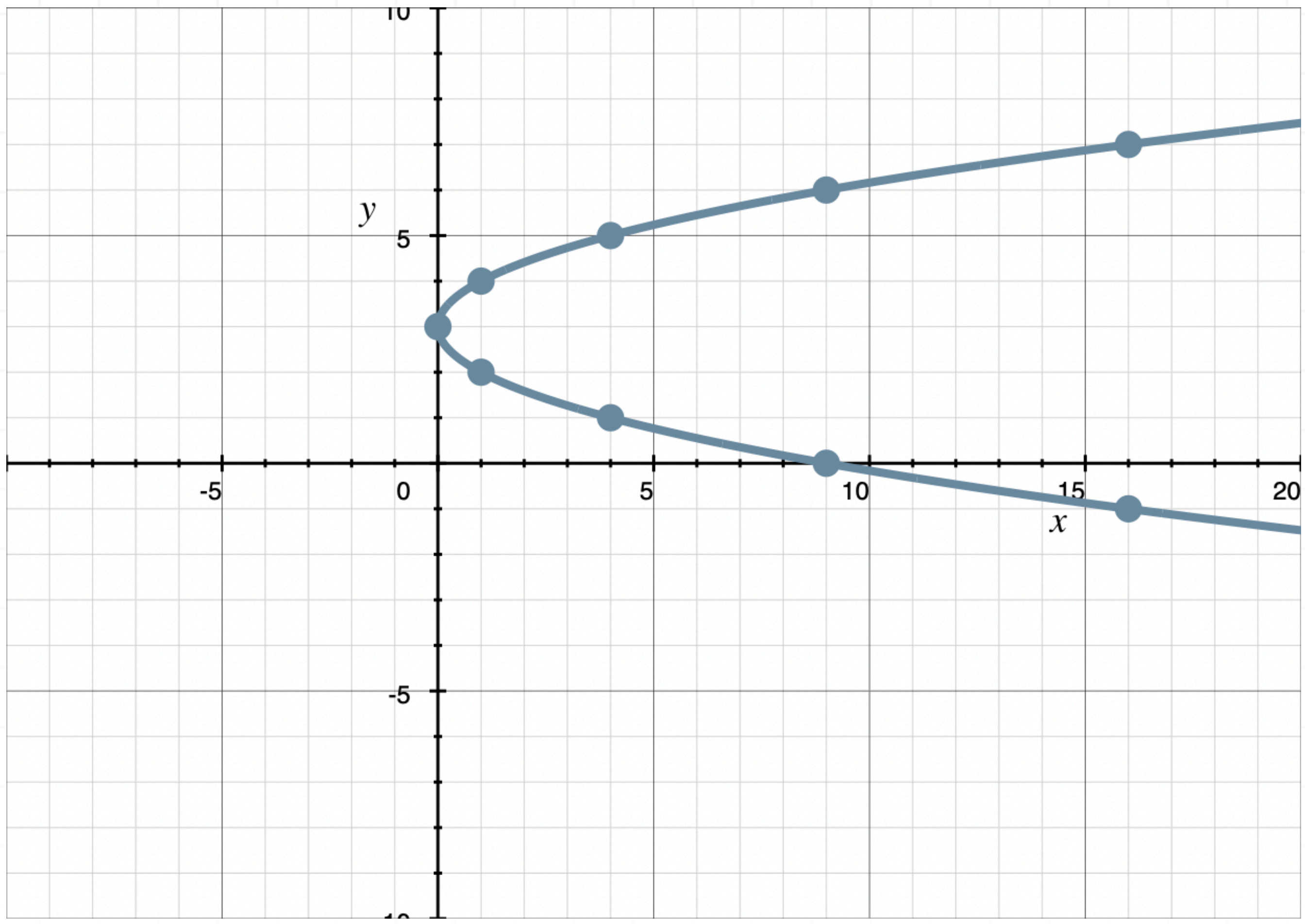
Let's explore what the curve looks like for values of the parameter starting at $t = 0$. If we plug integer values of t into the parametric equations for both x and y, we get

t	0	1	2	3	4
x	0	1	4	9	16
y	3	4	5	6	7

If we plot these (x, y) coordinate points, we see the shape of the curve starting to emerge.



If we include negative integer values of t as well, and connect the points with a smooth curve, the graph of the parametric equations is



Eliminating the parameter

To convert a parametric equation into a rectangular equation, we need to eliminate the parameter t. We can do this by substituting one parametric equation into the other, or by solving both equations for the parameter and then setting the curves equal to each other.

For instance, let's consider the parametric equations $x = t^2$ and $y = t + 3$ from the earlier example. If we solve $y = t + 3$ for t, we get $t = y - 3$. Then we can substitute $t = y - 3$ into $x = t^2$.

$$x = t^2$$

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

Now we have a rectangular equation in terms of x and y only, and the parameter t has been eliminated. If we sketch $x = (y-3)^2$, we'll see that it's an identical curve to the one we found when we sketched $x = t^2$ and $y = t + 3$ as a set of parametric equations.

Sometimes we can even eliminate the parameter by substituting the parametric equations into some third, related equation.

Example

Eliminate the parameter.

$$x = \cos t$$

$$y = \sin t$$

To solve either parametric equation for t would require us to introduce an inverse trig function, which gets a little messy. Instead, let's realize that we can use the Pythagorean identity $\sin^2 t + \cos^2 t = 1$.

If we square both parametric equations, they become $x^2 = \cos^2 t$ and $y^2 = \sin^2 t$. Now that the equations are in this form, we can make substitutions into the Pythagorean identity.

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

$$y^2 + x^2 = 1$$

$$x^2 + y^2 = 1$$

This is the equation of the circle centered at the origin with radius 1.

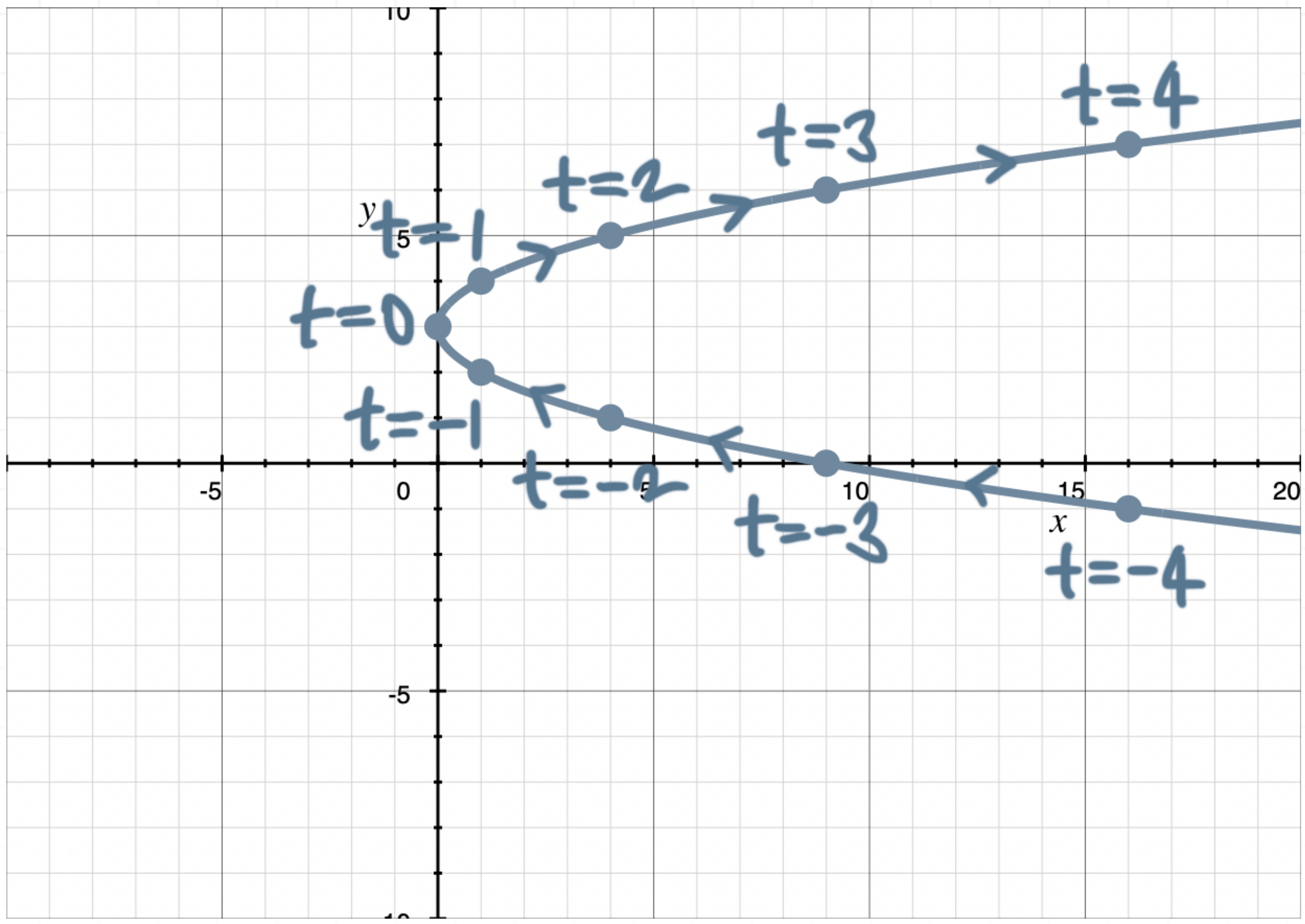
Direction of the parameter

In the last lesson, we showed how to convert a set of parametric equations $x = f(t)$ and $y = g(t)$ into a rectangular equation $y = f(x)$ by eliminating the parameter and rewriting the equation in terms of only x and y, without t.

The fact that we can convert like this between parametric and rectangular equations might trick us into thinking that the two equation types are fully equivalent.

And while they're equivalent in terms of the shape of the curve, the parametric equations actually come with more information than the equivalent rectangular equation, which is the direction, or orientation, of the parameter's value.

For instance, we know from a previous example that the parametric equations $x = t^2$ and $y = t + 3$ are equivalent to the rectangular equation $x = (y - 3)^2$. But if we attach the values of t to the corresponding points on the graph, we can identify the direction of increasing t, and indicate that direction with small arrows along the curve.



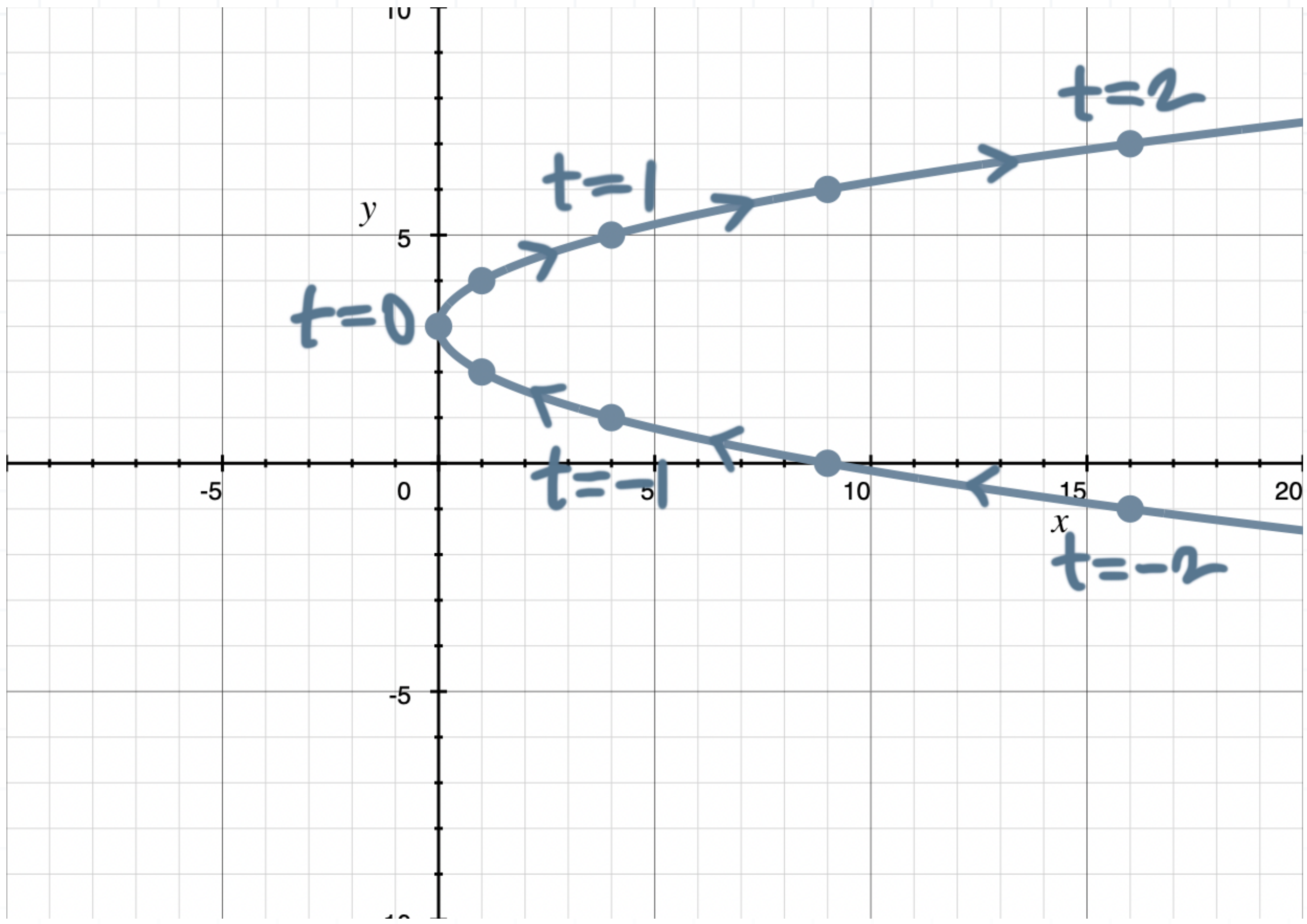
So expressing the curve as a set of parametric equations, instead of as a rectangular equation, gives us a way to include extra information about the curve.

The most obvious example is to think about the parameter t as representing time. Doing so allows us to replace t with $2t$,

$$x = t^2 \quad \rightarrow \quad x = (2t)^2 = 4t^2$$

$$y = t + 3 \quad \rightarrow \quad y = 2t + 3$$

to trace out the curve twice as fast,

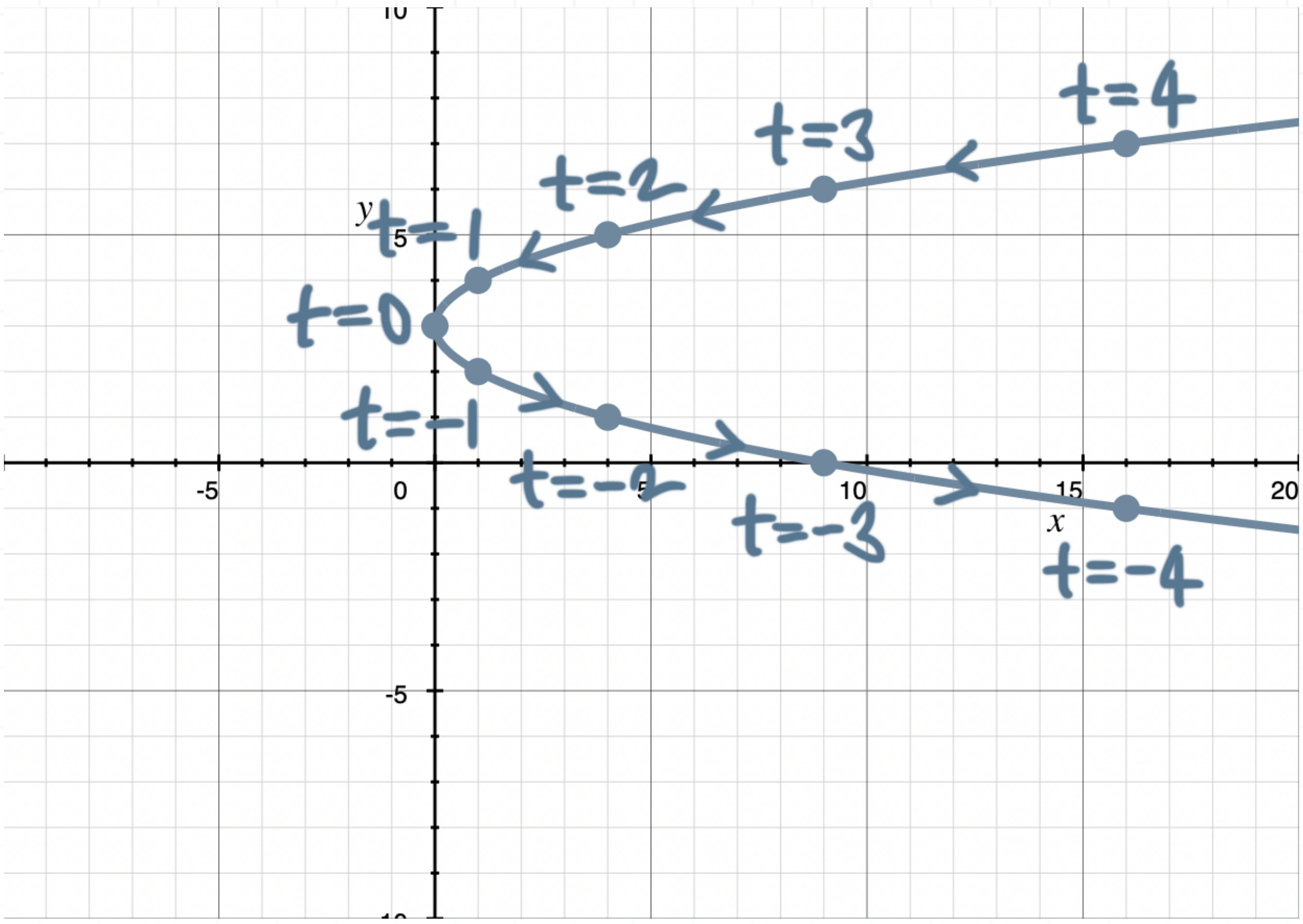


or to replace t with $-t$,

$$x = t^2 \qquad \to \qquad x = (-t)^2 = t^2$$

$$y = t + 3 \qquad \to \qquad y = -t + 3$$

to trace out the curve in the opposite direction.



Let's do an example where we identify the direction of increasing t.

Example

Sketch the graph of the parametric curve and indicate the direction of increasing t.

$$x = \frac{t}{2}$$

$$y = \frac{3t}{4}$$

Let's solve both equations for t,

$$2x = t$$

$$\frac{4y}{3} = t$$

so that we can set these two values of t equal to each other.

$$\frac{4y}{3} = 2x$$

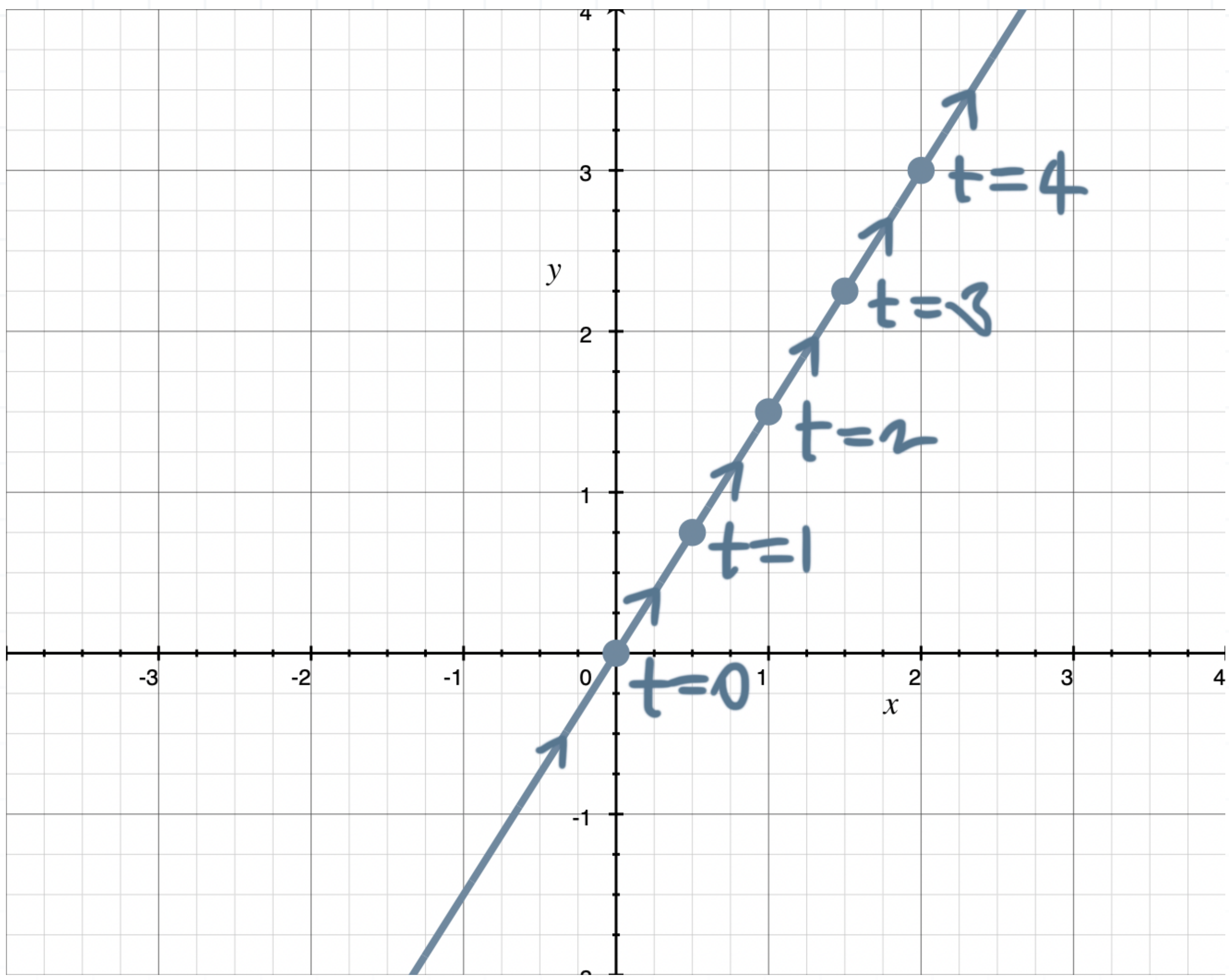
$$4y = 6x$$

$$y = \frac{3}{2}x$$

We know this rectangular equation represents the line with slope $3/2$ that passes through the origin. If we choose a few values of t and find their corresponding x and y values,

t	0	1	2	3	4
x	0	1/2	1	3/2	2
y	0	3/4	3/2	9/4	3

then we can sketch the line and include these specific points.



So we could say that t is increasing as x and y are increasing.

Let's do one more example with parametric equations defined in terms of trig functions.

Example

Sketch the graph of the parametric curve and indicate the direction of increasing t.

$$x = 4\cos t$$

$$y = 4\sin t$$

We'll solve both equations for the trig function, and then square each equation.

$$\cos t = \frac{x}{4} \qquad \to \qquad \cos^2 t = \frac{x^2}{16}$$

$$\sin t = \frac{y}{4} \qquad \to \qquad \sin^2 t = \frac{y^2}{16}$$

Then we can plug these results into the Pythagorean identity.

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

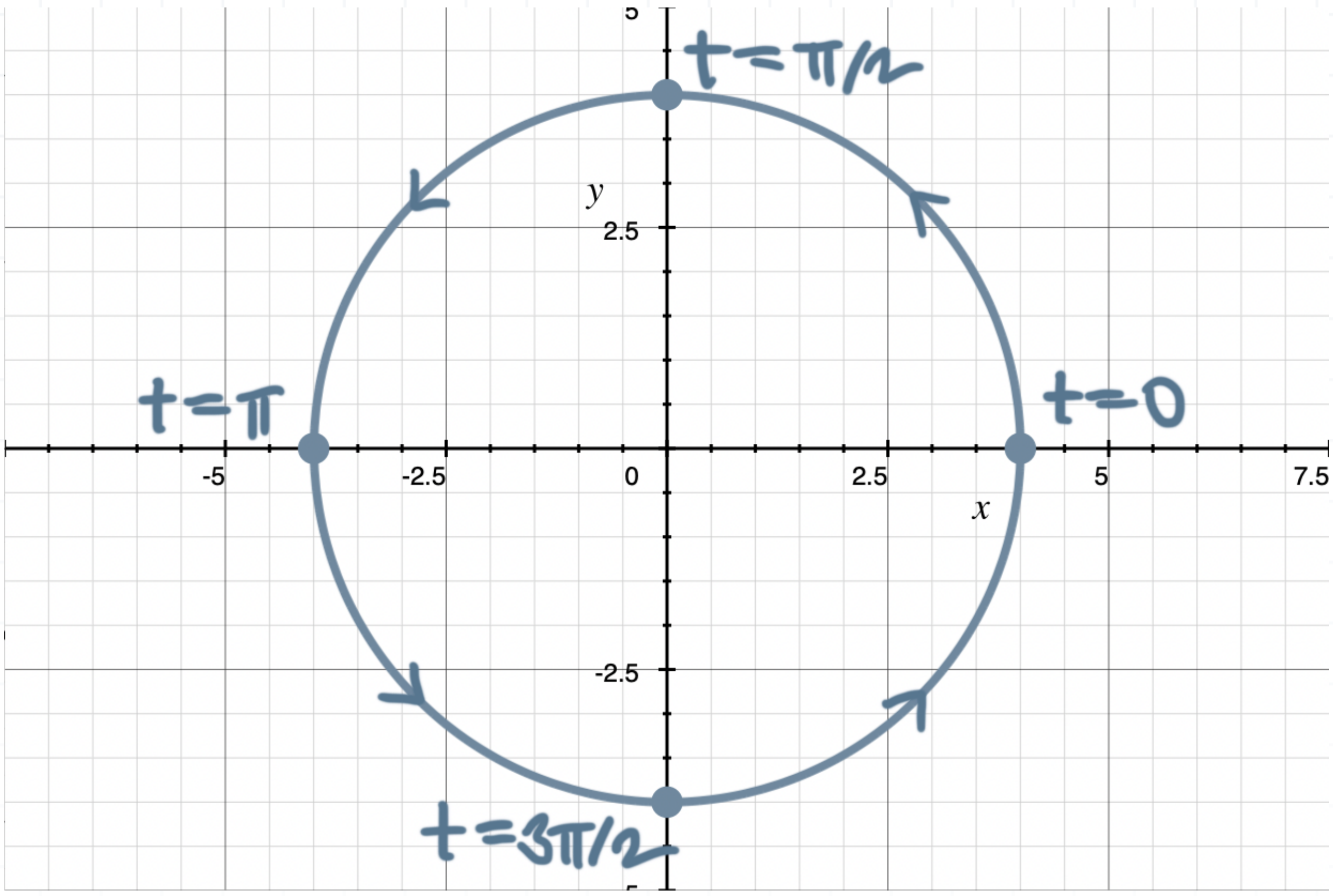
$$\frac{y^2}{16} + \frac{x^2}{16} = 1$$

$$x^2 + y^2 = 16$$

We know this is the equation of a circle centered at the origin with radius 4. If we choose a few values of t and find their corresponding x and y values,

t	0	π/2	π	3π/2	2π
x	4	0	-4	0	4
y	0	4	0	-4	0

then we can sketch the circle and include these specific points.



So we could say that t is increasing as we trace out the circle in a counterclockwise direction.

Finding the parametric representation

Now that we know how to convert a set of parametric equations into a rectangular equation by eliminating the parameter, we want to reverse the direction and understand how to change a rectangular equation into parametric equations.

Substituting the parameter for one variable

One way to do this simply is to define $x = t$, and then substitute into the rectangular equation. This plan will work, as long as y is a function of x, or x is a function of y. In other words, we couldn't use this method to convert the rectangular equation of a circle into parametric equations.

Let's look at an example with the $x = t$ substitution.

Example

Express the rectangular equation in parametric form.

$$x^3 - 6x^2 - y + 15 = 0$$

We can rewrite the equation by solving it for y.

$$x^3 - 6x^2 - y + 15 = 0$$

$$x^3 - 6x^2 + 15 = y$$

$$y = x^3 - 6x^2 + 15$$

For every real number x, the equation gives a unique value of y, so y is a function of x, and we can substitute $x = t$, which gives the parametric equations

$$x = t$$

$$y = t^3 - 6t^2 + 15$$

If x can be expressed as a function of y, we can substitute $y = t$ and follow the same process.

Example

Express the rectangular equation in parametric form.

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

We can rewrite the equation by solving it for x.

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

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

For every real number y, the equation gives a unique value of x, so x is a function of y, and we can substitute $y = t$, which gives the parametric equations

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

$$y = t$$

When a simple substitution won't work

When the rectangular equation isn't a function of y in terms of x, and isn't a function of x in terms of y, we have to find another way to manipulate the equation so that we can break it into a pair of parametric equations.

We'll often use trig functions to represent different parts of the rectangular equation, so let's look at an example.

Example

Express the rectangular equation in parametric form.

$$(x-2)^2 + (y+4)^2 = 10$$

Notice that $(x-2)^2 + (y+4)^2 = 10$ is the equation of the circle that's centered at $(2, -4)$ with radius $\sqrt{10}$.

If we rewrite the equation as,

$$\frac{(x-2)^2}{10} + \frac{(y+4)^2}{10} = 1$$

$$\left(\frac{x-2}{\sqrt{10}}\right)^2+\left(\frac{y+4}{\sqrt{10}}\right)^2=1$$

then we can match this equation to the Pythagorean identity $\sin^2 t+\cos^2 t=1$. We'll set

$$\cos t=\frac{x-2}{\sqrt{10}} \qquad \sin t=\frac{y+4}{\sqrt{10}}$$

$$\sqrt{10}\cos t=x-2 \qquad \sqrt{10}\sin t=y+4$$

to get the parametric equations

$$x=2+\sqrt{10}\cos t$$

$$y=-4+\sqrt{10}\sin t$$

WWW.KRISTAKINGMATH.COM