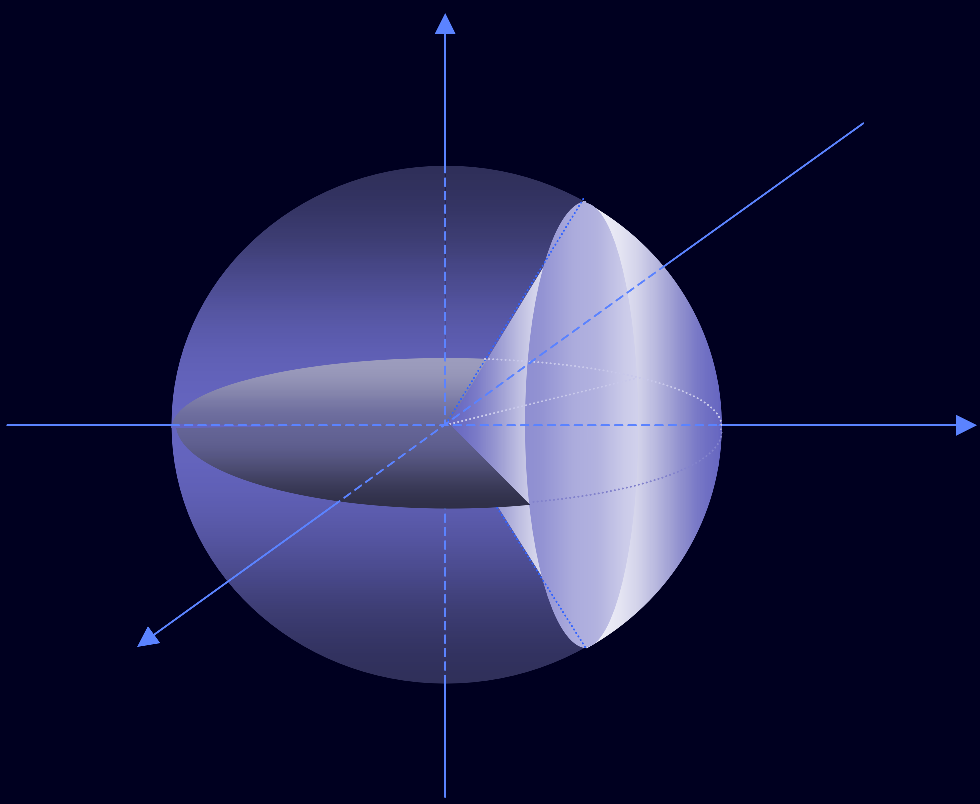
원뿔 곡선

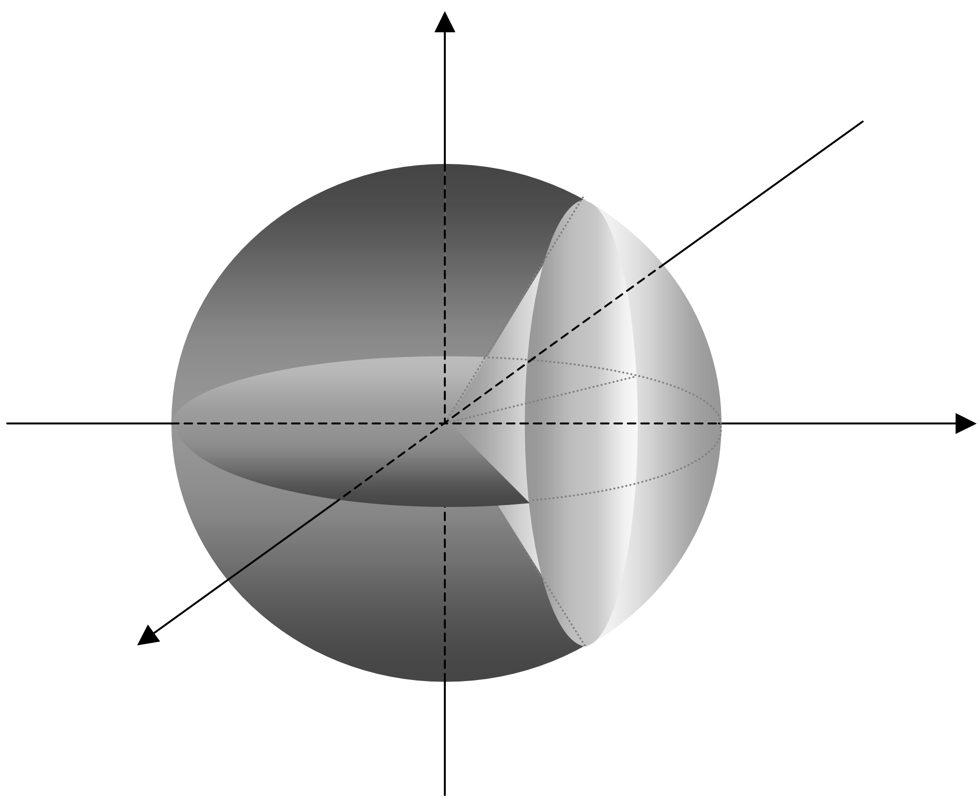
Conic Sections



김성렬

Seong R. Kim

Conic Sections



Seong R. KIM

Dear students,

Students need the best teacher, so you need examples, because examples are the best teacher. This is thus, a book of examples. All the examples here are fully worked, and explain **how** the basic but essential tools in math are made, together with **what** they are, **how** they work, and **how** to work with them. Such tools include numbers, formulas, identities, equations, laws, etc.

Doing math, we work with ideas and run ideas, because every thing in math is an idea. What idea though?

It's a concept. So we need to understand it if we want to use it. A number is an idea, for instance, and the same is true for a line or circle, too. And putting ideas together, we can build another, which becomes the base or an element of another, and each is connected. And that's the way your math grows.

So you get to build a circuit, and sometimes, need to fill the gap or repair the circuit so that you can get the sense of it. Your calculation thus, runs properly, and gets to the solution.

Getting the concept doing good examples, and gaining the fluency doing right examples, you can build the circuit strong.

This book is however, nothing but a bunch of examples until you get it powered. How then to get it powered, and make it run and work for you?

Read it closely at your pace, and then, do each example yourself. It is quite important to note that you do it in **your** own writing. Just watching someone doing it, you just only feel that you can do it, too, but can hardly do it when you need it. You do it, you get it. And you've got to do it over with different examples. It's a cliché, of course, but is always true that knowing is one thing and doing is another.

I've been helping students get conceptual understanding and fluency so that they can grow, take care of, and run their own math. The area covers the basics in algebra and geometry, and includes equations for unknowns and those for curves, functions and their graphs, and other math tools, which are the basic elements in calculus, which's been the core of my interest from my early age in high school.

Of my students, some are quite poor in math, and thus, are afraid of or even hate math, some require special education because of exceptional intelligence, some are smart enough, some are naïve and diligent, and many are clever but lazy. All the students are badly after though, one thing in common.

A Strong Foundation in Algebra

It is of course, the prime objective of my work, and I'm always happy and eager to help them achieve it. I can only help though. Your math grows as much as you grow and take care of it yourself. So does mine. What math then do students most often do or use in high schools or colleges?

Algebra and geometry. What algebra then?

Elementary algebra, of course. Doing the algebra, you should be able to not only work with, but manipulate, that is, change, alter, and convert math expressions the way you want, so for instance, you can break apart and factor polynomials in many kinds. Also, you can handle properly ratios and rates, trigonometry, factors or divisors, constants and variables, powers and logarithms or exponents, equations and functions, along with their graphs, equalities and inequalities, identities and formulas, laws and rules, etc.

And if you want to handle them properly, you want to know their natures and how they mingle with each other.

So studying those stated above, you want to know **what** they are, **how** they work, and **how** to work with them or **what** to do with them. What then about the geometry?

Basically, the geometry is about shapes, and has much to do with positions and angles in algebra. Why algebra?

You often need to compute lengths, areas, or volumes in many situations. The shapes begin with points, lines, triangles, and circles, and move on to rectangles, squares, parallelograms or rhombuses, trapezoids, tetragons, other polygons as pentagons, polyhedrons as hexahedrons, etc.

Doing the geometry, too, we need to do the algebra stated above such as trigonometry. So it is analytic geometry, often called coordinate geometry, too. And doing it, we can specify positions using coordinates. So in that geometry, basically, we work with graphs. Putting a math idea in a graph, we can not only effectively think about it, but actually see it, too, and therefore, can not just easily, but efficiently work with it, too. What idea then is it?

The idea begins with a point, a line, a parabola, a circle, an ellipse, and a hyperbola, called a conic section or basic curve, and then, moves on to other curves, planes, surfaces, volumes, and other objects in various dimensional spaces, together with vectors. And using an angle, we can specify and work with an amount of turning or change in direction.

So learning, using, or applying those ideas or tools in math, we get to solve problems. And this book can help. It can help you build and run your own math engine in your brain, and thus, can help achieve your solid math skill.

It is however, not a magic book giving you a math skill of high caliber overnight. And it’s just a book of examples, and can have many mistakes. I’m a human, too. There’s no mojo making a math genius, and math is science, and is full of facts and ideas. And it is after all, not me, not your teacher, and not any book but you who actually make you put together some of those facts and ideas, and understand it.

Putting facts and ideas together, understanding it, and taking care of what you have learned doing examples, you grow your math. And this book is ready to help and can help you help yourself.

Solving a problem, we put it many different ways. For instance, while expanding or reducing it, or modifying or converting it, we keep searching for the solution, keep approaching the solution, and eventually, can get there.

It is often the case a problem itself is the solution. We can put a problem in many different ways, and eventually, can end up with the solution. How come then is the solution no other than the problem?

For instance, the solution to 3232 ÷ 101 is 32. And we can put it this way:

3232 ÷ 101 $= \frac{3232}{101} = \frac{32 \text{ x } 101}{101} = \frac{32}{1} = 32$ $\Rightarrow$ 3232 ÷ 101 = 32.

And we can get this, too: 32 = 3232 ÷ 101. How?

$32 = \frac{32}{1} = \frac{32 \text{ x } 101}{101} = \frac{3232}{101} =$ 3232/101 = 3232 ÷ 101.

Too easy?

For another instance, the solution to $\boldsymbol{ax^2 + bx + c = 0}$ is

$$\boldsymbol{x} = \frac{-\boldsymbol{b} \pm \sqrt{\boldsymbol{b}^2 - 4\boldsymbol{ac}}}{2\boldsymbol{a}}$$, which is called the quadratic formula.

How come then is the solution no other than the problem?

We can put it this way:

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

$$\Rightarrow 2ax + b = \pm\sqrt{b^2 - 4ac} \Rightarrow (2ax + b)^2 = b^2 - 4ac$$

$$\Rightarrow 4a^2x^2 + 4abx + b^2 = b^2 - 4ac \Rightarrow 4a^2x^2 + 4abx = -4ac$$

$$\Rightarrow ax^2 + bx = -c \Rightarrow ax^2 + bx + c = 0.$$

And we can get this, too:

$$ax^2 + bx + c = 0 \quad \Rightarrow \quad x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} \quad \text{How?}$$

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

$$= a(x^2 + \frac{b}{a}x + \frac{b^2}{4a^2}) - \frac{b^2}{4a} + c = a(x + \frac{b}{2a})^2 - \frac{b^2 - 4ac}{4a} = 0$$

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

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

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

So when solving a problem, we expand or reduce it, modify or convert it, and keep searching for and approaching the solution. Then eventually, we can get there. What then do we do when doing the set of processes above?

We do algebra. You need a strong foundation in algebra.
So if a problem is well defined, that is, if it makes sense, we should be able to get it solved the way as follows.

A problem $\Rightarrow$ **…** $\Rightarrow$ **…** $\Rightarrow$ **the solution**,

and thus, **the problem** $\Rightarrow$ **the solution**.

So solving a problem, we put it many different ways so that we can get to the solution. And that's the way math runs.

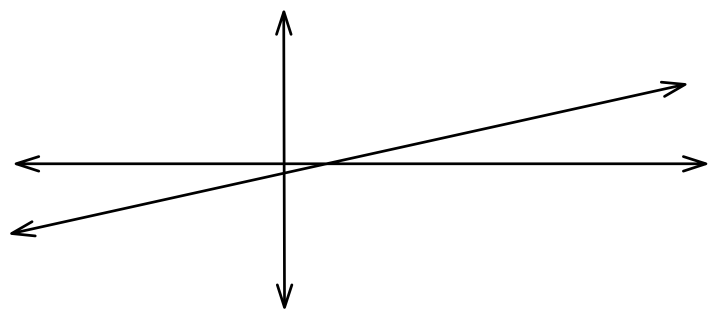
Many more examples are available for free, and you can get them downloaded at **www.runmath.com**. So visit the website, get them downloaded, and do the examples. The website is getting updated weekly or monthly.

May your math run very well.

Seong R. Kim
B.S. Math. Michigan Tech. Univ. M.S. Math. Rensselaer Polytechnic Institute

Contents

0.0. What is a line?



A line in math is in a plane, is of infinite length,
and is a set of points,
where every segment from one point to another is straight,
and has the same slope,
which is called the slope of the line.
And each segment is called a line segment, of course.

So in short, we can put a line the way as follows.

Every segment from one point to another in a line
shares the same slope as the line has.

And we can put the same this way, too:

No matter what two points we may choose in a line,
the slope of the segment between the two is constant,
that is, the same, and is the slope of the line.

So what is a line about?

It is about a slope.
Saying thus, a line in math, we can even say that we mean a slope.

Note: In this book, just saying geometry, we mean analytic geometry, often called coordinate geometry, too, and in geometry analytic, we put in a graph an object as a line, circle, triangle, etc., and study the object.

A line in math is a geometric object, and in geometry analytic, we put it in a graph. Putting an object in a graph, we graph it or make the graph of it.

Ordinarily, a straight line is simply called a line, and a curved line is just called a curve.
In math however, either curved or not, a line can be called a curve, too.
So a curve in a graph can be a line, parabola, circle, ellipse, hyperbola, etc.

And we have another object that looks like a line, but is not a line.
We can get it cutting a line. And it is called a ray or a half-line.

So we can make two different cuts in a line. One is a ray, and the other is a line segment.

A line segment has a length finite, but the length of a line or a ray is infinite. A ray is said to grow one way, but a line is said to grow two ways.

A ray begins with a point, and grows forever in one way only.
A line begins with a point, too, but grows forever in two ways, opposite of each other.

Fig. 0 lines

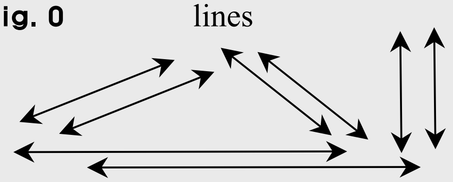
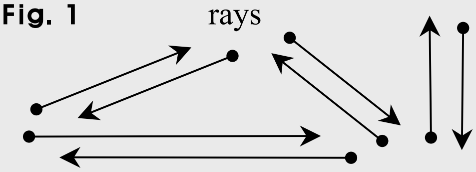


Fig. 1 rays



How can we get a line?

A line is one of conic sections, often just called conics.

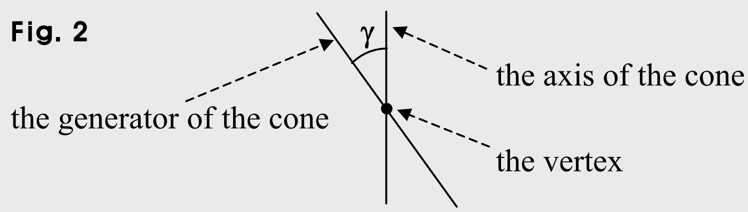
So it is a conic, and if cutting with a plane two right cones meeting at their vertices, we can get a cross section called a line. So it's a conic section, just called a conic, too.

What is a right cone though?

If a cone is a right cone, the line connecting the vertex of the cone and the center of the base is perpendicular to the base, and thus, makes a right angle (90^o) with the base. And the line stated above is called the axis of the cone. And in the study of conics, such a right cone is made or generated the way as follows.

First, take a line as the axis of the cone to be made.
Next, make another line meet the axis of the cone at a point at an angle γ, which is an acute angle, that is, $\mathbf{0 < \gamma < 90^o}$.

Fig. 2

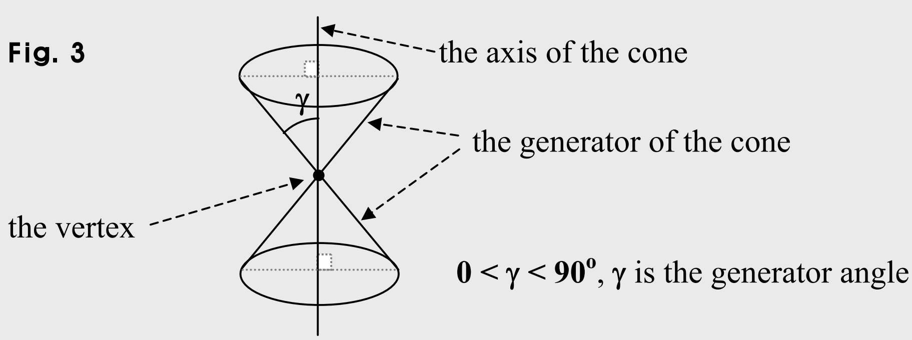


Then, we call the other line the *generator of the cone*, and the point where the generator of the cone meets the axis of the cone is the vertex of the cone.

So the vertex of the cone is the point where the generator meets the axis of the cone. And in the figure above, the angle γ is called the *generator angle*.

Then, rotating the generator around the axis fixing the generator at the vertex, we get a double-cone as shown in Fig. 3.

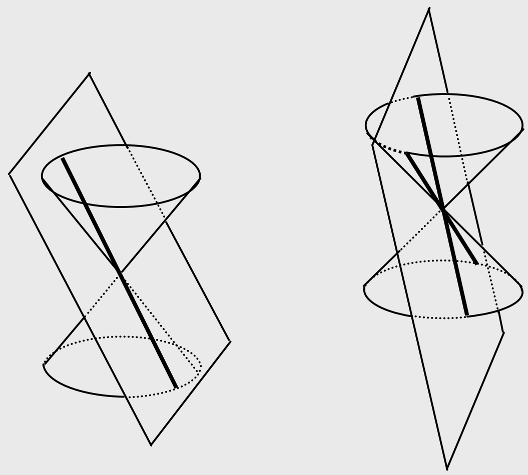
Fig. 3



And note that the generator and the axis are lines, and lines have infinite lengths, so the lengths of the cones are infinite, too. So we cannot show them all. Showing thus, such a cone or a curve in math, we just show some part of it.

How then can we get a line cutting the cone with a plane?

Fig. 4



In the figure above, the cross sections in black are lines. So if α is the angle between the plane and the axis of the cone, and $\mathbf{0} \le \alpha < \gamma$, and if the plane includes the vertex, the cross section is a pair of lines meeting at the vertex. And if the plane is tangent to the cone's side (called the generator of the cone, too), that is, if α is γ, only one line gets made as shown on the left in the figure above, and the line made is in fact, the generator of the cone.

What if a plane cuts through one cone only?

Then, if the plane includes the vertex, and $\mathbf{0} \le \alpha < \gamma$, we get two rays emitting from the vertex. And if the plane is tangent to the side of the cone, that is, if α is γ, only one ray gets created, and is emitting from the vertex.   What then about a line segment?

It is a part of a line, and has a finite length. In math, saying a line segment, we mean a straight line segment, and a distance means a length of a line segment.

So for instance, we indicate a distance between two points by means of the length of the line segment connecting the two points.

What then about a segment cut from a curved line?

It can be called a curved line segment.   What then is an arc?

It is a part of a circle, and is a curved line segment, too. Technically though, a circle is a curved line segment, too. A circle is however, a closed line segment, but an arc is not. For what then do we use a line? That is, what's the purpose of it?

A line has an idea of length or distance, for which we use a line segment.
Also, a line has another idea, called a direction. In geometry analytic or not, a line often means a direction. We use a line or ray to indicate a direction rather than a length or distance, for which we use a line segment.

In fact, a line has a direction. So we can indicate a direction by means of the direction a line or a ray has, and for instance, we can use a line or a ray indicating north or west.

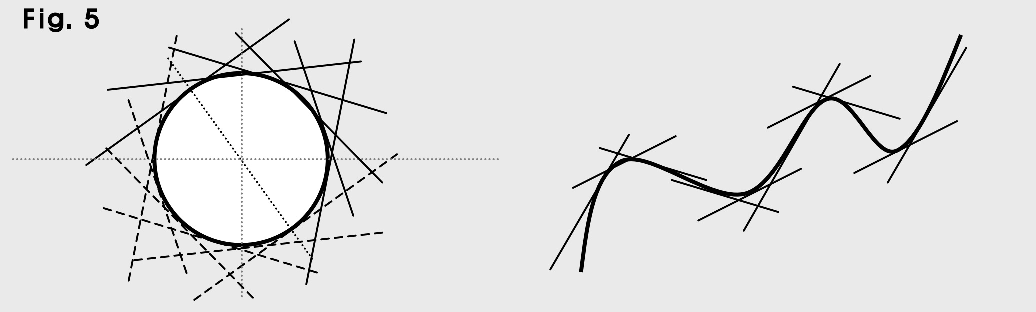
What then do we mean by a direction of a line?

Suppose a line is tangent to a circle, and for instance, the tangent line indicates two O'clock direction at the moment. Then, of course, the line can be said to indicate eight O'clock direction, too.

Suppose now, the line starts moving, and keeps moving being tangent to the circle.

Then, the line keeps changing its direction.
A line can be tangent to a circle at one point only.
So at every moment, the line is tangent to the circle at a different point. So the line gets a different direction as the line moves along the circle keeping itself tangent to the circle.
Thus, we can say that the tangent line can indicate every possible direction.

Fig. 5



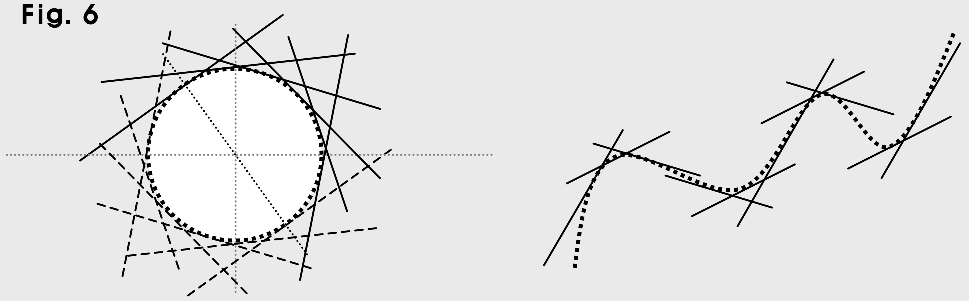
In fact, it doesn't have to be a whole circle, and can be a half circle, too.   How?

In a circle, there are two points, at which the two lines tangent to the circle have the same direction. What are the two points then?

The two points are called antipodal points, and the length of the line segment connecting the two points is the diameter.

And if a line is tangent to a curve, it can be tangent to it at one point only.
In other words, the curve and the tangent line share one point only.

Fig. 6



And there is no bending, turning, twisting, or such in a line, so a line is said to be straight. How?

Suppose a line is tangent to an object, but this time, the line is tangent to the object at all points in the object at the same time. So even moving along the object keeping itself tangent to the object, the line maintains the same direction. What then is the object?

It is a line, too, and in fact, is the same line. Thus, there is no bending, turning, twisting, or such in a line, so a line is straight. It's kind of too forced, isn't it?

So let's now be more specific about a line.

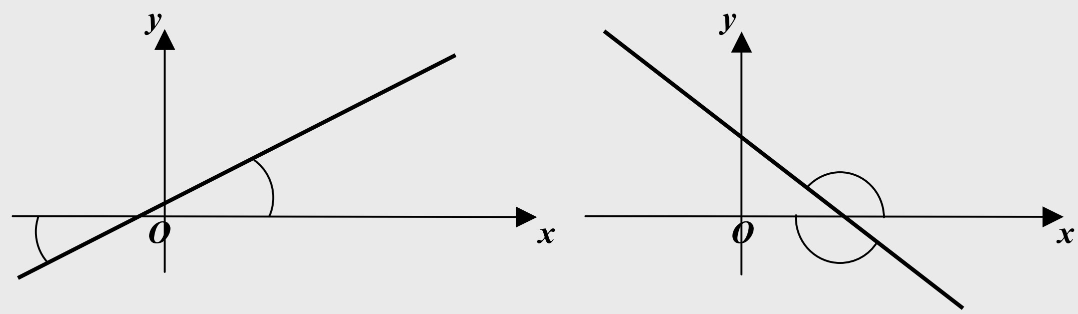
An object in a graph is a collection of points, so a line in a graph is a collection of points. So such a line or its segment is an object composed of points. What points, though?

No matter what two points we may choose in a line in a graph, a line segment connecting the two points has the same direction, which is the direction of the line.

Also, all lines parallel to each other share the same direction.
How then do we specify the direction of a line?

We can take a measurement of an angle, and can use it for the direction.
A line makes an angle against a coordinate axis in a graph, and the axis is normally horizontal. And usually, we put a graph in the x-y plane. So for instance, a line makes an angle against the x-axis, and we take the angle the way as follows.

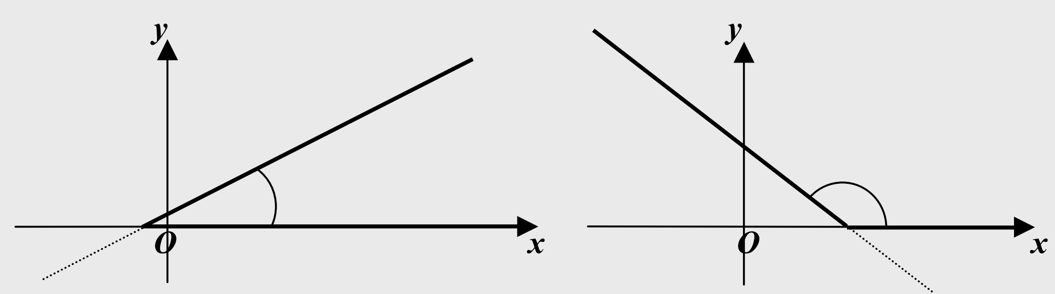
Fig. 7



Taking the measurement of such an angle though, we usually use a ray rather than a line.

Normally, we use as the ray the part of the line above the x-axis, and we use a part of the x-axis on the right of the point where the line meets the x-axis. So we normally take the measurement of the angle between the ray and the part of the x-axis as shown in Fig. 8.

Fig. 8

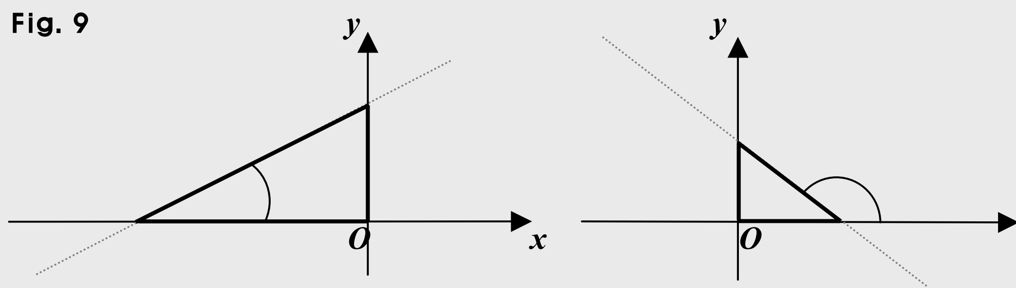


Taking such a measurement though, we do not normally use a protractor. How then can we take the measurement without such a device?

In fact, we usually specify the direction of a line with a different math tool.
The tool is called a slope, and a line has a slope. So we can specify the direction of a line by means of the slope of the line. What slope? What do we mean by the slope?

A slope is the ratio of the height to the base of a right triangle, that is, the ratio of the vertical leg to the horizontal leg.
Assuming a line is in the ***x-y*** plane, we can make a right triangle using parts of both coordinate axes and a part of the line as shown in Fig. 9 below.
So we can get a right triangle, where a part of the ***x***-axis is the base of the triangle, a part of the ***y***-axis is the height, and a part of the line is the hypotenuse.

Fig. 9



So the ratio of the height to the base is the slope of the line, and shows the degree of the steepness of the line. Therefore, the slope tells us how much the line is inclined with the ***x***-axis. However, the slope is often considered to be a rate rather than a ratio. We can take it as a rate of change, as well as a ratio. We will get to the detail in the next section.

Now, once we've got the slope, we can get the angle by means of a special ratio called a trig ratio, which is in this case called the tangent.
So the slope is the tangent of the angle. And parallel lines share the same angle. Thus, being parallel means the same slope and the same angle, as well as the same direction.

Working with a line, we can put the line in a graph. Then, we can actually see the line while working. A graph is not the only place we can put a line in, though.
Where else then can we put a line?

We can put it in an equation, too.
Putting a line in an equation, we express a line by means of an equation, that is, we indicate a line by an equation. And we call such an equation an equation of a line.
And we will get to see how we can get it in the next section.

0.1. Equations for Lines 1

Doing math, we solve problems, because solutions are what math is about. And doing problems, we often do problems with lines. A line seems to be simple, and is in fact, one of the simplest of all curves. When it comes to a problem though, it's not that simple.

Doing a problem with a line, we need to work with the line. Working with a line though, we can't just hold it in our hand, so we want to put it somewhere so that we can work with it. Where then can we put a line?

We can put it in a graph. Then, we can actually see the line working with it. So putting it in a graph, it is much easier to work with it. So doing a problem with a line, graph it.

A graph is not the only place though, we can put a line in. We can put it in an equation, too. And we call such an equation an equation of a line.

Usually thus, putting a line in a graph, we say that we put its equation in a graph, too.
So putting an equation of a line in a graph, we put the line in a graph, and we say that we graph the equation, or make the graph of it.

How then, can we put a line in an equation, that is, how can we get an equation of a line?

Getting an equation of a line, we express a line using an equation.
And expressing a line, we want to know what it's made of. So what is a line made of?

A line is basically a set of points, so it is made of points. What points though?

A set of points is not necessarily a line. Every geometric object put in a graph is in fact, composed of points. So depending on what points put together, the shape of the object gets determined. What points then, do we have to put together to form a line?

As stated in the previous section, we can explain what a line is the way as follows.

> No matter what two points we may pick from a line,
> the segment between the two has the same slope,
> which is the slope of the line.

Suppose for instance, $\boldsymbol{A}$, $\boldsymbol{B}$, and $\boldsymbol{C}$ are three points in a line.
Then, the slope of $\overline{\boldsymbol{AB}}$ is the same as that of $\overline{\boldsymbol{BC}}$, and is the same as that of $\overline{\boldsymbol{AC}}$, too, and thus, is the slope of the line.

So forming a line, we want to put together all such points.
We can't actually do so, of course, but can do it mathematically. How?

Putting all those points in an equation, we get the equation of the line.
So we are back again, to the question that how to get an equation of a line. And in fact, we can get such an equation using the explanation above, since it explains what a line is. That is, it can be the definition for lines. And we can put it the way as follows, too

> Every segment from one point to another in a line has
> the same slope as the line has.

So we can notice that every point in a line has to do with the slope.
Thus, a line is no other than a slope if you will.

So given a particular slope, can we define a particular line, that is, can we get the equation of it?

Many lines can have the same slope, so we can't. In fact, infinitely many lines can share one slope. How many lines then, can share a particular slope and a particular point?

Not even two can do so. Of infinitely many lines sharing one slope, only one can have a particular point. That is, only one line can have a particular slope and a particular point.

So defining a line, we want to get a point the line has, together with the slope.
That is, finding a line, we need at least one point in the line and the slope.

So suppose now that we are given the slope and a point of a line. How then can we get the line, that is, the equation of the line?

As stated earlier, we can get the equation using the definition for lines, which is below.

Every segment from one point to another in a line has the same slope as the line has.

And we can put the same this way, too:

No matter what two points we may choose in a line,
the slope of the line segment between the two is constant,
and thus, is the same.

So let's assume that the slope is $\boldsymbol{a}$, and that $\boldsymbol{L}$ is the line we want to define, and has a particular point $(\boldsymbol{x_1}, \boldsymbol{y_1})$.
What do we mean by a slope, though?

A slope is the ratio of the height to the base of a right triangle.
Assuming a line is in the ***x-y*** plane, we can make a right triangle using parts from the coordinate axes and a part of the line. We will see a specific example shortly.
So now getting back to the definition again, we have this:

No matter what two points we may choose in a line, the slope of the line segment between the two is constant, and thus, is the same.
What then do we mean by the phrase, ***no matter what two points***?

We mean an arbitrary pair of points.   How then can we make an arbitrary pair?

An arbitrary point in a line is a random point, and represents all the points in the line.
So we don't just pick any two points in the ***x-y*** plane, of course.

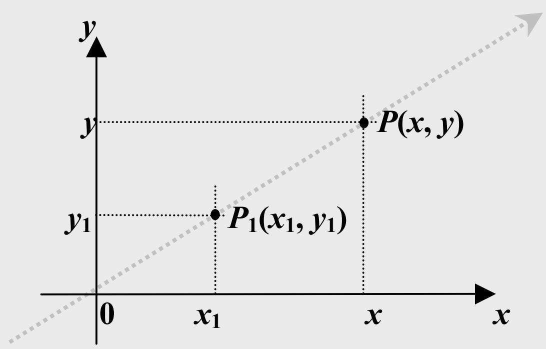
A pair of points has two points. Of the two, one is the point given.
What then is the other point?

It is the point that makes the pair arbitrary, so it has to be an arbitrary point in the line.
So since the line is in the ***x-y*** plane, we can specify the arbitrary point this way: (x, y).

Let's now, put in the ***x-y*** plane the two points stated above. One is (x_1, y_1), which is the point given, and the other is (x, y), which is the arbitrary point in the line.

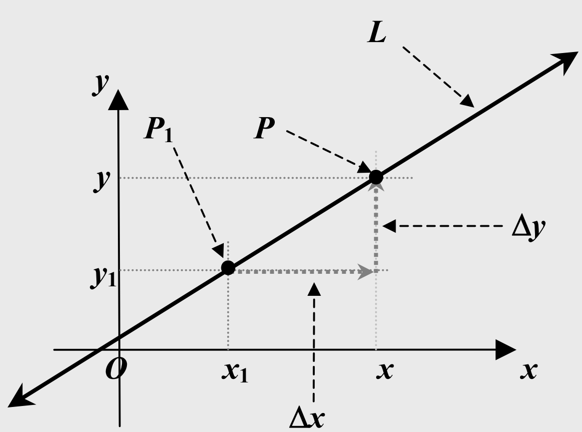
So putting the two in the ***x-y*** plane, we can get a graph as shown in Fig. 0.

Fig. 0



We know L is the line we are after, so the two points P and P_1 are in the line L.
And showing now the line L in the graph, we can show it the way as follows.

Fig. 1



In the graph above, the line segment indicated by Δx is parallel to the x-axis, and the line segment indicated by Δy is parallel to the y-axis.

So we can take Δx as a part of the x-axis, and take Δy as a part of the y-axis.
That's because we get $\Delta x = x - x_1$ and $\Delta y = y - y_1$.

So in the graph above, we can see a right triangle where the base is Δx, the height is Δy, and the hypotenuse is the line segment between P and P_1.
Now again, what do we mean by a slope?

A slope is the ratio of the height to the base of a right triangle.
We have a right triangle in the graph above, and also, a is the slope of the line L.
Thus, the slope of the line segment $\overline{P_1P}$ is a, too.

So we can put the slope of the line L the way as follows.

$\dfrac{\Delta y}{\Delta x} = \dfrac{y - y_1}{x - x_1} = a$, where a is constant.

So what do we have now?

We now have an equation, and the equation is $a = \dfrac{y - y_1}{x - x_1}$. What equation then is it?

It is the equation of the line L. How?

We know (x, y) is an arbitrary point in the line L, so it represents all the points in the line, and the equation above explains the coordinates of the arbitrary point. So?

So the equation explains each and every point in the line L.

For instance, putting a value into x, we can get the value of y, since a, x_1, and y_1 are known to us.

So using the equation, we can get each and every point in the line L. We can say thus, the equation indicates the line L. So we can call it the equation of the line L.
By the way, many people use different words for Δx and Δy in $\dfrac{\Delta y}{\Delta x} = \dfrac{y - y_1}{x - x_1} = a$.

They call Δy 'rise', and call Δx 'run', so the slope is often called 'rise-over-run'.

It makes sense, because Δy is ***the change*** in y-coordinate between two points in a line, and is the change in the vertical direction, and Δx is ***the change*** in x-coordinate, and is the change in the horizontal direction.

In the equation above, Δy is the change in y-coordinate between (x, y) and (x_1, y_1), and Δx is the change in x-coordinate between the two points, which are in the line for which the equation is made.
Usually though, we don't put the equation the way above. It has a bit different look.

Putting it in more conventional form, we get $a = \dfrac{y - y_1}{x - x_1} \Rightarrow y - y_1 = a(x - x_1)$.

So we usually put the equation this way: $y - y_1 = a(x - x_1)$.

And we just call the equation an equation of a line. So indicating a line in general, we can use the equation above, and just call it an equation of a line.
It's not the only form though, we use indicating a line in general.

Solving the equation above for y, we get $y - y_1 = a(x - x_1) \Rightarrow y = ax + y_1 - ax_1$.

And we know a, x_1, and y_1 are all constant. So $y_1 - ax_1$ is constant, too.

So assuming b is constant, and replacing $(y_1 - ax_1)$ with b, we can put the equation above this way too: $y = ax + b$.

So we now have two forms for the same equation. And the two are as follows.

$y - y_1 = a(x - x_1)$ $\qquad$ $y = ax + b$, where $b = y_1 - ax_1$

So all the equations above indicate the same line, and the line has a slope of a, and is passing through a point (x_1, y_1). Thus, given a slope and a point, we can get the line.
By the way, we call each equation above the ***connective equation*** between x and y, because it connects x and y, and explains specifically the relation between the two.

That is, it explains how x and y are related (or connected) in a specific manner.
So we usually call it the connective equation between the two.

And in this case, x and y are the coordinates of the arbitrary point (x, y) in a line.
What then do we need to find finding an equation of a curve as a circle or parabola?

It is the connective equation, which connects the coordinates of the arbitrary point in the curve. And we can find it using the definition for those curves as circles, for instance.

So for instance, using the definition for circles, we can find the connective equation, connecting the coordinates of an arbitrary point (x, y) in a circle. And the connective equation for circles is $(x - a)^2 + (y - b)^2 = c^2$, where (a, b) is the center, and c is the radius.

So the connective equation for certain curves indicates such a curve in general.
Thus, the connective equation for lines indicates a line in general.

And finding a connective equation using such a definition, we say that we ***derive*** the equation. Then, depending on the use of it, we change the form as we did to the connective equation for lines. And then, we give a name to each form.

Now, getting back to the forms we made for lines, we have these:

$$y - y_1 = a(x - x_1). \qquad y = ax + b$$

The first form can be called the ***point-slope*** form, because a is the slope, and (x_1, y_1) is a point in the line.
And the second can be called the ***slope-intercept*** form, because a is the slope, and b is the y-intercept of the line.
The y-intercept is the y-coordinate of the point where the line meets the y-axis.

And of course, there can be many other forms that can indicate a line in a general manner. It all depends on the use of the form, that is, the situation where the form is used.

Many options are not always good though. You might get confused, and there are quite a few things to learn besides things in math.
And another bad news is the names of the forms change from people to people, from country to country, from teachers to teachers, from time to time, etc.
It doesn't really matter though.

What matters is you can do the derivation of the equation from the definition.

At least, you understand the derivation processes, that is, how the derivation can be done. How then can we find the equation of a particular line?

We can find it using one of the forms above, i.e., one of the equations for lines.

So assuming for instance, the slope of a line is 2, and a point (1, 3) is in the line, we can use the slope form, and can put the equation of the line this way: $\boldsymbol{(y - 3) = 2(x - 1)}$.

And for another instance, if the slope is 2 and the $\boldsymbol{y}$-intercept is 1, we can use the intercept form, and can put the equation of the line this way: $\boldsymbol{y = 2x + 1}$.

In fact, the two equations above indicate the same line, so both are actually the same equations, and only have different looks.

So depending on the problem we do, or the situation where the equation is used, we choose one of the forms, and can get the equation putting the specifics into the form. And the specifics are information as the slope, the point in the line, the intercept, or something else, which depends on the form we choose.

So there are some more forms, and you will get to see them in the next section.

0.2. Equations for Lines 2

To begin with, we can put the definition for lines the way as follows.

> Every segment from one point to another in a line has
> the same slope as the line has.

And we can put the same this way, too:

> No matter what two points we may choose in a line,
> the slope of the line segment between the two is constant,
> and thus, is the same.

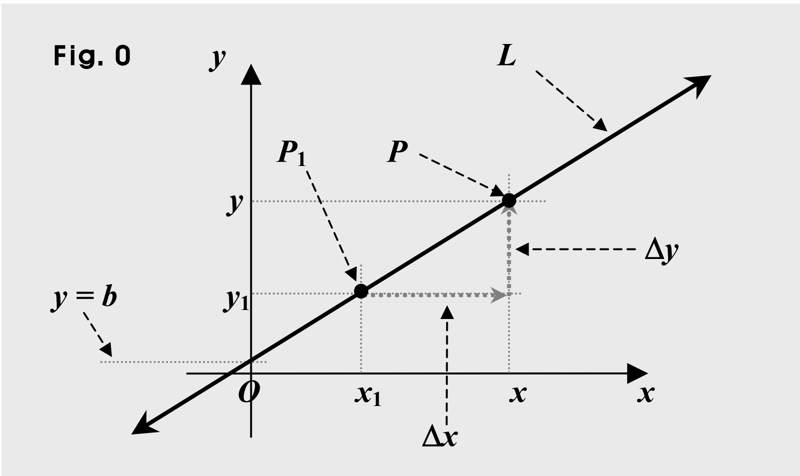
And using the definition above, we can get the connective equation for lines, and can put it in either of the two forms as follows.

$$\boldsymbol{y - y_1 = a(x - x_1)} \qquad \boldsymbol{y = ax + b}, \text{ where } \boldsymbol{b = y_1 - ax_1}$$

So both the two above indicate the same line, and the line has a slope of $\boldsymbol{a}$, and is passing through a point $\boldsymbol{(x_1, y_1)}$. Thus, given a slope and a point, we can get the line.

Of the two above, taking a close look at the second one, we can notice that we get $\boldsymbol{y = b}$ setting $\boldsymbol{x = 0}$ in the equation $\boldsymbol{y = ax + b}$. And we call $\boldsymbol{b}$ the $\boldsymbol{y}$-intercept, which is the $\boldsymbol{y}$-coordinate of the point where the line meets the $\boldsymbol{y}$-axis.

So if the line passes though the origin $\boldsymbol{O}(0, 0)$, the $\boldsymbol{y}$-intercept is 0.
And assuming the equation above indicates a line $\boldsymbol{L}$, we can graph it the way as follows.



Then, we can see that the line $\boldsymbol{y = ax + b}$ cuts through the $\boldsymbol{y}$-axis at a point $\mathbf{(0,\ \boldsymbol{b})}$.

So when $\boldsymbol{x} = \mathbf{0}$, we get $\boldsymbol{y = b}$. Note however, not $\boldsymbol{b}$ but $|\boldsymbol{b}|$ is the distance from the origin to the point where the line crosses the $\boldsymbol{y}$-axis. That's because it can be the case a line can meet the $\boldsymbol{y}$-axis at a point below the origin, that is, below the $\boldsymbol{x}$-axis, so $\boldsymbol{b}$ can be negative.

What then about the $\boldsymbol{x}$-intercept?

We can see it in the graph above, and we can get it setting $\boldsymbol{y} = \mathbf{0}$ in the equation, because when the line crosses the $\boldsymbol{x}$-axis, we get $\boldsymbol{y} = \mathbf{0}$. So setting $\boldsymbol{y} = \mathbf{0}$ in $\boldsymbol{y = ax + b}$, we get

$\mathbf{0} = \boldsymbol{ax + b} \Rightarrow \boldsymbol{x} = -\dfrac{\boldsymbol{b}}{\boldsymbol{a}}$. Thus, the $\boldsymbol{x}$-intercept of the line $\boldsymbol{L}$ is $-\dfrac{\boldsymbol{b}}{\boldsymbol{a}}$.

So for now, we have two forms we can choose from finding an equation of a line.

$$\boldsymbol{y - y_1 = a(x - x_1)} \qquad \boldsymbol{y = ax + b}$$

So if given a slope and a point, we can use this form: $\boldsymbol{y - y_1 = a(x - x_1)}$.

And if given a slope and an intercept, we can use the form of this: $\boldsymbol{y = ax + b}$.

What if we are not given a slope, though?

A line is no other than a slope so to speak. So does no slope mean no line?

There is no rule without an exception. So some lines are said to have no slope.

Lines with no slope are lines vertical, and thus, are perpendicular to the x-axis.
So if lines are perpendicular to the x-axis, the lines have no slope. Why not?

It's because of the definition of the slope.
By definition, if a is a slope, we get $a = \frac{\Delta y}{\Delta x}$.
And we know Δx is the change in x-coordinate between two points in the line.
So we get $\Delta x = 0$ if the line is vertical, that is, perpendicular to the x-axis. How?

No matter what two points we may choose in a vertical line, we get no change in x-coordinate in the two points. So we get $\Delta x = 0$, which is however, the denominator.
And we know that no division by 0 is allowed. So we get no slope if the line is vertical.

So in short, a line vertical has no slope. What then about lines horizontal?

Lines horizontal have slopes, but have to share one slope only, which is 0. How?

Lines parallel to the x-axis are horizontal. So if parallel to the x-axis, their slopes are 0.

It's again because of the definition of the slope. In the definition for slopes above, Δy is the change in y-coordinate in two points in the line.
So we get $\Delta y = 0$ if the line is horizontal, that is, parallel to the x-axis. How?

No matter what two points we may choose in a horizontal line, we get no change in y-coordinate in the two points. So we get $\Delta y = 0$, which is the numerator.
And a division of 0 by any nonzero is 0. So if a line is horizontal, the slope is 0.

What if we are not given any slope, and still need to find a line neither vertical nor horizontal?

If we are asked to find a line not vertical, we are given the slope anyway. That is to say that we need to find somehow the slope from the information given by the problem.

In other words, the slope is hidden somewhere in the problem.
So we've got to play some hide and seek.
Having only to have two points in fact, we can still find the line. How?

Where are the two points?

They are in the line we want to find, and ***two points can define a slope***.
So using the two points, we can find the slope.
Then, using one of the two points, together with the slope, we can find the line.

So for instance, given two points (x_1, y_1) and (x_2, y_2), we get $\dfrac{y_2 - y_1}{x_2 - x_1}$, which is the slope.

And using the slope form $y - y_1 = a(x - x_1)$, since a is the slope, we get

$y - y_1 = \dfrac{y_2 - y_1}{x_2 - x_1}(x - x_1)$, which can be put this way: $(y - y_1)(x_2 - x_1) = (x - x_1)(y_2 - y_1)$,

which is the equation of the line passing through the two points (x_1, y_1) and (x_2, y_2), and can be called the ***two-point*** form.

And we can put it this way, too: $\dfrac{y_2 - y_1}{x_2 - x_1} = \dfrac{y - y_1}{x - x_1}$, which is closer to the definition.

What if the two points have the same x-coordinates, for instance, (1, 2) and (1, 3)?

Then, the line is vertical. A line vertical has no slope, but is still a line, so it has its equation, too. What then is the equation of the line?

It was mentioned above that no slope means no change in x-coordinate between two points in a line.
So at all the points in a vertical line, all the x-coordinates are same,
and thus, the line gets defined to be $x = x$-coordinate which is given in the point given.

So if given (1, 2) and (1, 3), we get $\boldsymbol{x = 1}$, which is the equation of the line passing through the two points (1, 2) and (1, 3).
And for instance, some of the other points in the line $\boldsymbol{x = 1}$ can be (1, -2) and (1, 5).
So the arbitrary point in the line is $\boldsymbol{(1, y)}$.

What if this time, the two points given have the same $\boldsymbol{y}$-coordinates, for instance, (1, 2) and (3, 2)?

Then, the line is horizontal. So its slope is 0. Using thus, the form of $\boldsymbol{y - y_1 = a(x - x_1)}$, called the slope-point form, and using (1, 2), we get $\boldsymbol{y - 2 = 0(x - 2) \Rightarrow y = 2}$, which is the equation of the line passing through the two points (1, 2) and (3, 2).

And we can put the idea this way, too:

If the slope is 0, it means no change in $\boldsymbol{y}$-coordinate between two points in a line.
So at all the points in a line horizontal, all the $\boldsymbol{y}$-coordinates are same,
and thus, the line gets defined to be $\boldsymbol{y = y}$-coordinate which is given in the point given.

So we now have some forms we can use finding a particular line.

Given a slope and a point, we can use the form of $\boldsymbol{y - y_1 = a(x - x_1)}$ or this: $\boldsymbol{a = \dfrac{y - y_1}{x - x_1}}$.

Given a slope and an intercept, we can use the form of this: $\boldsymbol{y = ax + b}$.

Given two points, we can use the form of $\boldsymbol{y - y_1 = \dfrac{y_2 - y_1}{x_2 - x_1}(x - x_1)}$, or the one below.

$$\boldsymbol{(y - y_1)(x_2 - x_1) = (x - x_1)(y_2 - y_1)}.$$

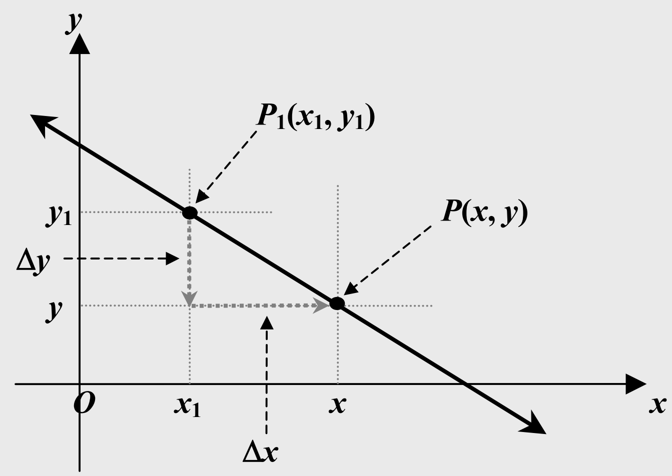
If a line is vertical and has a point $\boldsymbol{(x_1, y_1)}$, its equation is $\boldsymbol{x = x_1}$.

And if a line is horizontal and has a point $\boldsymbol{(x_1, y_1)}$, its equation is $\boldsymbol{y = x_1}$.

What if the slope $\boldsymbol{a}$ is negative?

We can put in a graph a line with a negative slope the way as follows.

Fig. 3



So a slope says how a line slants, and explains the amount and direction a line is inclined with the $\boldsymbol{x}$-axis.
The amount is the magnitude of the slope.
Taking the absolute value of the slope, we can get the magnitude.

And the direction a line slants depends on the sign of its slope.
If the slope is positive, the line slants to the left, and if the slope is negative, it slants to the right.
And if the slope is 0, the line is horizontal, and if no slope, the line is vertical.
What then is the slope?

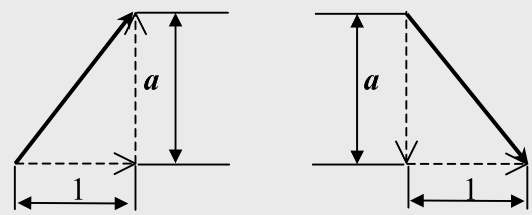
The slope is the ratio of the vertical change to the horizontal change between two points in a line.
The vertical change is the change in $\boldsymbol{y}$-coordinate between two points, and the horizontal change is the change in the $\boldsymbol{x}$-coordinate. So assuming $\boldsymbol{\Delta y}$ is the vertical change, $\boldsymbol{\Delta x}$ is the horizontal change, and $\boldsymbol{a}$ is the slope, we can put the slope $\boldsymbol{a}$ the way as follows.

$a = \dfrac{\Delta y}{\Delta x}$. So if (x_1, y_1) and (x_2, y_2) are two points in a line, the slope is as follows.

$$a = \frac{\Delta y}{\Delta x} = \frac{y_2 - y_1}{x_2 - x_1} = \frac{y_1 - y_2}{x_1 - x_2}.$$

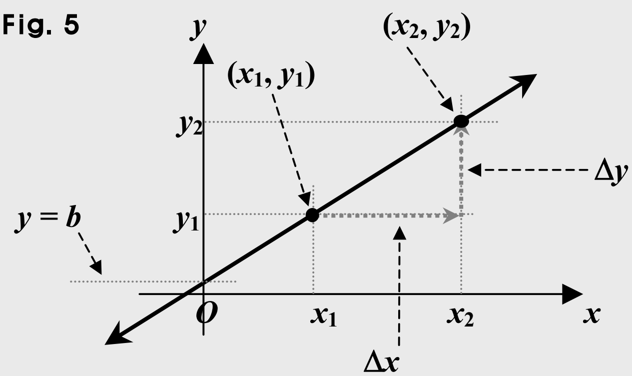
And Δy is often called 'rise', and Δx is called 'run', so it's often called rise-over-run, too.

Fig. 4



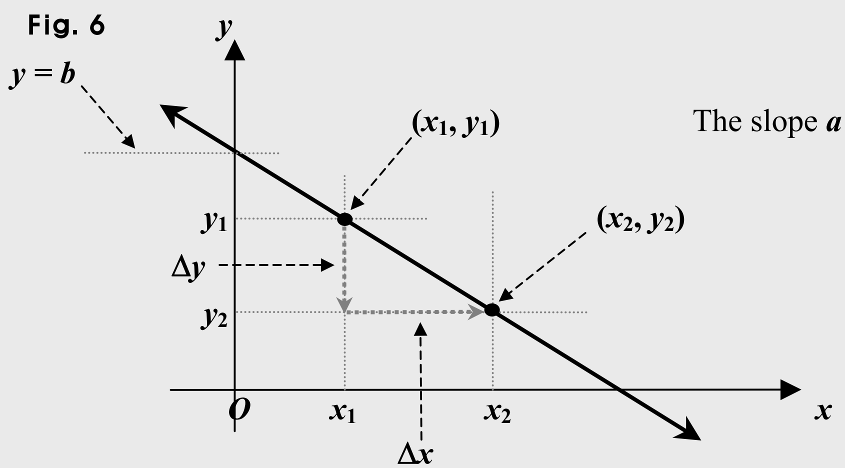
So if we put in a graph a line indicated by the equation $y - y_1 = a(x - x_1)$, depending on the sign of a, the equation indicates either of the two lines as follows.

Fig. 5



The slope $a > 0$.

Fig. 6



The slope $a < 0$.

Normally, a line is said to proceed from the left to the right in the $\boldsymbol{x}$-$\boldsymbol{y}$ plane.
Thus, as a line proceeds, the $\boldsymbol{x}$-coordinate at each point increases, so $\boldsymbol{\Delta x}$ is positive.

Next, if the slope is positive, the $\boldsymbol{y}$-coordinate at each point increases, so $\boldsymbol{\Delta y}$ is positive, too. If therefore, the slope is positive, we get the graph in Fig. 5 above.

If the slope is negative, the $\boldsymbol{y}$-coordinate at each point decreases, so $\boldsymbol{\Delta y}$ is negative, too. So if the slope is negative, we get the graph in Fig. 6 above.

What then about the magnitude?

The bigger the magnitude of the slope, the steeper the line slants.

And the sign of the slope shows the direction the line slants. If the slope is positive, the line slants to the left, and if negative, the line slants to the right.

All the lines parallel to each other share the same slope, and therefore, have the same direction, so they all have the same angle (against the $\boldsymbol{x}$-axis if they are in the $\boldsymbol{x}$-$\boldsymbol{y}$ plane).

Also, we can notice that the slope in a line is not simply a ratio but actually a rate.
How?

Each point in a line has an $\boldsymbol{x}$-coordinate and a $\boldsymbol{y}$-coordinate.
And we call the $\boldsymbol{x}$-coordinate the $\boldsymbol{x}$-value, and call a $\boldsymbol{y}$-coordinate the $\boldsymbol{y}$-value.

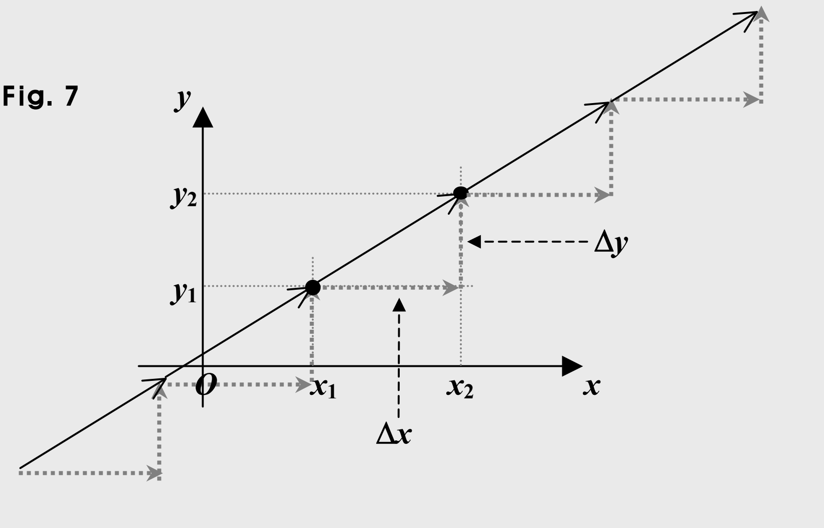
So the slope tells how fast or slow the $\boldsymbol{y}$-value changes as $\boldsymbol{x}$-value changes.
In other words, the slope is the rate of change in $\boldsymbol{y}$-value.

So for instance, assuming $\boldsymbol{y}$-value is an amount of distance, and $\boldsymbol{x}$-value is an amount of time, we can say that the slope is ***the time rate of change in distance***, because it shows how fast or slow the distance changes as the time changes.

Let's now again, put the line $\boldsymbol{y - y_1 = a(x - x_1)}$ in a graph.

Then, assuming $\boldsymbol{a > 0}$, we can graph the line the way as follows.

Fig. 7



And the slope of the line above is as follows. $\dfrac{\Delta y}{\Delta x} = \dfrac{y_2 - y_1}{x_2 - x_1} = \boldsymbol{a}$, where $\boldsymbol{a}$ is constant.

Now, assuming a point (x, y) is moving along the line above, we can see that Δx is the change in the x-coordinate, that is, the change in the direction of the x-axis, and also, see that Δy is the change in the y-coordinate, that is, the change in the direction of the y-axis.

So we can put the idea above the way as follows, too.

As x-coordinate changes by Δx, y-coordinate changes by Δy.

So we can put the expression $\dfrac{\Delta y}{\Delta x}$ the way as follows, too.

It is the ratio of the ***change*** in the y-coordinate to the ***change*** in the x-coordinate.

What then do we mean by such a ratio?

It is a rate. A rate explains a change in one object with respect to a change in the other.

So a rate specifies how the amount of one object changes with respect to change in the other object.

For instance, a speed is a rate, which is a time rate of change in distance.

So a speed specifies how a distance changes with respect to change in time. That is, it can specify the change in distance with respect to a particular amount of change in time.

For instance, an amount of distance can increase every second as 7 m/sec.

So assuming the value of $\boldsymbol{\Delta x}$ indicates time, and that of $\boldsymbol{\Delta y}$ indicates a distance, we can see that the expression above indicates change in distance with respect to change in time, which is a speed, which is a rate, more specifically, a rate of change.

For instance, we can get $\dfrac{\boldsymbol{\Delta y} \text{ meters}}{\boldsymbol{\Delta x} \text{ seconds}} = \dfrac{\frac{\boldsymbol{\Delta y}}{\boldsymbol{\Delta x}} \text{ meters}}{\mathbf{1} \text{ second}} = \dfrac{\boldsymbol{\Delta y}}{\boldsymbol{\Delta x}} \; \dfrac{\text{meters}}{\text{second}}$

That is, the slope of the line specifies the time rate of change in distance, and thus, is a *rate of change*.

Now, the *slope* itself can be positive, 0, or negative, and thus, can be taken as a *velocity*. So the *magnitude* of the slope can be called a *speed*.

In fact, a rate of change, a velocity, and a slope can be synonyms in math.

What then about acceleration?

It’s a time rate of change in velocity. So it is a rate of change in rate of change.

So it can indicate how fast or slow a rate changes, that is, how fast or slow a velocity changes.
If the rate of change increases, the velocity increases, so the slope increases, too.
And if the slope decreases, the velocity decreases, so the rate of change decreases, too.

So no acceleration is in a line. There is no change in slope in a line, so the velocity doesn’t change, that is, we have no change in the rate.
And no change in the rate of change means that acceleration is 0.

So if we collect data a moving body produces, and put the data in a graph, we can get a line as the curve if the body is moving at a constant velocity, and thus, has 0 acceleration.

In sum:

A line is a collection of points in a plane, and the points are as follows.

No matter what two points we may pick from a line, the line segment between the two points has the same slope as the line has. In short, every line segment from one point to another in a line has the same slope as the line has. If a line is not vertical, it has a slope.

The magnitude of the slope shows how steep the line slants.
The bigger the amount of the slope, the steeper the line slants.

The sign of the slope shows the direction the line slants. If the slope is positive, the line slants to the left, and if negative, the line slants to the right.

All the lines parallel to each other have the same slope, and therefore, have the same direction, so they all have the same angle (against the $\boldsymbol{x}$-axis if they are in the $\boldsymbol{x}$-$\boldsymbol{y}$ plane).

And we have some forms we can use finding a particular line.

Given a slope and a point, we can use this form: $\boldsymbol{y - y_1 = a(x - x_1)}$ or this: $\boldsymbol{a = \dfrac{y - y_1}{x - x_1}}$.

Given a slope and an intercept, we can use this form: $\boldsymbol{y = ax - b}$.

Given two points, we can use the form of $\boldsymbol{y - y_1 = \dfrac{y_2 - y_1}{x_2 - x_1}(x - x_1)}$, or the one as follows.

$$\boldsymbol{(y - y_1)(x_2 - x_1) = (x - x_1)(y_2 - y_1)}.$$

0.3. Equations for Lines 3

To begin with, we can put the definition for lines this way: every segment from one point to another in a line has the same slope as the line has.

And we can put the same this way, too: no matter what two points we may choose in a line, the slope of the line segment between the two is constant, and thus, is the same.

And using the definition above, we can get the connective equation for lines, and can put it in either of the two forms as follows.

$$y - y_1 = a(x - x_1) \qquad y = ax + b, \text{ where } b = y_1 - ax_1$$

So all the equations above indicate the same line, and the line has a slope of a, and is passing through a point (x_1, y_1). Thus, given a slope and a point, we can get the line.

And if given two points (x_1, y_1) and (x_2, y_2), using one of the first two forms above, we can get $y - y_1 = \dfrac{y_2 - y_1}{x_2 - x_1}(x - x_1)$, which can be put this way, too: $\dfrac{y_2 - y_1}{x_2 - x_1} = \dfrac{y - y_1}{x - x_1}$, or this way: $(y - y_1)(x_2 - x_1) = (x - x_1)(y_2 - y_1)$.

So putting threads together, we have three forms as follows.

$$y - y_1 = a(x - x_1) \qquad y = ax + b \qquad y - y_1 = \frac{y_2 - y_1}{x_2 - x_1}(x - x_1)$$

So using one of the three forms above, together with the given information as the slope and a point, or two points, we can get the equation of a particular line.

And usually, we call the three forms the equations for lines, or call each form an equation of a line, which means each indicates a line in general.

None of the three forms however, cannot indicate lines of a particular kind.
What kind then is it?

The lines of the kind are vertical lines as $\boldsymbol{x = 1}$.
Examining each of the three forms above, we can notice that for any values of the constants, $\boldsymbol{a}$, $\boldsymbol{b}$, $\boldsymbol{x_1}$, $\boldsymbol{y_1}$, $\boldsymbol{x_2}$, and $\boldsymbol{y_2}$, no form can indicate a vertical line.

It is in fact, because a vertical line has no slope. Each of the three forms has a slope.
In the first two forms, $\boldsymbol{a}$ is the slope, and in the third, the slope is $\boldsymbol{\dfrac{y_2 - y_1}{x_2 - x_1}}$.
Isn't there any form then, that can indicate a vertical line, as well as a line with a slope?

We have one, and it is called the general form, which is thus, called a general equation of a line. And we can put the general form the way as follows.

$$\boldsymbol{ax + by + c = 0},$$

where $\boldsymbol{a}$, $\boldsymbol{b}$, and $\boldsymbol{c}$ are constant,
but $\boldsymbol{a}$ and $\boldsymbol{b}$ cannot be 0 at the same time.

So for instance, setting $\boldsymbol{a = 1}$, $\boldsymbol{b = 0}$, and $\boldsymbol{c = -1}$, we get

$$\boldsymbol{ax + by + c = 0 \Rightarrow x + 0 - 1 = 0 \Rightarrow x = 1},$$

which is a line perpendicular to the $\boldsymbol{x}$-axis, and thus, is vertical.

And for another instance, if putting $\boldsymbol{4x + 2y + 1 = 0}$ in the slope-intercept form, we can simply put it this way: $\boldsymbol{4x + 2y + 1 = 0 \Rightarrow y = 2x} - \dfrac{1}{2}$.

Assuming for instance, $\boldsymbol{ax + by + c = 0}$ indicates a line $\boldsymbol{L}$, we can put it this way, too:

$\boldsymbol{L} = \{(\boldsymbol{x, y})|\ \boldsymbol{ax + by + c = 0}$, where $\boldsymbol{a^2 + b^2 \neq 0}\}$ Not quite sure?

We know a line is a set of points. So $\boldsymbol{L}$ is a set of points, too, and we can indicate $\boldsymbol{L}$ the way above using a set notation. So the set expression above is saying that $\boldsymbol{L}$ is the set of every point $(\boldsymbol{x, y})$ that satisfies the equation $\boldsymbol{ax + by + c = 0}$.

So for instance, if $\boldsymbol{K} = \{(\boldsymbol{x, y})|\ \boldsymbol{2x - y + 1 = 0}\}$, $\boldsymbol{K}$ is a line $\boldsymbol{2x - y + 1 = 0}$.

And we can put the line $\boldsymbol{K}$ this way, too, of course: $\boldsymbol{y = 2x + 1}$.

What do we mean by this though: $\boldsymbol{a^2 + b^2 \neq 0}$?

It means that $\boldsymbol{a}$ and $\boldsymbol{b}$ cannot be 0 at the same time.
And using the set notation, we can indicate some points in a line, too.
For instance, we can indicate a set of three points in such a way as follows.

$\boldsymbol{A} = \{(\boldsymbol{x, y})|\ \boldsymbol{y = 2x}$, and $\boldsymbol{x = 2, 3}$, or $\boldsymbol{4}\} = \{(\boldsymbol{2, 4}), (\boldsymbol{3, 6}), (\boldsymbol{4, 8})\}$.

Also, we can indicate the line $\boldsymbol{L}$ this way, too: $\boldsymbol{L(x, y) = ax + by + c = 0}$.

We can read $\boldsymbol{L(x, y)}$ as ‘L of x and y’, or ‘L of xy’ for short.
So $\boldsymbol{L(x, y)}$ is the name of the equation of the line $\boldsymbol{L}$, and the equation is $\boldsymbol{ax + by + c = 0}$.
And in this case, the letter $\boldsymbol{L}$ is called the equation designator, which means, of course, the name of the equation.
So note that $\boldsymbol{L(x, y)}$ is not a function but an equation, made of two variables.

And the conics is in fact, about equations of degree 2 with two variables.
So putting an equation of a line in full general manner, we can put it the way as follows.

$$\boldsymbol{ax^2 + by^2 + cxy + ux + vy + w = 0}$$

For ***some*** and not all values of the constants, the equation above can indicate a line or two lines at the same time.

For instance, $\boldsymbol{x^2 + 3y^2 + 4xy + 3x + 7y + 2 = 0}$ is an equation of two lines.

One is $\boldsymbol{x + 3y + 1 = 0}$, and the other is $\boldsymbol{x + y + 2 = 0}$.

So we have quite a few ways to put an equation of a line.

Taking one more example, we can put a line in this form, too: $\boldsymbol{\frac{x}{a} + \frac{y}{b} = 1}$.

Then, of course, $\boldsymbol{a}$ and $\boldsymbol{b}$ are nonzero constants, since they are denominators.

In the form above, $\boldsymbol{a}$ is in fact, the $\boldsymbol{x}$-intercept, and $\boldsymbol{b}$ is the $\boldsymbol{y}$-intercept. How?

Setting $\boldsymbol{x = 0}$ in the equation, we get the $\boldsymbol{y}$-intercept, and setting $\boldsymbol{y = 0}$, we get the $\boldsymbol{x}$-intercept. So setting $\boldsymbol{x = 0}$ in $\boldsymbol{\frac{x}{a} + \frac{y}{b} = 1}$, we get $\boldsymbol{y = b}$, and setting $\boldsymbol{y = 0}$, we get $\boldsymbol{x = a}$.

So the form above can be called the ***intercept*** form.

Note however, that the form above cannot indicate a line passing through the origin.

Modifying the form though, we can get $\boldsymbol{\frac{x}{a} + \frac{y}{b} = 1 \Rightarrow bx + ay = ab}$.

Then, in the form $\boldsymbol{bx + ay = ab}$, setting $\boldsymbol{a = 0}$, we get $\boldsymbol{x = 0}$, which is the $\boldsymbol{y}$-axis, and setting $\boldsymbol{b = 0}$, we get $\boldsymbol{y = 0}$, which is the $\boldsymbol{x}$-axis.

So if knowing the $\boldsymbol{x}$-intercept is $\boldsymbol{a}$, and the $\boldsymbol{y}$-intercept is $\boldsymbol{b}$, using this form: $\boldsymbol{bx + ay = ab}$, we can quickly get the equation of the line.

Usually though, if a line passes through the origin, we put the line in this form: $\boldsymbol{y = ax}$, where $\boldsymbol{a}$ is the slope, of course. We can in fact, get the form simply putting (0, 0) into the form as follows. $\boldsymbol{y - y_1 = a(x - x_1)}$. That is, setting $\boldsymbol{(x_1, y_1) = (0, 0)}$, we get $\boldsymbol{y = ax}$.

And in fact, we can get this: $\boldsymbol{y - y_1 = a(x - x_1)}$ applying to the form $\boldsymbol{y = ax}$, a graph operation called a parallel transformation.

In such a parallel transformation, we translate the line $\boldsymbol{y = ax}$ in the amount of $\boldsymbol{x_1}$ in the direction of the $\boldsymbol{x}$-axis, and in the amount of $\boldsymbol{y_1}$ in the direction of the $\boldsymbol{y}$-axis.

Then, we get $y - y_1 = a(x - x_1)$, of course.

For the details on transformations, refer to the book **Function Transformations**.

So now, putting threads together, we have five forms, and the forms are as follows.

The point-slope form, $y - y_1 = a(x - x_1)$, which needs the slope and a point.

The slope-intercept form, $y = ax + b$, which needs the slope and x or y intercept.

The two-point form, $y - y_1 = \frac{y_2 - y_1}{x_2 - x_1}(x - x_1)$, which needs two points.

The general form, $ax + by + c = 0$, which is used in general, and can be put in one of the forms above.

And the intercept form, $\frac{x}{a} + \frac{y}{b} = 1$, where a is the x-intercept, and b is the y-intercept.
Or another intercept form, $bx + ay = ab$, where a is the x-intercept, of course, and b is the y-intercept.

So depending on the problem with a line or the situation where the equation is used, we can choose one of the forms above, and can get the equation of the line we want.

For instance, if the slope is 3 and a point (1, -2) is given,
we can use $y - y_1 = a(x - x_1)$,
and then, we get $y - (-2) = 3(x - 1) \Rightarrow y + 2 = 3(x - 1)$,
which can be put this way, too: $y = 3x - 5$.

Next, if the slope is –2, and the x-intercept is 3,
we can use $y = ax + b$.
Then, since the x-intercept is 3, we get $y = 0 \Rightarrow x = 3$.
And $a = -2$. So we get $y = ax + b \Rightarrow 0 = -2 \cdot 3 + b \Rightarrow b = 6$.
Thus, the line is $y = -2x + 6$.

Next, if the slope is 3, and the y-intercept is 2, we can use this again $y = ax + b$.
And we know that we get the y-intercept when $x = 0$.
So in the form above, b is the y-intercept, since we get $y = b$ if $x = 0$ in $y = ax + b$.

Thus, since the slope is 3, we get $a = 3$, and since the y-intercept is 2, we get $b = 2$.
So the line is $y = 3x + 2$. And of course, we can get the line the way as follows, too.

Since the y-intercept is 2, we get $x = 0 \Rightarrow y = 2$.
So we get $y = ax + b \Rightarrow 2 = a \cdot 0 + b \Rightarrow b = 2$. And $a = 3$. So the line is $y = 3x + 2$.

Next, if we are given (1, -2) and (-3, 4), we can use this form: $y - y_1 = \dfrac{y_2 - y_1}{x_2 - x_1}(x - x_1)$.

Then, setting $(x_1, y_1) = (1, -2)$, and $(x_2, y_2) = (-3, 4)$, we get

$$y - y_1 = \frac{y_2 - y_1}{x_2 - x_1}(x - x_1) \Rightarrow y - (-2) = \frac{4 - (-2)}{-3 - (-1)}(x - 1) \Rightarrow y + 2 = \frac{4 + 2}{-3 + 1}(x - 1)$$

$$\Rightarrow y + 2 = \frac{6}{-2}(x - 1) = -3(x - 1) \Rightarrow y + 2 = -3(x - 1)$$, which is the line.

For some reason though, some people or teachers prefer this form: $y = ax + b$.
So if that is the case, you can keep them happy the way as follows.

$y + 2 = -3(x - 1) \Rightarrow y + 2 = -3x + 3 \Rightarrow y = -3x + 1$, which is the line.

And next, if 2 is the x-intercept, and -3 is the y-intercept, we can use either of the two as follows. $\dfrac{x}{a} + \dfrac{y}{b} = 1$, or $bx + ay = ab$, where a is the x-intercept, and b is the y-intercept.

Then, using the first, we get $\dfrac{x}{a} + \dfrac{y}{b} = 1 \Rightarrow \dfrac{x}{2} + \dfrac{y}{-3} = 1 \Rightarrow -3x + 2y = -6$, which is the line,

and can be put this way: $2y - 3x + 6 = 0$, can be put this way, too, of course: $y = \dfrac{3}{2}x - 3$.

And using the second form, we get $-3x + 2y = -6$, which is the one above.

0.4. Equations for Lines 4

Normally, we put one line in one equation, and yet several lines can be put in one equation, too. In other words, an equation can indicate several lines at once.

In fact, we can specify by one equation as many lines as we want. Putting a line in an equation, we often use either of two forms, one is the point-slope form, and the other is the general form. Putting however, in an equation more than one line, we use the general form, which is $ax + by + c = 0$ where a, b, and c are constant.

If an equation indicates more than one line, the equation is composed of as many equations as the number of the lines, and each of the equations is an equation of a line. So let's first, put two lines in one equation.

Suppose U is a line $ax + by + c = 0$, and V is another line $px + qy + r = 0$.

Then, we can put the two lines U and V in one equation the way as follows.

$(ax + by + c)(px + qy + r) = 0$. How though, does it indicate two lines?

Assuming for instance, $AB = 0$, we get $A = 0$ or $B = 0$. Likewise, we can get this:

$(ax + by + c)(px + qy + r) = 0 \Rightarrow ax + by + c = 0$ or $px + qy + r = 0$.

So the one equation $(ax + by + c)(px + qy + r) = 0$ can indicate two lines, one of which is $ax + by + c = 0$, which is the line U, and the other is $px + qy + r = 0$, which is the line V.

And yet, there is another way.
We can put the two lines U and V in this equation: $U^2 + V^2 = 0$.
Then, we get $U = 0$ and $V = 0$.
So the one equation $(ax + by + c)^2 + (px + qy + r)^2 = 0$ can indicate two lines, one of which is $ax + by + c = 0$, which is the line U, and the other is $px + qy + r = 0$, which is the line V.

Assuming next, $ABC = 0$, we get $A = 0$, $B = 0$, or $C = 0$.
So assuming W is another line $ux + vy + w = 0$, we can get this:

$(ax + by + c)(px + qy + r)(ux + vy + w) = 0 \Rightarrow ax + by + c = 0$, $px + qy + r = 0$, or $ux + vy + w = 0$.
So the one equation $(ax + by + c)(px + qy + r)(ux + vy + w) = 0$ can indicate three lines.

And of course, we can use the other way.
Assuming now, $A^2 + B^2 + C^2 = 0$, we get $A = 0$, $B = 0$, and $C = 0$.
So assuming W is another line $ux + vy + w = 0$, we can get this:

$(ax + by + c)^2 + (px + qy + r)^2 + (ux + vy + w)^2 = 0 \Rightarrow ax + by + c = 0$, $px + qy + r = 0$, and $ux + vy + w = 0$.
So the one equation $(ax + by + c)^2 + (px + qy + r)^2 + (ux + vy + w)^2 = 0$ can indicate three lines.

And by the same token, we can put in an equation more than three lines, too.
So let's this time, put n lines in one equation.

Suppose first, $L_1L_2L_3 \dots L_n = 0$, where $L_k = a_kx + b_ky + c_k$, where a_k, b_k, and c_k are constant, where $k = 1, 2, \dots n$.

That is, we have $L_1 = a_1x + b_1y + c_1$, $L_2 = a_2x + b_2y + c_2$, . . ., and $L_n = a_nx + b_ny + c_n$.

Then, we get

$L_1L_2L_3 \dots L_n = 0 \Rightarrow (a_1x + b_1y + c_1)(a_2x + b_2y + c_2) \dots (a_nx + b_ny + c_n) = 0$.

So we get $L_1 = 0$, $L_2 = 0$, . . . or $L_n = 0$.
That is, we get $L_k = 0$, where $k = 1, 2, \dots n$.

In other words, we get $a_1x + b_1y + c_1 = 0$, $a_2x + b_2y + c_2 = 0 \dots$ or $a_nx + b_ny + c_n = 0$.

Therefore, $L_1L_2L_3 \dots L_n = 0$ is an equation that indicates n lines.

And of course, we can use the other way, too.

We can get $L_1^2 + L_2^2 + L_3^2 + \ldots + L_n^2 = 0$

$\Rightarrow (a_1x + b_1y + c_1)^2 + (a_2x + b_2y + c_2)^2 + \ldots + (a_nx + b_ny + c_n)^2 = 0$.

So we get $L_1 = 0$, $L_2 = 0$, . . . and $L_n = 0$.

That is, we get $L_k = 0$, where $k = 1, 2, \ldots n$.

In other words, we get $a_1x + b_1y + c_1 = 0$, $a_2x + b_2y + c_2 = 0 \ldots$ and $a_nx + b_ny + c_n = 0$.

Therefore, $L_1L_2L_3 \ldots L_n = 0$ is an equation that indicates n lines.

Then, can we put two or more lines in one set of points, too?

Yes, we can. We can put in one set of points, as many lines as we want.
For instance, putting a line L in a set of points, we can put the line L in a set of points the way as follows.
$L = \{(x, y)|\ ax + by + c = 0$ where a, b, and c are constant, but $a^2 + b^2 \neq 0$.$\}$.

What do we mean by $a^2 + b^2 \neq 0$, though?

It means that a and b can be any real number, but cannot be 0 at the same time. So either of a and b is not 0. If $a^2 + b^2 = 0$, a and b both have to be 0 at the same time.

And next, assuming for instance, that A is a set of points indicating the two lines U and V stated earlier, we can put the set A the way as follows.

$A = \{(x, y)|\ (ax + by + c)(px + qy + r) = 0$ where a, b, c, p, q, and r are constant.$\}$

And of course, $a^2 + b^2 \neq 0$, and $p^2 + q^2 \neq 0$, either.
And by the same token, we can put in a set of points more than two lines, too.
Suppose B is a set of points that indicates all those n lines stated above. Then, we get

$B = \{(x, y)|\ (a_1x + b_1y + c_1)(a_2x + b_2y + c_2) \ldots (a_nx + b_ny + c_n) = 0$ where a_k, b_k, and c_k are constant, $k = 1, 2, \ldots n$, and $a_k^2 + b_k^2 \neq 0$.$\}$

Of course, $a_k^2 + b_k^2 \neq 0 \Rightarrow a_k$ and b_k for $k = 1, 2, \ldots n$ can be any real number, but cannot be 0 at the same time. So neither of a_k and b_k is 0.

And lines can make a line, too.

When making, indicating, finding, or defining a line, we put it in an equation.
Then, defining or finding a particular line, what do we need at least?

Two points in the particular line, or one point and the slope of the line
So two points can define a line, and thus, can make a line if you will.

Given two points though, we can define (or make) one line only, because one particular line only can pass through the two points, and no other line can do so.

Let's now, go over briefly how to define a line with two points.

Suppose in the $\boldsymbol{x}$-$\boldsymbol{y}$ plane, two points $(\boldsymbol{s}, \boldsymbol{t})$ and $(\boldsymbol{u}, \boldsymbol{v})$ are in a line.
Then, we can put the line in an equation the way as follows.

$$\boldsymbol{y - t = a(x - s)}, \text{ or } \boldsymbol{y - v = a(x - u)}, \text{ where } \boldsymbol{a = \frac{v - t}{u - s} = \frac{t - v}{s - u}}.$$

Both equations above are equivalent to each other, that is, they only look different.
So for instance, given **(-1, 3)** and **(-2, 5)**, we can get the slope the way as follows.

$$\mathbf{\frac{5-3}{-2+1} = \frac{2}{-1} = \text{-}2 = \frac{3-5}{-1+2} = \frac{-2}{1}}.$$

So we get $\boldsymbol{y - 3 = -2(x + 1) \Rightarrow y = -2x + 1}$, and also, $\boldsymbol{y - 5 = -2(x + 2) \Rightarrow y = -2x + 1}$.

Thus, two points can define one line only, which is one particular line.

Next, a point and a slope can define a particular line, too.

Given a point and a slope, we can define or make one line only, since one particular line only can have the slope and pass through the point. So for instance, if a line of slope $\boldsymbol{m}$ has a point $(\boldsymbol{s}, \boldsymbol{t})$, we can define the line by an equation as follows.

$$\boldsymbol{y - t = m(x - s)}$$

Thus, two points can define one line only, and so do a point and a slope.

So we can make a line by means of two points or a point and a slope if you will.
That's not the only way though, we can make a line.
We can make a line by means of lines, too. That is, lines too, can make a line if you will.
Putting lines together, we can make another line.

We can use two different lines to make another line. Calling the two lines, original lines, we can make a new line by means of the two original lines.

Suppose first, the original lines are parallel to each other. Then, the new line, too, is parallel to the lines original. That is, all the three lines are parallel to each other.

Suppose next, the original lines meet each other at a particular point. Then, the new line, too, meets the originals at the particular point. That is, all the three lines meet each other at the particular point altogether. So they share the particular point altogether.
How then do we make the new line?

If we take the average of two original lines, the average can be the new line.
What do we mean by the average though?

Taking the sum of two originals, and then dividing the sum by 2, we get a line, which is the average.
So for instance, making a new line with $\boldsymbol{y = a_1x + b_1}$, and $\boldsymbol{y = a_2x + b_2}$, we get

$$\boldsymbol{y + y = (a_1x + b_1) + (a_2x + b_2) \Rightarrow 2y = (a_1 + a_2)x + (b_1 + b_2) \Rightarrow y = \frac{a_1 + a_2}{2}x + \frac{b_1 + b_2}{2}}.$$

So we get a new line, which is the average of the two lines original.
However, the new line above is not the only new line we can get.
That is, such a new line doesn't have to be the average. How?

Suppose $\boldsymbol{A}$ is a line $\boldsymbol{ax + by + c = 0}$, and $\boldsymbol{B}$ is another line $\boldsymbol{ux + vy + w = 0}$.

Then, taking the sum of the two, we get $\boldsymbol{A + B}$, which is $\boldsymbol{(a + u)x + (b + v)y + c + w = 0}$, which, too, can be a new line. Also, taking $\boldsymbol{2A + B}$, we get

$$\boldsymbol{2(ax + by + c) + ux + vy + w = (2a + u)x + (2b + v)y + 2c + w = 0},$$

which can be a new line, too. What do we mean by such a new line, though?

If $\boldsymbol{A}$ is parallel to $\boldsymbol{B}$, the new line is parallel to $\boldsymbol{A}$, too, that is, if $\boldsymbol{C}$ is the new line, $\boldsymbol{A}$, $\boldsymbol{B}$, and $\boldsymbol{C}$ are parallel to each other. If however, $\boldsymbol{A}$ meets $\boldsymbol{B}$ at a point $\boldsymbol{P}$, the new line $\boldsymbol{C}$ meets $\boldsymbol{B}$ at the point $\boldsymbol{P}$, too, that is, all the three lines $\boldsymbol{A}$, $\boldsymbol{B}$, and $\boldsymbol{C}$ share the point $\boldsymbol{P}$.

Suppose next, $\boldsymbol{A}$ is a line $\boldsymbol{y = ax + b}$, and $\boldsymbol{B}$ is another line $\boldsymbol{y = ax + c}$.
Then, $\boldsymbol{A}$ and $\boldsymbol{B}$ are parallel to each other, and taking the sum of the two, we get

$\boldsymbol{(y = ax + b) + (y = ax + c) \Rightarrow 2y = 2ax + b + c \Rightarrow y = ax + \frac{b+c}{2}}$, which is a new line,

which is a line of slope $\boldsymbol{a}$, too. So the new line, also, is parallel to the original two lines.
Suppose this time, $\boldsymbol{A}$ is a line $\boldsymbol{y = x + 1}$, and $\boldsymbol{B}$ is another line $\boldsymbol{y = -2x + 5}$.
Then, first, finding the point where the two originals $\boldsymbol{A}$ and $\boldsymbol{B}$ meet each other, we get

$$\boldsymbol{x + 1 = -2x + 5 \Rightarrow 3x = 4 \Rightarrow x = \frac{4}{3} \Rightarrow y = x + 1 = \frac{4}{3} + 1 = \frac{7}{3}}$$

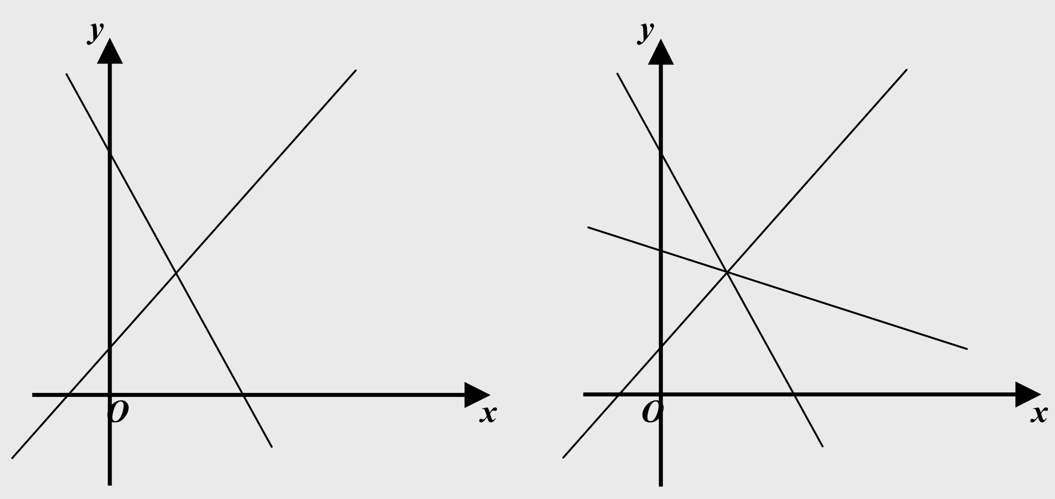
So the original two meet at a point $\boldsymbol{(\frac{4}{3}, \frac{7}{3})}$. Next, taking the average of the two originals,

we get $\boldsymbol{(y = x + 1) + (y = -2x + 5) \Rightarrow 2y = -x + 6 \Rightarrow y = -\frac{1}{2}x + 3}$, which is a new line.

Now, checking to see if the new line passes through the point $\boldsymbol{(\frac{4}{3}, \frac{7}{3})}$, too, we get

$\boldsymbol{y = -\frac{1}{2}x + 3 \Rightarrow \frac{7}{3} = -\frac{1}{2}\cdot\frac{4}{3} + 3 = -\frac{2}{3} + 3 = \frac{7}{3}}$. It does, so the three lines share the point.

Fig. 0



In fact, any multiples of the two lines ***A*** and ***B*** can make a new line, and the new line, too, includes the point $(\frac{4}{3}, \frac{7}{3})$.

For instance, taking the sum of the line ***A*** and twice the line ***B***, we get a new line passing through the point $(\frac{4}{3}, \frac{7}{3})$. Not quite sure?

Making the new one stated above, we get

$$(y = x + 1) + (2y = -4x + 10) \Rightarrow 3y = -3x + 11 \Rightarrow y = -x + \frac{11}{3}$$

Checking to see if it includes the point $(\frac{4}{3}, \frac{7}{3})$, too,

we get $y = -x + \frac{11}{3} \Rightarrow \frac{7}{3} = -\frac{4}{3} + \frac{11}{3} = \frac{7}{3}$.

Let's multiply the line ***A*** by -10, and the line ***B*** by 20, and take the sum of the product.

$(-10y = -10x - 10) + (20y = -40x + 100) \Rightarrow 10y = -50x + 90 \Rightarrow y = -5x + 9$, a new one.

Checking to see if the new one above includes the point $(\frac{4}{3}, \frac{7}{3})$, too,

we get $y = -5x + 9 = -5 \cdot \frac{4}{3} + 9 = \frac{7}{3}$.

Let's take the sum of the line ***A*** and 20 times the line ***B***, and check it again.

$(y = x + 1) + (20y = -40x + 100) \Rightarrow 21y = -39x + 101 \Rightarrow y = -\frac{39}{21}x + \frac{101}{21}$, another new.

Checking to see if the other new one above includes the point $(\frac{4}{3}, \frac{7}{3})$, too,

we get $y = -\frac{39}{21}x + \frac{101}{21} = -\frac{39}{21} \cdot \frac{4}{3} + \frac{101}{21} = \frac{-13 \cdot 4 + 101}{21} = \frac{49}{21} = \frac{7}{3}$.

Let's next, take the sum of the line $\boldsymbol{B}$ and twice the line $\boldsymbol{A}$.

$(\boldsymbol{y} = \mathbf{-2}\boldsymbol{x} + \mathbf{5}) + (\mathbf{2}\boldsymbol{y} = \mathbf{2}\boldsymbol{x} + \mathbf{2}) \Rightarrow \mathbf{3}\boldsymbol{y} = \mathbf{7} \Rightarrow \boldsymbol{y} = \frac{\mathbf{7}}{\mathbf{3}}$, which is a line parallel to the $\boldsymbol{x}$-axis

and passing through all the points $(\boldsymbol{x}, \frac{\mathbf{7}}{\mathbf{3}})$, so the line includes the point $(\frac{\mathbf{4}}{\mathbf{3}}, \frac{\mathbf{7}}{\mathbf{3}})$, too.

Let's this time, take the sum of the line $\boldsymbol{A}$ and the line $\boldsymbol{B}$ multiplied by -1.

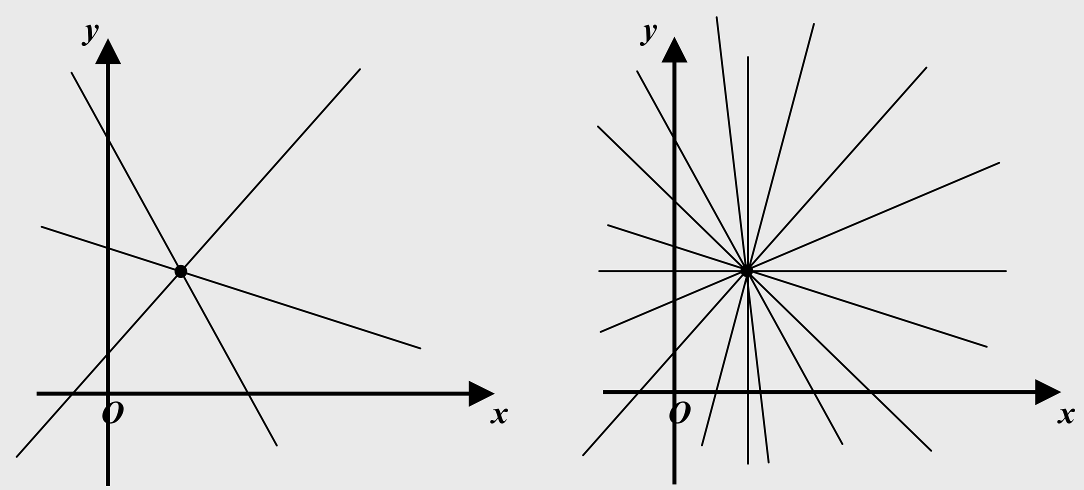
$(\boldsymbol{y} = \boldsymbol{x} + \mathbf{1}) + (\mathbf{-}\boldsymbol{y} = \mathbf{2}\boldsymbol{x} - \mathbf{5}) \Rightarrow \mathbf{0} = \mathbf{3}\boldsymbol{x} - \mathbf{4} \Rightarrow \boldsymbol{x} = \frac{\mathbf{4}}{\mathbf{3}}$, which is a line parallel to the $\boldsymbol{y}$-axis

and passing through all the points $(\frac{\mathbf{4}}{\mathbf{3}}, \boldsymbol{y})$, so the line includes the point $(\frac{\mathbf{4}}{\mathbf{3}}, \frac{\mathbf{7}}{\mathbf{3}})$, too.

Now, it seems that given two lines meeting each other at a point $\boldsymbol{P}$, we can make as many new lines as we want, each of which passes through the point $\boldsymbol{P}$, too. In fact, every linear combination of two lines meeting at a point $\boldsymbol{P}$ passes through the point $\boldsymbol{P}$, too.

So infinitely many such new lines can be made.
We will cover such a fact in one of the examples followed by this section.

Fig. 1



The same is true for the lines parallel to each other, too. So given two lines parallel to each other, we can make infinitely many new lines parallel to the two parallel lines.

That is to say that with any linear combinations of two lines parallel to each other, we can make a new line, and the new line is parallel to the two, also.

So for instance, taking the average of three lines $y = ax + b$, $y = ax + c$, and $y = ax + \frac{b+c}{2}$, which is the average of the first two, we get

$$(y = ax + b) + (y = ax + c) + (y = ax + \frac{b+c}{2}) \Rightarrow 3y = 3ax + \frac{3(b+c)}{2} \Rightarrow y = ax + \frac{b+c}{2}.$$

Of course, adding the average back to the sum and taking the average again, we get the same average. The same is true, too, for two lines meeting at a point. So for instance

$$(y = ax + b) + (y = cx + d) + (y = \frac{a+c}{2}x + \frac{b+d}{2})$$

$$\Rightarrow 3y = \frac{3(a+c)}{2}x + \frac{3(b+d)}{2} \Rightarrow y = \frac{a+c}{2}x + \frac{b+d}{2}.$$

Let's this time, make a new line with three lines parallel to each other as follows

$$y = ax + b,\ y = ax + c, \text{ and } y = ax + d.$$

Then, we get $\{y = ax + b\} + \{y = ax + c\} + \{y = ax + d\} \Rightarrow 3y = 3ax + (b + c + d)$

$\Rightarrow y = ax + \frac{b+c+d}{3}$, which is a line of slope a, which is therefore, parallel to the original three.

And the same is true, also, for three lines meeting each other at a particular point. Suppose s, t, a, b, and c are constant.
Then, three lines $y - t = a(x - s)$, $y - t = b(x - s)$, and $y - t = c(x - s)$ pass through a point (s, t) in the x-y plane, so the three lines meet each other at (that is, share) the point (s, t).

Now, setting $Y = y - t$ and $X = x - s$, we get $Y = aX$, $Y = bX$, and $Y = cX$.

So we get $(Y = aX) + (Y = bX) + (Y = cX) \Rightarrow 3Y = (a + b + c)X \Rightarrow Y = \frac{a+b+c}{3}X$.

Thus, we get $y - t = \frac{a+b+c}{3}(x - s)$, which, too, is a line passing through the point (s, t).

Now, the same idea applies to four or more lines meeting at a particular point, too.
In fact, with any linear combinations of lines meeting at a particular point, we can make a new line, and the new line, also, passes through the particular point.
And the same is true for four or more lines parallel to each other, too.
What then is the point of all the stories above?

The point is 'a point and a slope'. So there are two main ideas in a line.

One is that we can have infinitely many lines meeting at (sharing) a particular point, and each of all those many lines can be made (a sum) of two or more lines, which pass through the particular point, too.   This is about a point in a line.

And the other is that we can have infinitely many lines parallel to each other (sharing a particular slope), and each of all those many lines can be made (a sum) of two or more lines, which are parallel to (share the particular slope with) those many lines, too.

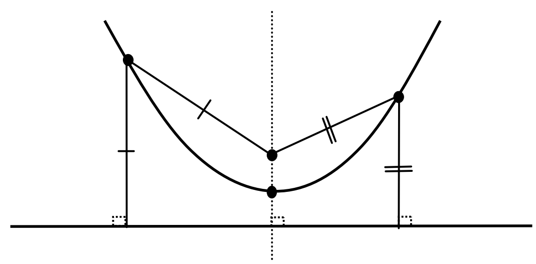
There is another important idea, too.
There is only one line with a particular slope passing through a particular point.
So a point and a slope can define a line.
Thus, a line in math is made of a point and a slope, so to speak.

As other objects in math, a line is a concept, and is not like a string on a guitar or a rope in real world no matter how thin it may be. A line in math has no thickness, and its length is infinite.

Getting the concept of a line can help, and it does very often when we do algebra solving problems with lines when learning not only basic math at high school but higher math at a collage, too, as calculus.

The strong foundation on lines is vital to the early stage of learning calculus.
As a matter of fact, calculus begins with a line.

1.0. What is a parabola?



> A parabola is in a plane, and is a set of points, from each of which, the distance to the focus is the same as the distance to the directrix.

A parabola is one of the basic curves called conic sections, often called conics for short. So it is a conic, and we can call it a path a baseball makes. So it looks like a half circle, but is not, of course. What curve then is it?

A parabola can also be quite close to a half ellipse or a hyperbola. And it can even look like a line or a wedge. So it's often the case it's hard to tell if we just look at it.

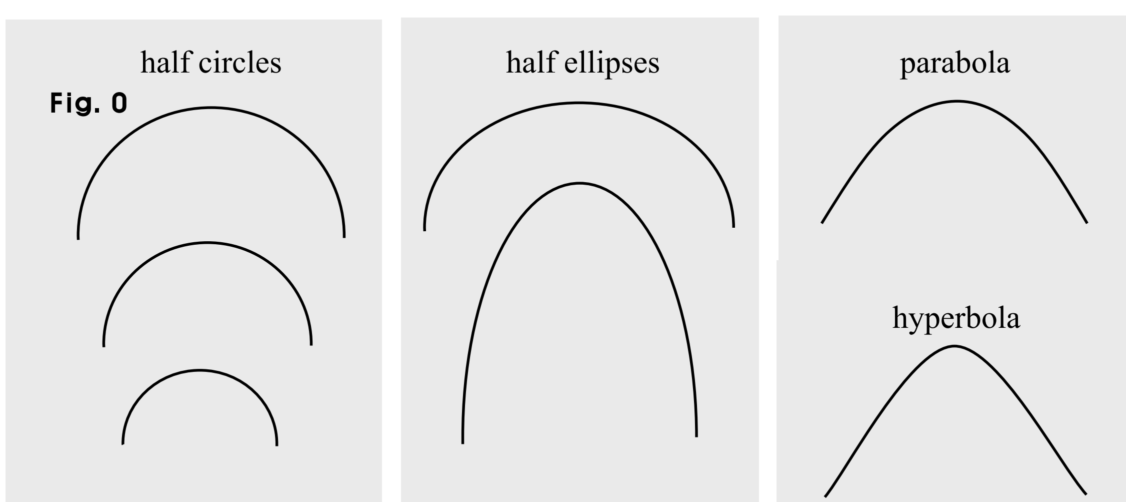
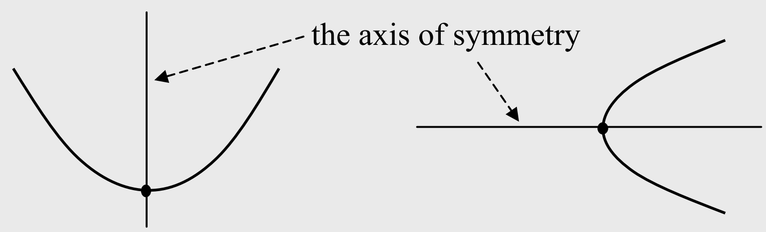


Fig. 0

Next, unlike other conics, each and every line tangent to a parabola has a different direction (slope). Like other conics though, a parabola is symmetric, too.

It is however, symmetric about one line only, and the line is passing though its vertex. So it has only one axis of symmetry, and the axis of symmetry passes through the vertex.

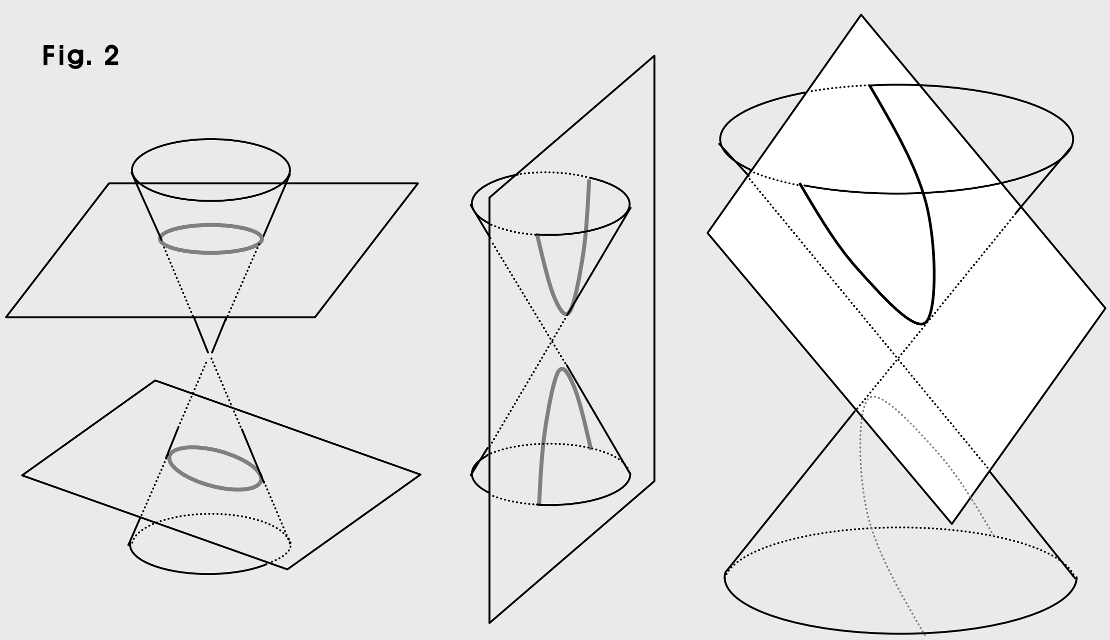
Fig. 1



As shown above, the axis of symmetry passes through the point called the vertex.
So the axis of symmetry has the vertex, which is important.
How then can we get a parabola?

As shown in Fig. 2 below, if cutting a right cone with a plane, we can get a cross section called a parabola. So it's called a conic section, just called a conic for short.

Fig. 2



What is a right cone though?

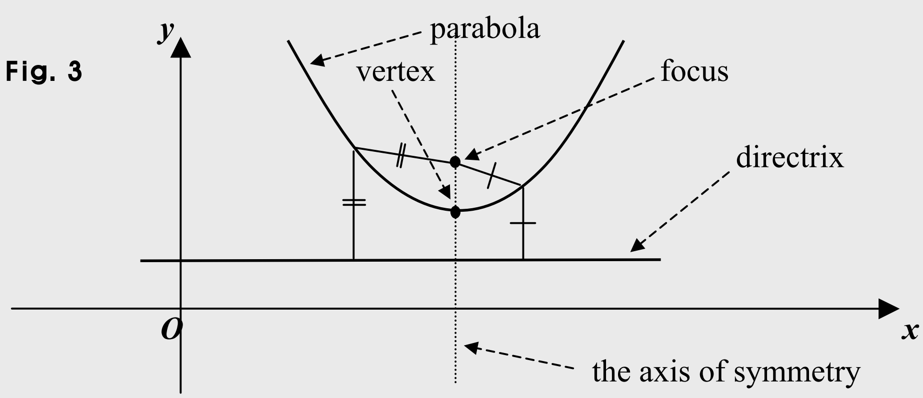
If a cone is a right cone, the line connecting the vertex of the cone and the center of the base is perpendicular to the base, and thus, makes a right angle (90^o) with the base.

And in the figure above, the cross section in black is a parabola. So if the plane is parallel to the side of the cone (called the generator of the cone, too), and does not include the vertex, the cross section is a parabola.

And we can explain a parabola the way as follows.

A parabola is a collection of points as in the case of a line.
Suppose we designate a particular point and a line in the ***x-y*** plane as shown in Fig. 3 below. Then, a parabola is a set of points, from each of which, the distance to the particular point is the same as the distance to the line. And the particular point is called the focus of the parabola, and the line is called the directrix.

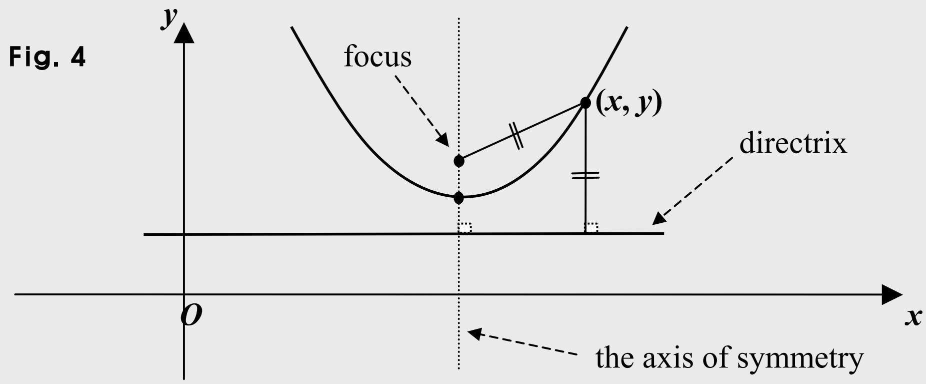
Fig. 3



And we can also explain a parabola the way as follows.

Suppose a point is moving along a curve, and the distance from the moving point to a particular point is the same as the distance to a particular line. Then, the particular point is the focus, the particular line is the directrix, and the curve is a parabola.
So if in the ***x-y*** plane, a point **(*x*, *y*)** is in a parabola, no matter where the point **(*x*, *y*)** may be in the parabola, the distance from the point **(*x*, *y*)** to the point called the focus is the same as the distance to a line called the directrix.

Fig. 4

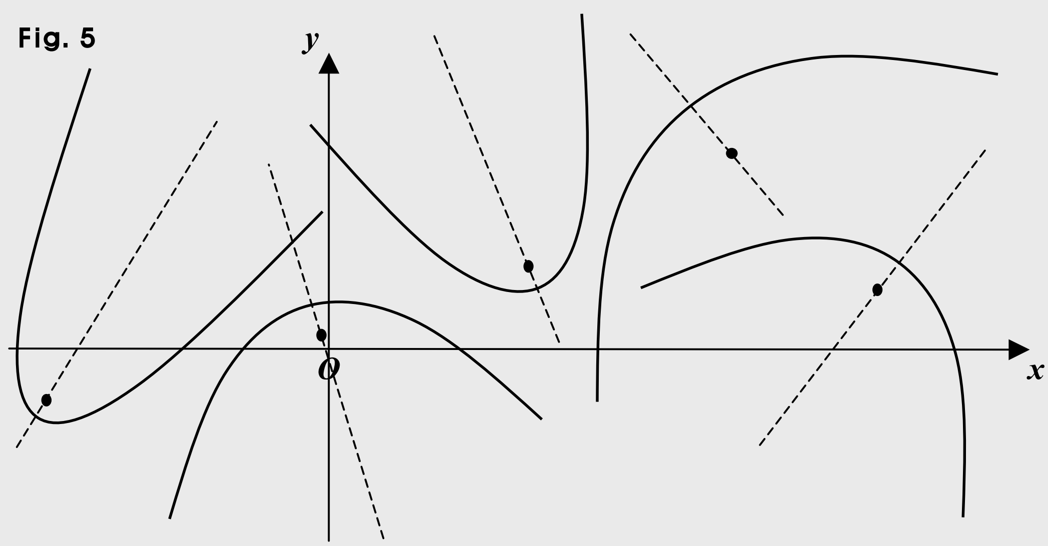


And we can notice that the axis of symmetry is perpendicular to the directrix.

In the figure above, the axis of symmetry is perpendicular to the $\boldsymbol{x}$-axis. And in that case, of course, the directrix is parallel to the $\boldsymbol{x}$-axis.

It can be the case, of course, the axis of symmetry is neither perpendicular nor parallel to the $\boldsymbol{x}$-axis. Such a parabola can be said to be skewed or tilted, and some examples can be as follows.

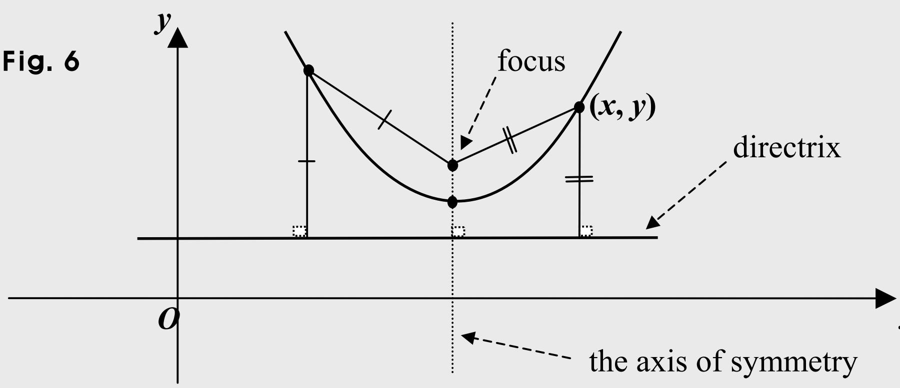
Fig. 5



Usually though, in high school math or basic courses in college math, if we use a parabola, the axis of symmetry is perpendicular to a coordinate axis. So such a parabola can be called a perpendicular parabola. And in this book, such a parabola is said to be perpendicular, and thus, is called a perpendicular parabola. Note however, in other books, it may not be called a perpendicular parabola, and can be called differently.

Anyway, putting a perpendicular parabola in the ***x-y*** plane, we can get this:

Fig. 6



So no matter what point we may take in a parabola, the distance from the point to the focus is the same as the distance to the line called the directrix.

And the directrix is perpendicular to the axis of symmetry.
And the axis of symmetry passes through the focus and the vertex.
So the focus and the vertex are in a line called the axis of symmetry.

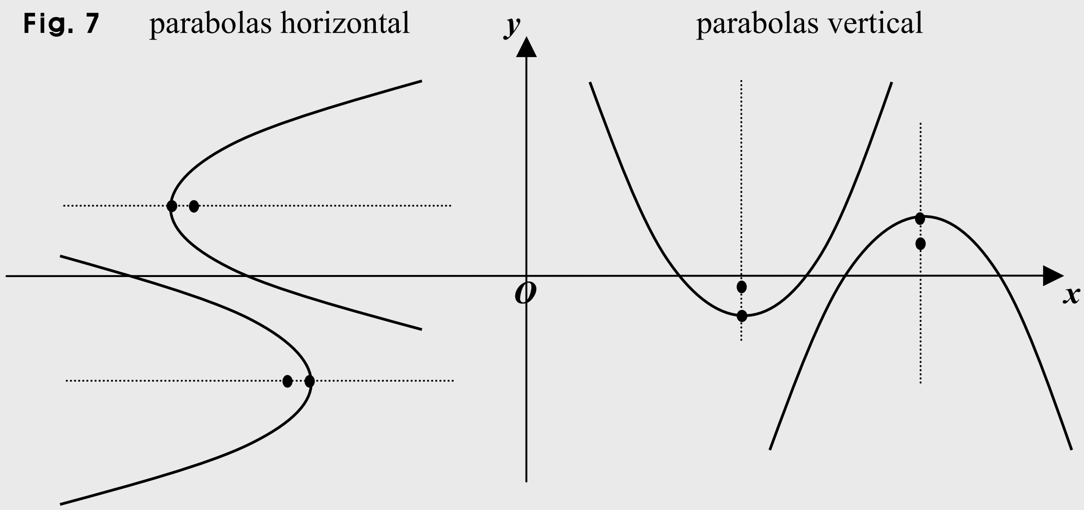
And more importantly, we can see that a curve called a parabola changes its behavior at the point called the vertex. If a parabola is perpendicular, the vertex is the point where a maximum or a minimum occurs.

And using the fact above, we solve many problems with parabolas. That is to say that we often use the idea of parabolas doing problems with maximums or minimums.

And in this book, we have two kinds in parabolas perpendicular.

One is horizontal, and the other is vertical.

Fig. 7



So in a parabola horizontal, the axis of symmetry is horizontal, that is, parallel to the ***x***-axis, and in a parabola vertical, the axis of symmetry is vertical, that is, perpendicular to the ***x***-axis.
Thus, if the axis of symmetry is horizontal, the parabola is horizontal, and if the axis of symmetry is vertical, the parabola is vertical.

(Note that though, in other books, parabolas may not be classified the way above, and thus, may not be said to be horizontal or vertical.)

Either way, the vertex is in the axis of symmetry, and is the key point in a parabola.

Where can we find a parabola though?

We use a parabola a lot. Parabolas have much to do with simple trajectories. That is, a parabola is often used for a path a projectile makes through space under the action of a constant force as gravity. Such a path is called a trajectory, and is a parabola segment. For instance, if we throw a basketball, the basketball travels in the air along a path that is a parabola segment.

For another instance, we use a parabola when we design a tool for watching satellite TV. Such a tool is called a satellite dish. If a plane cuts the dish, the cross section is a parabola segment. And the same is true, too, for a reflector in a flashlight or such. That's because of the property of a parabola. And the property is covered in the next section.

1.1. Properties in Parabolas

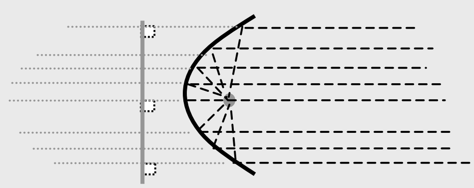
A parabola has a focus, which is a point, and it has a directrix, which is a line.
And the relation between the two is the definition for parabolas, and is as follows.

> No matter what point we may choose in a parabola,
> the distance from the point chosen to the focus is
> the same as the distance from the point chosen to the directrix.

So a parabola is bent toward the focus, and has some properties. One is as follows.
Suppose a parabola reflects lights, and light rays are coming into the parabola, and are perpendicular to the directrix.

Then, all the light rays reflecting off the parabola concentrate at the focus.
And for that reason, too, if we put a light source (such as a light bulb) at the focus, all the light rays reflecting off the parabola will be perpendicular to the directrix.

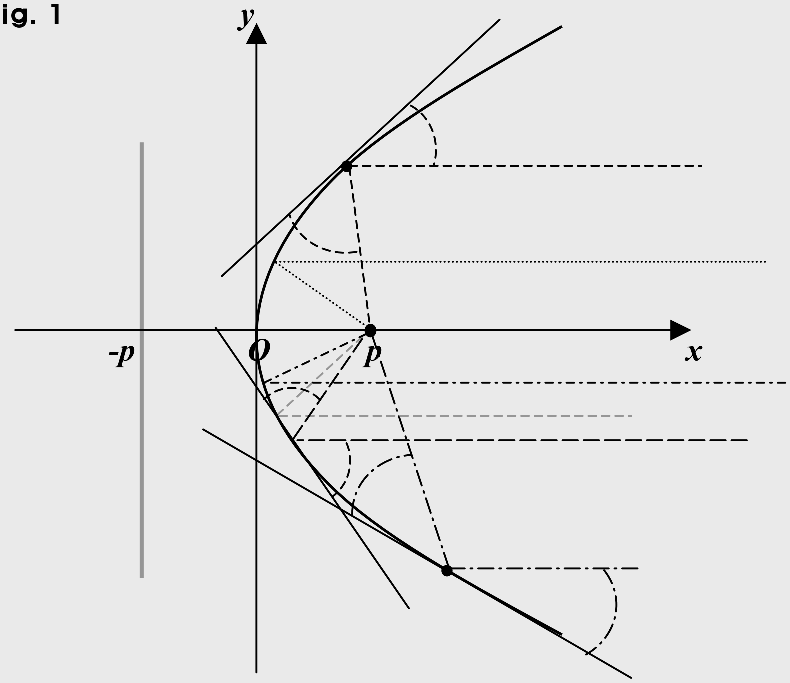
Fig. 0



So rays coming into a parabola and perpendicular to the directrix concentrate at the focus, and reversely, rays emitting from the focus and reflecting off the parabola are perpendicular to the directrix.

That's because when a ray reflects off a parabola, ***the incidence angle*** is the same as ***the angle of reflection***. And the same is true for other waves such as sound and radio, too.

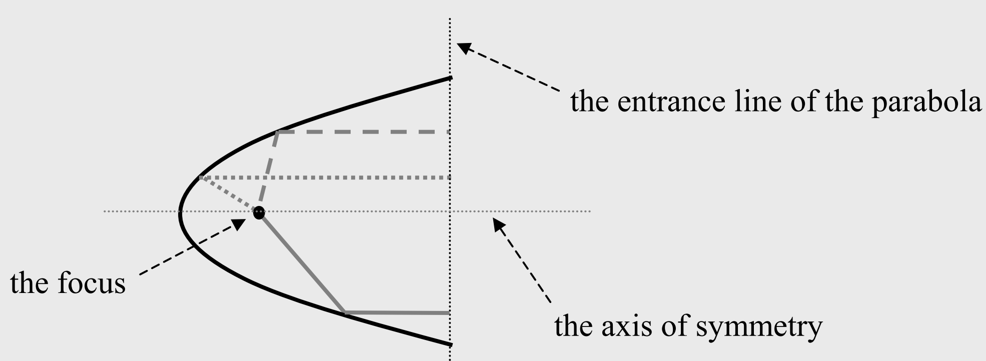
Fig. 1



We can see in Fig. 1, that for each ray reflecting off the parabola, the incidence angle is the same as the angle of reflection. We'll see how it's the case, the proof later.
And we have another property as follows.

Suppose ***A*** is the length of a ray from the entrance line of a parabola to the parabola, and ***B*** is the length of a ray from the parabola to the focus. Then, ***A*** + ***B*** is constant.

Fig. 2



So the sum of the lengths of the two line segments dotted is the same as that of the two line segments solid, and is the same as that of the two line segments dashed, too.

How then can we actually make a parabola?

We know the fact that a parabola is a set of points, from each of which, the distance to the focus is the same as the distance to the directrix.

And because of the fact above, a parabola has the property stated above.

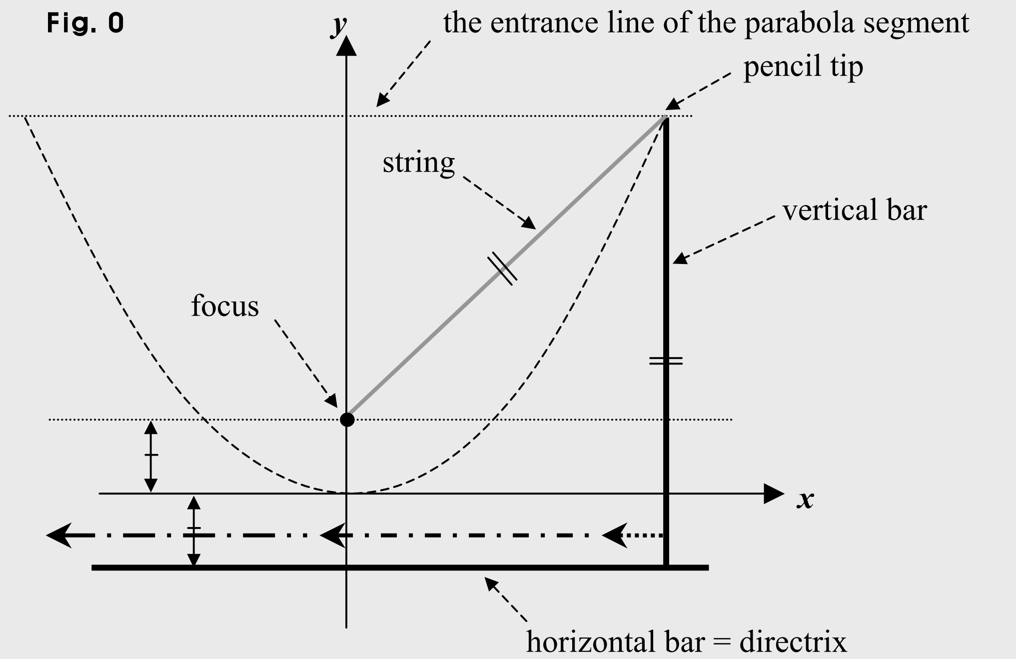
The property is in fact, no other than the definition for parabolas. And using the property, that is, the definition for parabolas, we can actually construct a parabola. And in fact, a German astronomer, Johannes Kepler came up with a method. So the method can be called Kepler's method, which is covered in the next section.

1.2. Kepler's Method

This method is for construction of a parabola, and uses the property saying that the sum of two line segments is constant, one line segment is from the entrance line of a parabola to the parabola, and the other is the line segment from the parabola to the focus.

First, as in Fig. 0 below, pick a position of the focus, and then, put a vertical bar on a horizontal bar that serves as the parabola's directrix.
Next, pin to the focus one end of a string that has the length of the vertical bar, and pin the other end to the top of the vertical bar, and keep the sting taught. The string is in gray in the figure below.
Beginning at the top of the vertical bar, holding the string taught against the bar with a pencil tip, slide the vertical bar along the horizontal bar. A parabola just has begun.

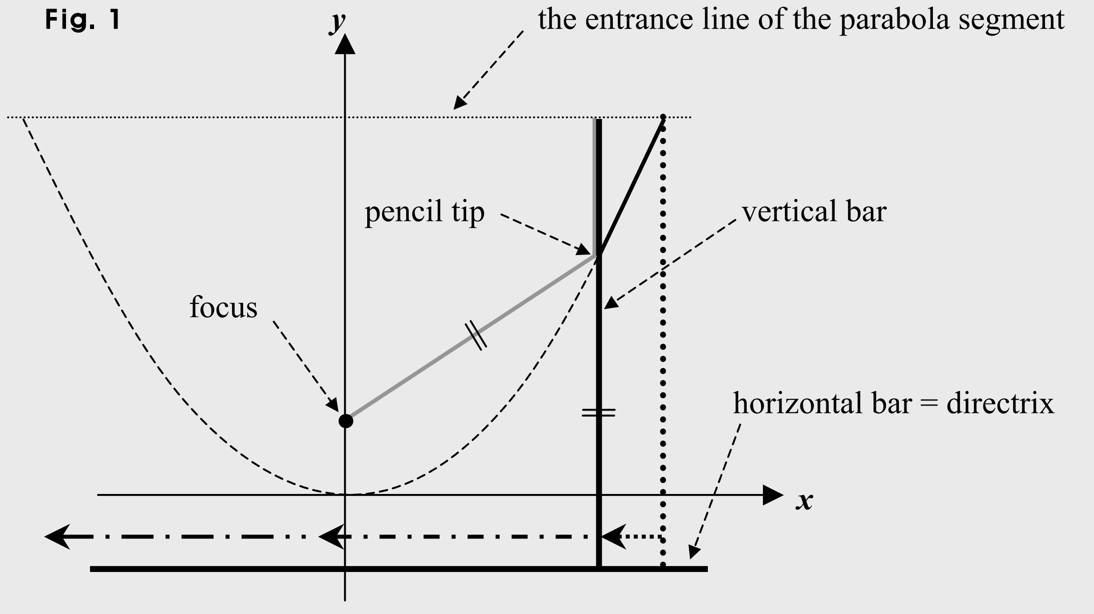
Fig. 0



The distance from the pencil tip to the focus is the same as the one from the pencil tip to the horizontal bar. That is, the length of the string is the same as that of the vertical bar.

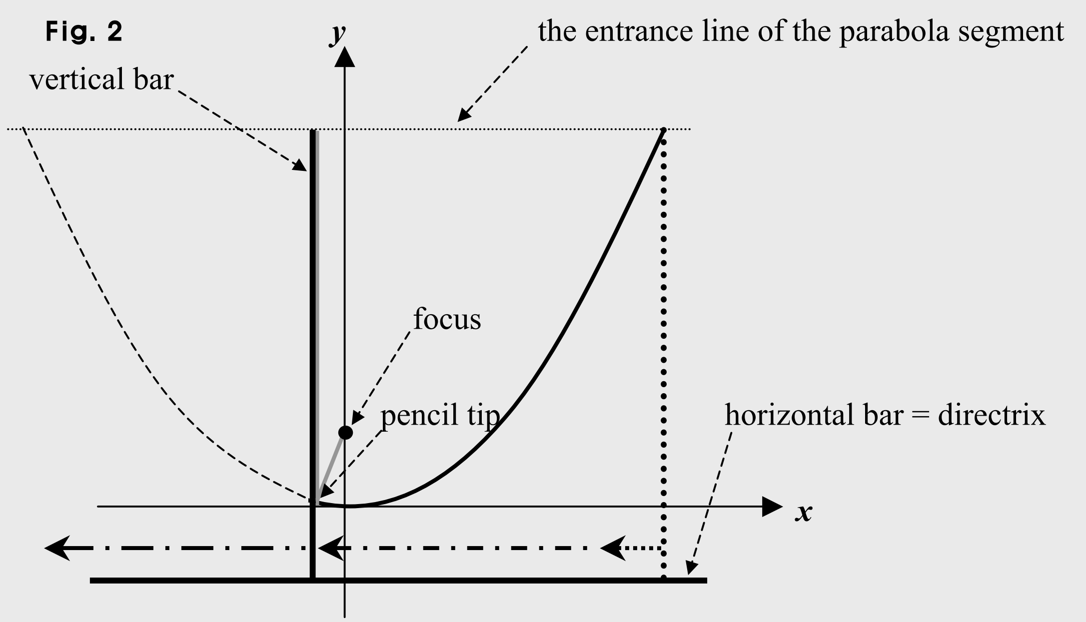
Keep sliding the vertical bar to the left along the horizontal bar while holding the string taught against the vertical bar with the pencil tip. Then, the pencil tip keeps moving downward along the vertical bar, and the parabola segment keeps getting longer.

Fig. 1

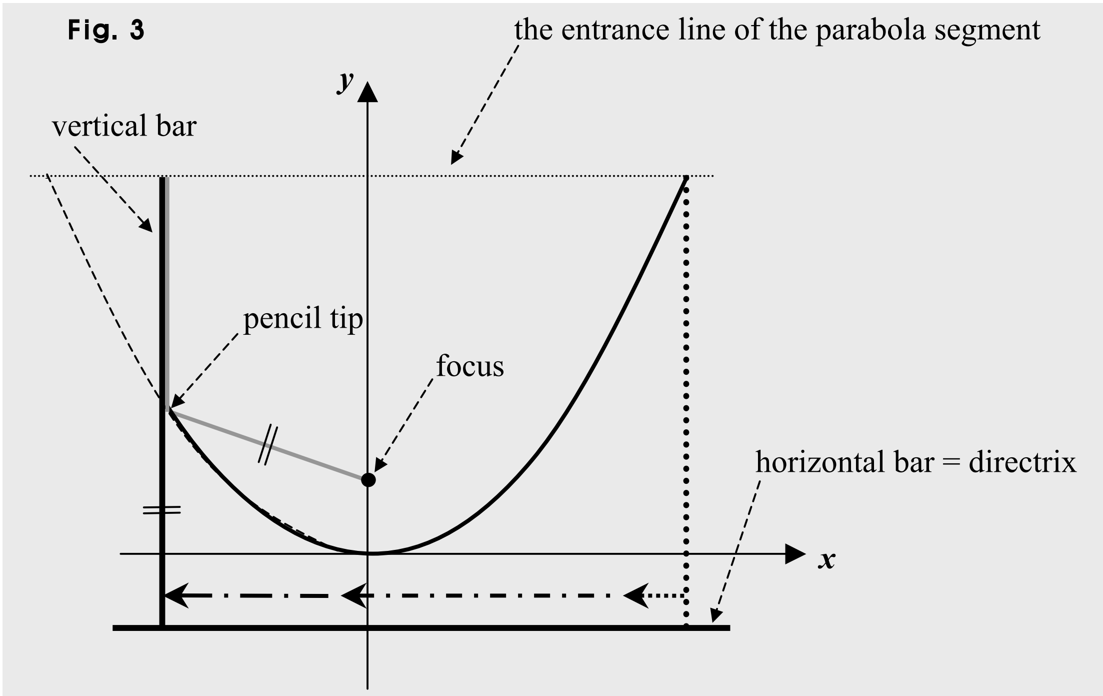


In the figure above, we can see that the pencil tip keeps the same distances from both the focus and the point where the vertical bar meets the horizontal bar.

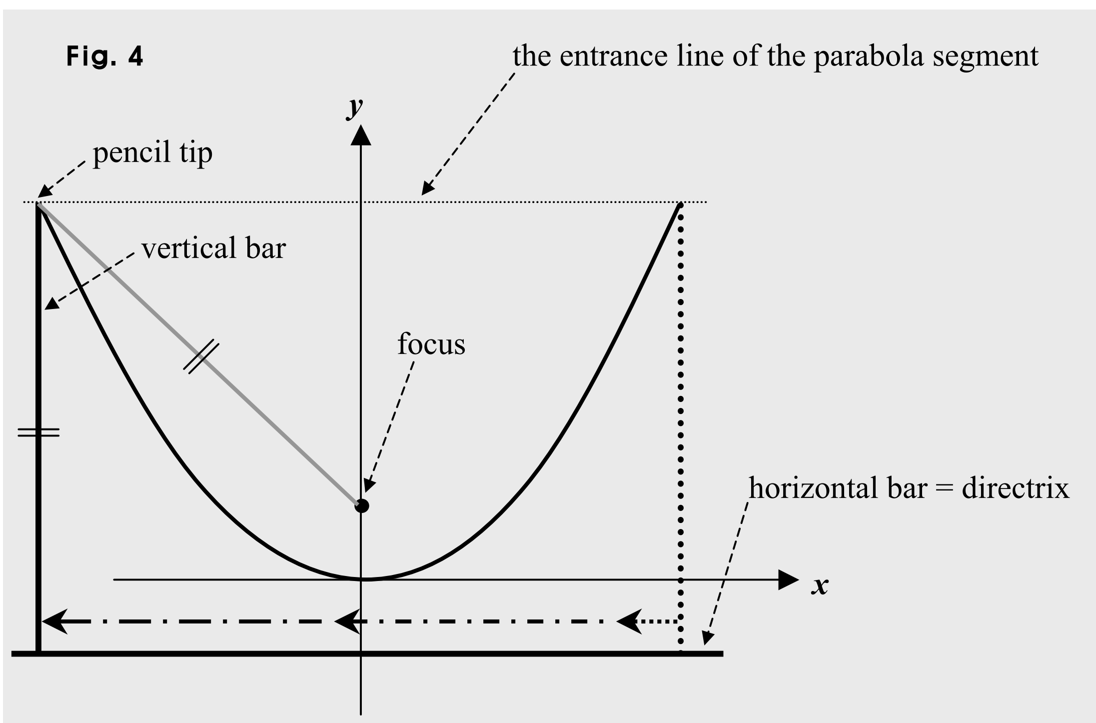
Fig. 2



In the figure above, the sum of the two line segments is the length of the string, and the length is the length of the vertical bar, too. One of the two line segments is from the top of the vertical bar to the pencil tip, and the other is from the pencil tip to the focus.



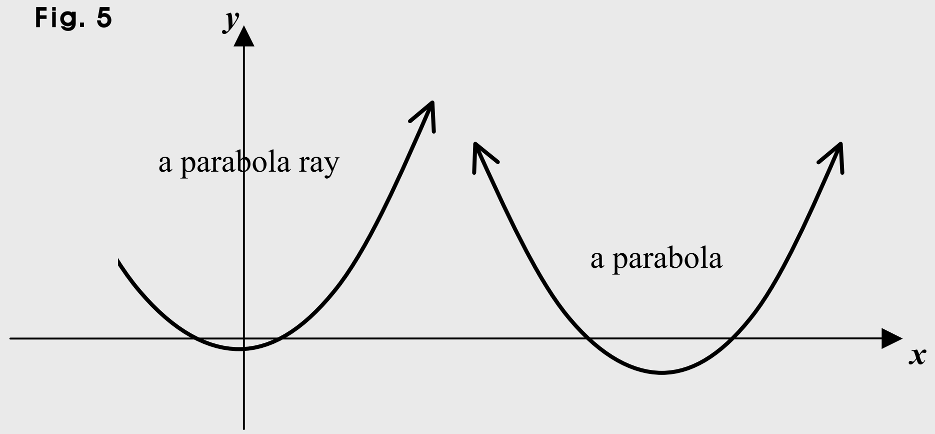
In both figures above and below, the pencil tip still keeps the same distances from both the focus and the point where the vertical bar meets the horizontal bar.



Note that the parabola constructed in the figure above is a parabola segment, and thus, has a finite length.

A parabola is a curve of length infinite, because its both ends proceed infinitely.
And so is a parabola ray, because its one end proceeds infinitely.

Fig. 5



How then can we define a particular parabola?

Normally, we put a parabola in the ***x-y*** plain, and the parabola is perpendicular, so the axis of symmetry is perpendicular to the ***x***-axis or the ***y***-axis. And we can define such a parabola specifying the vertex and the focus, or specifying the vertex and the directrix. That's because knowing the two, we can get the equation of the parabola.

So defining a parabola, we produce the equation of it.
And such equations are covered in several sections beginning the next section.

1.3. Equations for Parabolas 1

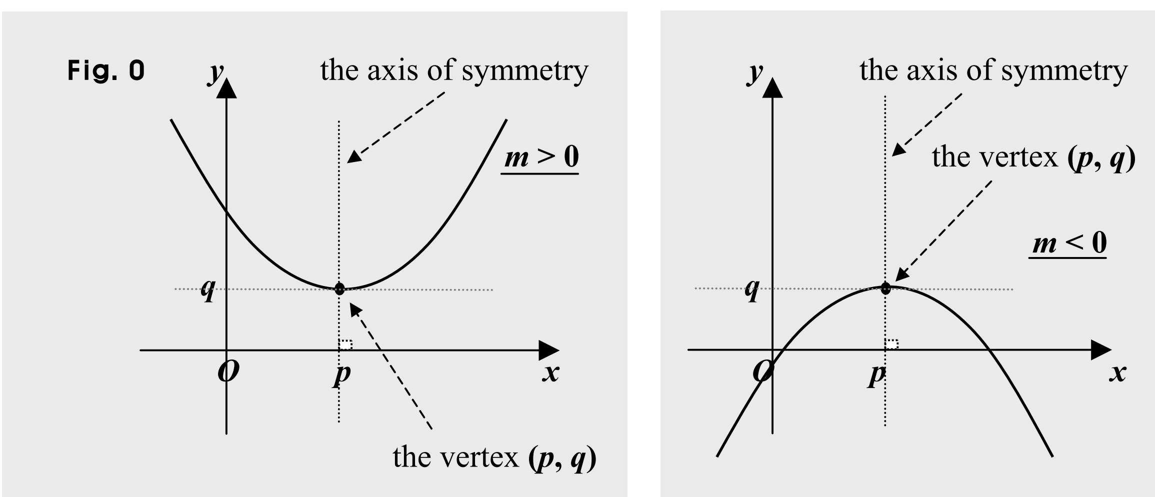
Normally, we put a parabola in the ***x-y*** plain, and in high school math, the parabola is perpendicular, so the axis of symmetry is perpendicular to the ***x***-axis or the ***y***-axis. And producing an equation to indicate such a parabola, we get to use one of two simple forms. In this book, one is called the ***vertex form***, and the other is called the ***non-vertex form***. (So in other books, they might be called differently.)

To begin with, if the axis of symmetry is perpendicular to the ***x***-axis, the vertex form is

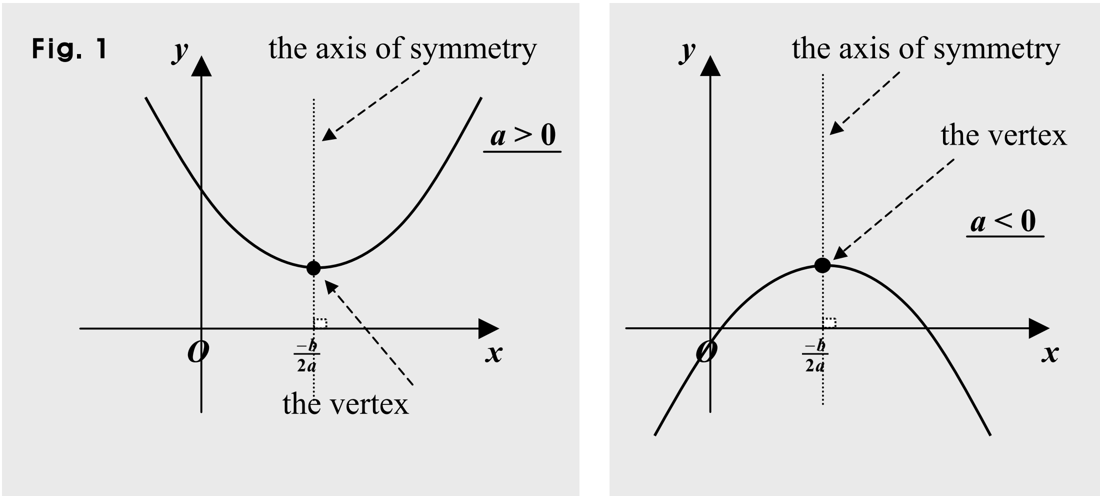
$$y = m(x - p)^2 + q,$$

because (p, q) is the vertex. Of course, we can use it as an equation of a parabola, too.

And the axis of symmetry is a line $x = p$, because the axis of symmetry passes through the vertex, and is perpendicular to the ***x***-axis.



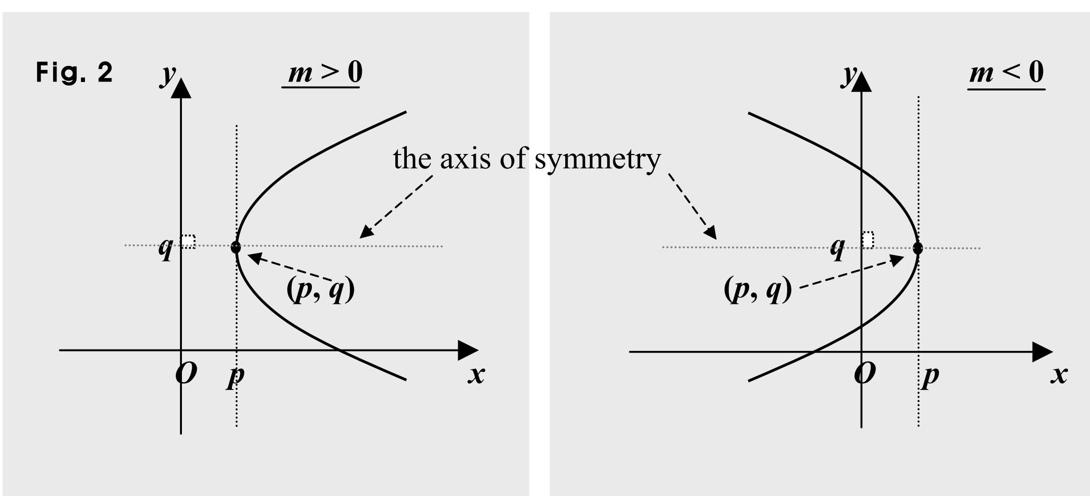
Next, the non-vertex form is $y = ax^2 + bx + c$, which can be also, used as an equation of a parabola, of course. And in this case, the axis of symmetry is a line $x = \dfrac{-b}{2a}$.



And if we use either of the two forms above, the parabola is said to be ***vertical***, because the axis of symmetry is vertical, that is, perpendicular to the $\boldsymbol{x}$-axis.
What then about parabolas horizontal?

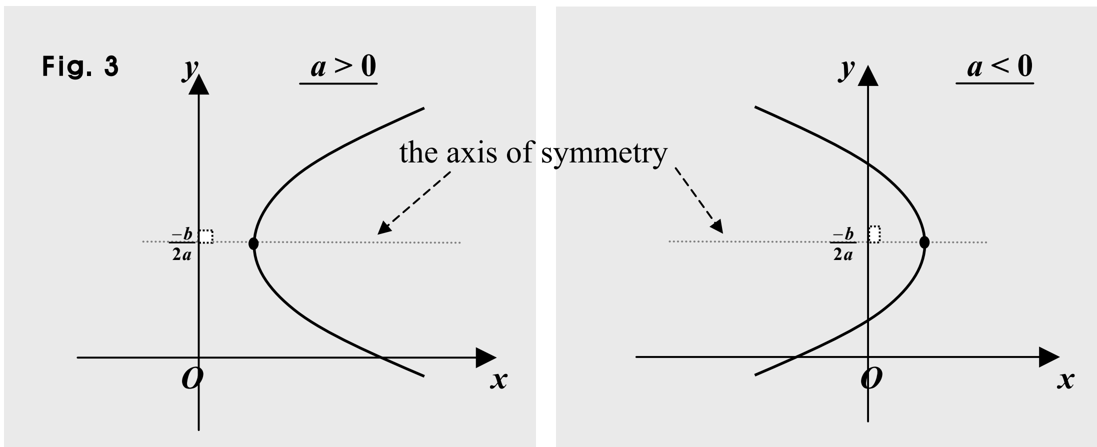
For parabolas horizontal, we use two forms, too, one is the vertex form, and the other is the non-vertex form.

The vertex form is $\boldsymbol{x = m(y - q)^2 + p}$, and the vertex is at $\boldsymbol{(p, q)}$.
And the axis of symmetry is a line $\boldsymbol{y = q}$, because the axis of symmetry includes the vertex, and is horizontal, that is, parallel to the $\boldsymbol{x}$-axis.



And the non-vertex form is $\boldsymbol{x = ay^2 + by + c}$, and in this case, the axis of symmetry is a line $\boldsymbol{y = \frac{-b}{2a}}$.

Fig. 3



So in sum,

$\boldsymbol{y = m(x - p)^2 + q}$ is often called the vertex form, because $\boldsymbol{(p, q)}$ is the vertex, and of course, we can use it as an equation of a parabola, too.

$\boldsymbol{y = ax^2 + bx + c}$ is called the non-vertex form, and also, we can use it as an equation of a parabola. And in this case, the axis of symmetry is a line $\boldsymbol{x = \frac{-b}{2a}}$.

And each form is used if the parabola is ***vertical***, because the axis of symmetry is vertical, that is, perpendicular to the $\boldsymbol{x}$-axis.

If however, the parabola is horizontal, because the axis of symmetry is horizontal, that is, parallel to the $\boldsymbol{x}$-axis, we use either of the two forms as follows.

$\boldsymbol{x = m(y - q)^2 + p}$, which is often called the vertex form, because $\boldsymbol{(p, q)}$ is the vertex.
$\boldsymbol{x = ay^2 + by + c}$, and in this case, the axis of symmetry is a line, $\boldsymbol{y = \frac{-b}{2a}}$.

How then can we get those equations above?

The vertex forms above are from these: $(y - v)^2 = 4p(x - u)$ and $(x - u)^2 = 4p(y - v)$.
And the equations above are in the vertex form, too.
Modifying the forms above, we can get the vertex forms.
And modifying the second of the two above, we can get
$(x - u)^2 = 4p(y - v) \Rightarrow y - v = \frac{1}{4p}(x - u)^2 \Rightarrow y = \frac{1}{4p}(x - u)^2 - v$.

So setting $m = \frac{1}{4p}$, and using k and h as constants, we get $y = m(x - h)^2 + k$.

And also, simplifying an equation in the vertex form above, we can get an equation in the non-vertex form as follows. $y = ax^2 + bx + c$.

How then can we get these equations: $(y - v)^2 = 4p(x - u)$ and $(x - u)^2 = 4p(y - v)$?

A parabola is a set of points. Then, a set of what points?

A parabola is a set of points, from each of which, the distance to a particular point is the same as the distance to a particular line. And the particular point is called the focus, and the particular line is called the directrix.

Thus, no matter what point we may choose in a parabola, the distance from the point to the focus is the same as the distance from the point to directrix.   So?

So using the fact above, we can come up with an equation of a parabola.

And in fact, the form $(x - u)^2 = 4p(y - v)$ is from a form $x^2 = 4py$, which is the simplest vertex form. And setting $p = \frac{1}{4}$, we get $y = x^2$, which is the simplest equation of a parabola, and thus, can be called the prototype in equations for parabolas.

And translating the curve of $x^2 = 4py$ in the amount of u along the x-axis, and in the amount of v along the y-axis, we get the curve of $(x - u)^2 = 4p(y - v)$.

And in the next section, you will get to see how you can get the simplest vertex form as follows. $x^2 = 4py$.

1.4. Vertex Forms 1

Usually, we use as the vertex form, the equation as follows.

$$y = a(x - k)^2 + h, \text{ where } a, k \text{ and } h \text{ are constant}$$

For instance, $y = 2(x - 1)^2 + 3$ is in the vertex form. It's not the only form, though, we can use when producing an equation of a parabola in the vertex form. We can put it this way, too: $y - h = a(x - k)^2$. And some other ways, too.

So for instance, we can have $y = 2x^2$, $y = x^2 + 1$, $3y = x^2 + 1$, $2(y - 3) = 3(x - 1)^2$, etc. It's because looking at the equation only, we can quickly see the vertex. So we have quite a few versions of vertex forms. And in fact, they are all from this form: $x^2 = my$, which can be thus, taken as the most basic or the simplest vertex form.

And we are going to get the form now, going through derivation processes. And the derivation begins with the definition for parabolas. And the definition can be as follows.

A parabola is in a plane, and is a set of points, from each of which, the distance to the focus is the same as the distance to the directrix.

So no matter what point we may choose in a parabola, the distance from the point to the focus is the same as the distance from the point to the directrix.

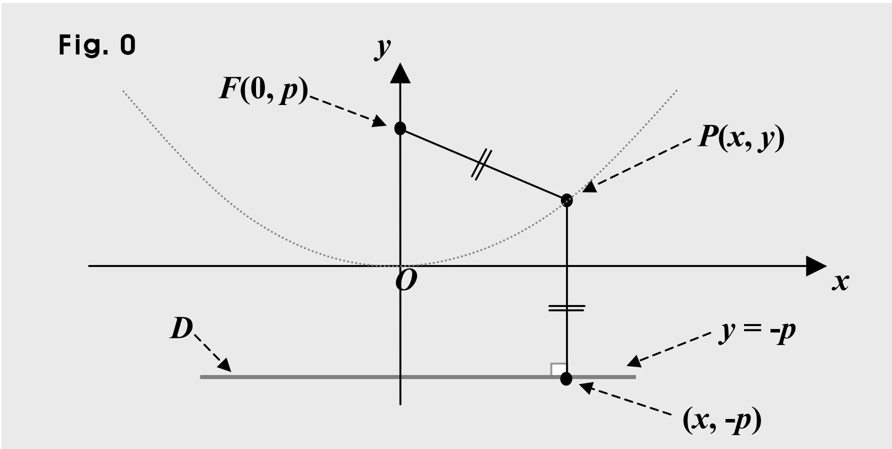
Using thus, the fact above, we can come up with the simplest vertex form.

So suppose first, $F(0, p)$ is a point in the x-y plane, A is a curve, and D is a line $y = -p$.

Suppose next, $P(x, y)$ is an arbitrary point in the curve A. So the point P represents all the points in the curve A.

And suppose now, P is moving along the curve A covering all the points in A, and the point P keeps the same distance away from both the point F and the line D.

Then, capturing a moment while P is moving, we can get a snapshot as follows.



Suppose now, U is the distance between the two points $P(x, y)$ and $F(0, p)$, and V is the distance from the point $P(x, y)$ to the line D.
We can get the distances using the distance formula, often called Pythagorean Theorem.
How then can we apply the distance formula?

Assuming d is the distance between two points, Δx is the difference in x-coordinates, and Δy is the difference in y-coordinates, we can put the formula the way as follows.

$$d^2 = (\Delta x)^2 + (\Delta y)^2$$

So using the distance formula, we can get
$U^2 = (x - 0)^2 + (y - p)^2$ and $V^2 = (x - x)^2 + \{y - (-p)\}^2$.

So we now have $U^2 = x^2 + (y - p)^2$, and $V^2 = (y + p)^2$.

Next, we know P keeps the same distance away from both F and D. So we get
$U^2 = V^2 \Rightarrow x^2 + (y - p)^2 = (y + p)^2 \Rightarrow x^2 + y^2 - 2py + p^2 = y^2 + 2py + p^2 \Rightarrow x^2 = 4py$.

What then did we get?

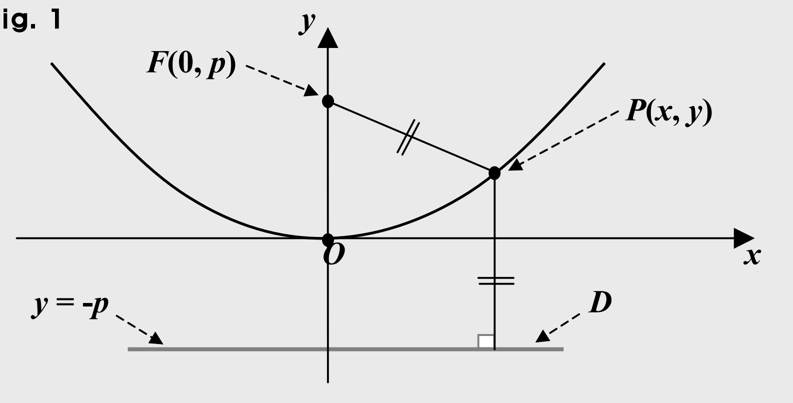
We got an equation expressed in terms of the coordinates of the point $P(x, y)$, that is, we got an equation in terms of x and y. And we call such an equation a connective equation. So in this case, the connective equation is the equation that connects the two variables used as the coordinates of the arbitrary point $P(x, y)$ in the curve A.

The equation explains every point in the curve ***A***, that is, it indicates the curve ***A***.
And we call ***A*** a parabola. So the equation is called the equation of the parabola ***A***.

Thus the equation of the parabola ***A*** is $x^2 = 4py$.

And assuming $p > 0$, and putting the parabola ***A*** in the ***x-y*** plane, we get this:

Fig. 1



And ***F*** is called the focus, and ***D*** is called the directrix.
So the focus is at $(0, p)$, and the directrix is a line $y = -p$, which is parallel to the ***x***-axis.

And in the case above, we can notice that the point at the origin is ***closest*** to the focus, and is the ***midpoint*** between the directrix and the focus. And we call the point at the origin ***O***, the ***vertex***. So the vertex of the parabola $x^2 = 4py$ is at the origin $O(0, 0)$.
What then is the vertex about?

First, the vertex of a parabola is the ***midpoint*** between the directrix and the focus.

Next, of all the points in a parabola, the vertex is the ***closest*** to the focus, and the directrix, too, of course.

The vertex is the point where the parabola changes its behavior, and is often called the maximum or the minimum point.
Also, the vertex is the point the axis of symmetry passes through.
So the vertex is an important point, and is a critical point in a parabola.
What then about ***p*** at the focus $(0, p)$?

The parabola $\boldsymbol{A}$ has the focus, and we know that the vertex in the parabola $\boldsymbol{A}$ is at $\boldsymbol{O}(0, 0)$.
So we can notice that $\boldsymbol{p}$ is the distance from the focus and the vertex.
Thus, $\boldsymbol{p}$ can be called the ***focal distance*** of the parabola $\boldsymbol{A}$.

And of course, since the directrix of the parabola $\boldsymbol{A}$ is the line $\boldsymbol{y = -p}$, and the vertex is the midpoint between the focus and the directrix, the distance from the vertex to the directrix is $\boldsymbol{p}$, too.

Next, setting $\boldsymbol{p = \frac{1}{4}}$ in $\boldsymbol{x^2 = 4py}$, we get $\boldsymbol{y = x^2}$, which is quite familiar, isn't it?

The equation $\boldsymbol{y = x^2}$ is the simplest of all equations indicating parabolas.

And all the variations can be easily made from the equation $\boldsymbol{y = x^2}$.

So we can call it the prototype in parabolas. (Note however, that in no other book, it is called that way.)
And we can put $\boldsymbol{x^2 = 4py}$ this way: $\boldsymbol{y = \frac{1}{4p}x^2}$. Then, setting $\boldsymbol{a = \frac{1}{4p}}$, we get $\boldsymbol{y = ax^2}$.

What then is the vertex of the parabola $\boldsymbol{y = x^2}$?

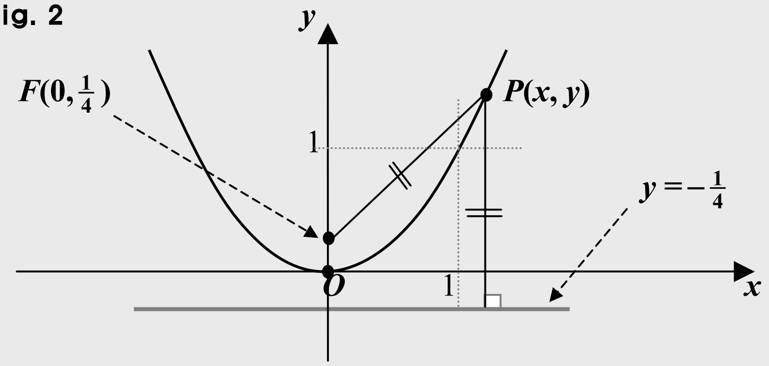
We know that the vertex is the midpoint between the focus and the directrix.
And we know that the focus of $\boldsymbol{x^2 = 4py}$ is $\boldsymbol{(0, p)}$, and the directrix is $\boldsymbol{y = -p}$.
So the vertex of $\boldsymbol{x^2 = 4py}$ is at $\boldsymbol{O}(0, 0)$.

And we know that setting $\boldsymbol{p} = \frac{1}{4}$ in $\boldsymbol{x^2 = 4py}$, we get $\boldsymbol{y = x^2}$.

So the focus of $\boldsymbol{y = x^2}$ is $\boldsymbol{(0, \frac{1}{4})}$, and the directrix is a line $\boldsymbol{y = -\frac{1}{4}}$.

And the midpoint between the focus and the directrix above is $\boldsymbol{O}(0, 0)$, the origin.
So as in the case of $\boldsymbol{4py = x^2}$, the vertex of the parabola $\boldsymbol{y = x^2}$ is at the origin $\boldsymbol{O}(0, 0)$, too.

Fig. 2

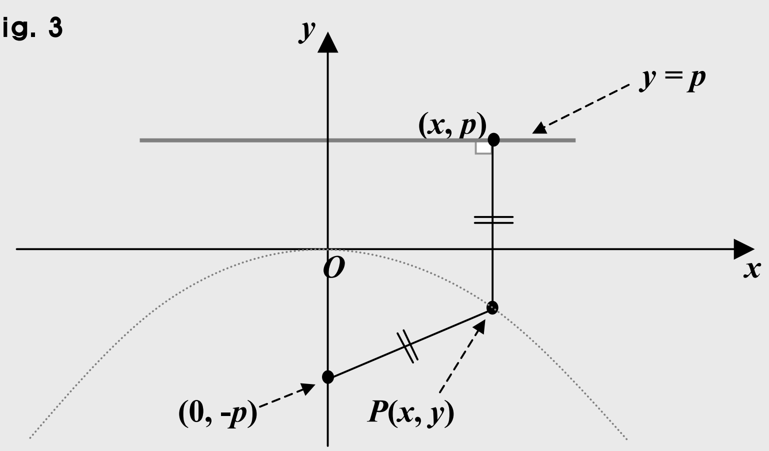


Also, the parabola above is symmetric about the y-axis.
And the axis of symmetry is perpendicular to the directrix, and includes the vertex (0, 0).
Thus, the axis of symmetry of $y = x^2$ is a line $x = 0$, which is the y-axis. Besides, in this case, the parabola is said to be open upwardly, and is often said to be ***concave-up***.
Now, what if this time, the directrix is a line $y = p$, and the focus is at $(0, -p)$?

Then, the parabola is $x^2 = -4py$. How then can we get the equation?

Assuming first, ***B*** is the parabola, and ***P*(*x*, *y*)** is an arbitrary point in the parabola ***B***, we can say that the point ***P*** represents all the points in the parabola ***B***.
So next, assuming now, ***P*** is moving along the parabola ***B*** covering all the points in ***B***, we can say that the point ***P*** keeps the same distance away from both the focus $(0, -p)$ and the directrix $y = p$.
Then, capturing a moment while ***P*** is moving, we can get a snapshot as follows.

Fig. 3



Then again, assuming $\boldsymbol{U}$ is the distance from $\boldsymbol{P(x, y)}$ to the focus $\mathbf{(0, -}\boldsymbol{p}\mathbf{)}$, and $\boldsymbol{V}$ is the distance between $\boldsymbol{P(x, y)}$ and the directrix $\boldsymbol{y = p}$, we get

$$U^2 = (x - 0)^2 + \{y - (-p)\}^2 = x^2 + (y + p)^2, \text{ and } V^2 = (x - x)^2 + (y - p)^2 = (y - p)^2.$$

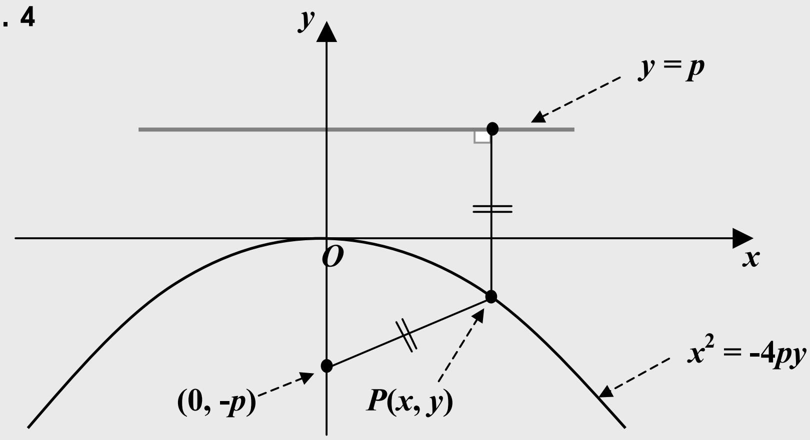
Next, we know $\boldsymbol{P}$ keeps the same distance away from the focus and directrix. So we get

$$U^2 = V^2 \Rightarrow x^2 + (y + p)^2 = (y - p)^2 \Rightarrow x^2 + y^2 + 2py + p^2 = y^2 - 2py + p^2 \Rightarrow x^2 = -4py.$$

Thus, the connective equation between $\boldsymbol{x}$ and $\boldsymbol{y}$ is $\boldsymbol{x^2 = -4py}$, and $\boldsymbol{x}$ and $\boldsymbol{y}$ are the coordinates of the arbitrary point $\boldsymbol{P(x, y)}$ of the parabola $\boldsymbol{B}$.

So $\boldsymbol{x^2 = -4py}$ is the equation of the parabola $\boldsymbol{B}$, which is in the figure below.

Fig. 4



And of course, in this case, too, we can put $\boldsymbol{x^2 = -4py}$ this way: $\boldsymbol{y = -\frac{1}{4p}x^2}$.

Then, setting $\boldsymbol{a = -\frac{1}{4p}}$, we get $\boldsymbol{y = ax^2}$, which is quite familiar, too, isn't it?

And next, setting $\boldsymbol{p = \frac{1}{4}}$ in $\boldsymbol{x^2 = -4py}$ or in $\boldsymbol{y = -\frac{1}{4p}x^2}$, we get $\boldsymbol{y = -x^2}$.

What then is the vertex of the parabola $\boldsymbol{y = -x^2}$?

We know that the vertex is the midpoint between the focus and the directrix.
And we know that the focus of $x^2 = -4py$ is $(0, -p)$, and the directrix is $y = p$.

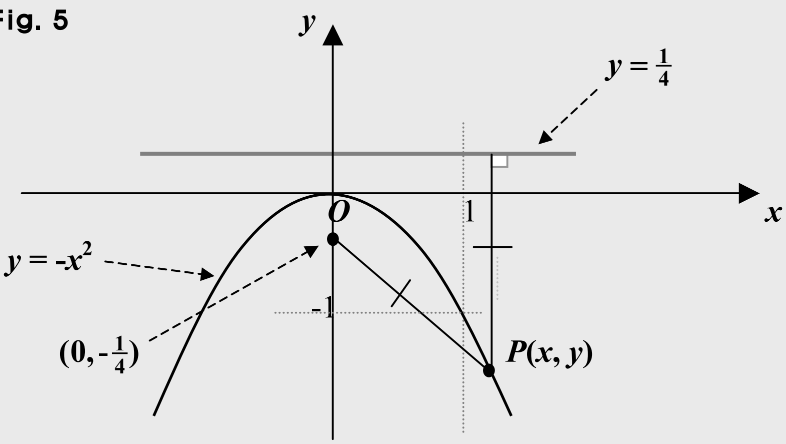
So the vertex of $x^2 = -4py$ is at $O(0, 0)$.

And we know that setting $p = \frac{1}{4}$ in $x^2 = -4py$, we get $y = -x^2$.

So the focus of $y = -x^2$ is $(0, -\frac{1}{4})$, and the directrix is a line $y = \frac{1}{4}$.
And the midpoint between the focus and the directrix above is $O(0, 0)$, the origin.

So as in the case of $-4py = x^2$, that is, $4py = -x^2$, the vertex of the parabola $y = -x^2$ is at the origin $O(0, 0)$, too.

Fig. 5

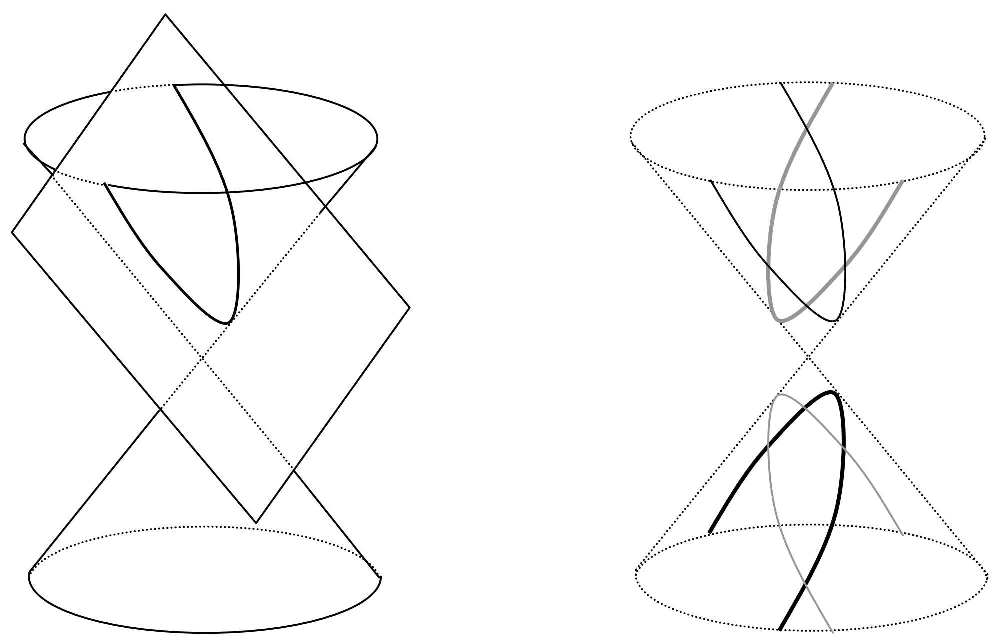


Also, the parabola above is symmetric about the y-axis.

And the axis of symmetry is perpendicular to the directrix, and includes the vertex (0, 0).

Thus, the axis of symmetry of $y = -x^2$ is a line $x = 0$, too, which is the y-axis.

Besides, in this case, the parabola is said to be open downwardly, and is often said to be ***concave-down***.



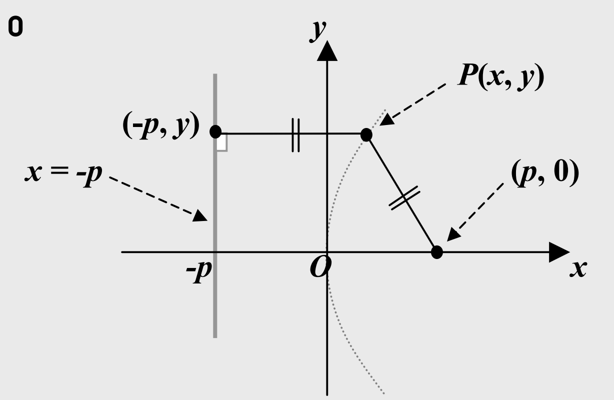
1.5. Vertex Forms 2

A parabola is a set of points, from each of which, the distance to a particular point called the focus is the same as the distance to a particular line called the directrix.
So no matter what point we may choose in a parabola, the distance from the point to the focus is the same as the distance from the point to the directrix.
Using thus, the fact above, we can come up with an equation of a parabola.

Suppose this time, the focus is at $(p, 0)$, and the directrix is a line $x = -p$.
Then, the parabola is $y^2 = 4px$. How then can we get the equation?

Assuming first, A is the parabola, and $P(x, y)$ is an arbitrary point in the parabola A, we can say that the point P represents all the points in the parabola A.
So next, assuming now, P is moving along the parabola A, and taking a snapshot of P, we can put it the way as follows.

Fig. 0



Next, assuming U is the distance from $P(x, y)$ to the focus $(p, 0)$, and V is the distance between $P(x, y)$ and the directrix $x = -p$, we get

$U^2 = (x - p)^2 + (y - 0)^2 = (x - p)^2 + y^2$, and $V^2 = \{x - (-p)\}^2 + (y - y)^2 = (x + p)^2$.

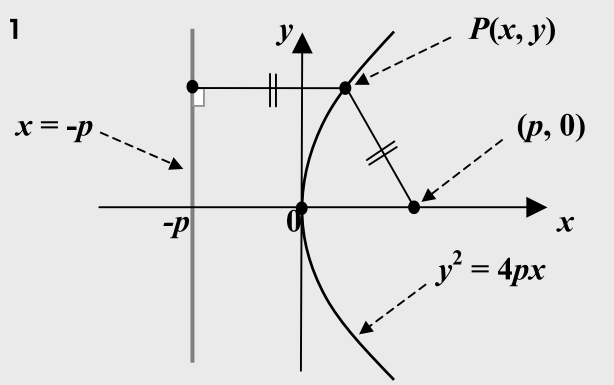
And next, since $\boldsymbol{P}$ keeps the same distance away from both the focus and the directrix, we get

$$U^2 = V^2 \Rightarrow (x-p)^2 + y^2 = (x+p)^2 \Rightarrow x^2 - 2px + p^2 + y^2 = x^2 + 2px + p^2 \Rightarrow y^2 = 4px.$$

So the connective equation between x and y is $y^2 = 4px$, and x and y are the coordinates of the arbitrary point $P(x, y)$ of the parabola A.

Thus, $y^2 = 4px$ is the equation of the parabola A, which is in the figure below.

Fig. 1



And of course, we can put $y^2 = 4px$ this way: $x = \dfrac{1}{4p}y^2$.

Then, setting $a = \dfrac{1}{4p}$, we get $x = ay^2$.

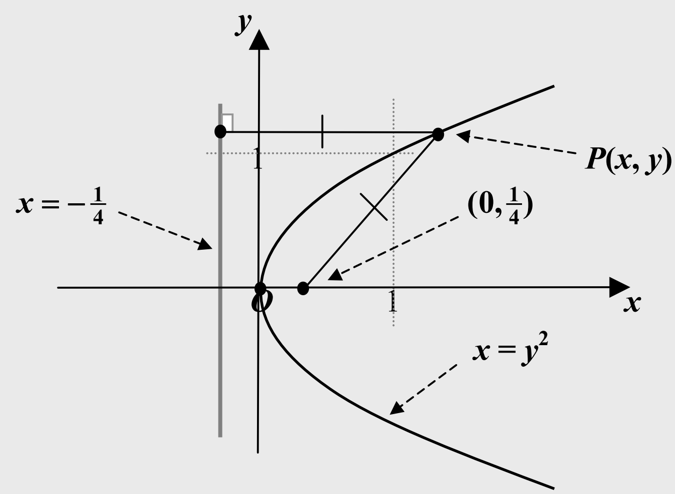
And next, setting $p = \dfrac{1}{4}$ in $y^2 = 4px$ or in $x = \dfrac{1}{4p}y^2$, we get $x = y^2$.
What then is the vertex of the parabola $x = y^2$?

We know that the vertex is the midpoint between the focus and the directrix.
And we know that the focus of $y^2 = 4px$ is $(p, 0)$, and the directrix is $x = -p$.
So the vertex of $y^2 = 4px$ is at $O(0, 0)$.
And we know that setting $p = \dfrac{1}{4}$ in $y^2 = 4px$, we get $x = y^2$.

So the focus of $x = y^2$ is $(0, \dfrac{1}{4})$, and the directrix is a line $x = -\dfrac{1}{4}$.

And the midpoint between the focus and the directrix above is $\boldsymbol{O}(0, 0)$, the origin.
So as in the case of $\boldsymbol{y^2 = 4px}$, the vertex of the parabola $\boldsymbol{x = y^2}$ is at the origin $\boldsymbol{O}(0, 0)$, too.

Fig. 2



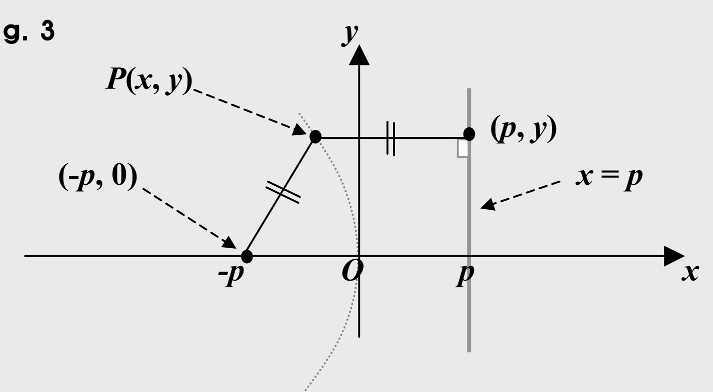
Also, the parabola above is symmetric about the $\boldsymbol{x}$-axis.
And the axis of symmetry is perpendicular to the directrix, and includes the vertex (0, 0).

Thus, the axis of symmetry of $\boldsymbol{x = y^2}$ is a line $\boldsymbol{y = 0}$, too, which is the $\boldsymbol{x}$-axis. Besides, in this case, the parabola is said to be open to the right, and can be said to be ***concave-right***.
What if this time, the focus is at $\boldsymbol{(-p, 0)}$, and the directrix is a line $\boldsymbol{y = p}$?

Then, the parabola is $\boldsymbol{y^2 = -4px}$.   How then, can we get the equation?

Assuming again, ***B*** is the parabola, and $\boldsymbol{P(x, y)}$ is an arbitrary point in the parabola ***B***, and is now, moving along the parabola ***B***, we can put a snapshot of ***P*** the way below.

Fig. 3



Next, assuming $\boldsymbol{U}$ is the distance from $\boldsymbol{P(x, y)}$ to the focus $\mathbf{(-}\boldsymbol{p}\mathbf{, 0)}$, and $\boldsymbol{V}$ is the distance between $\boldsymbol{P(x, y)}$ and the directrix $\boldsymbol{x = p}$, we get

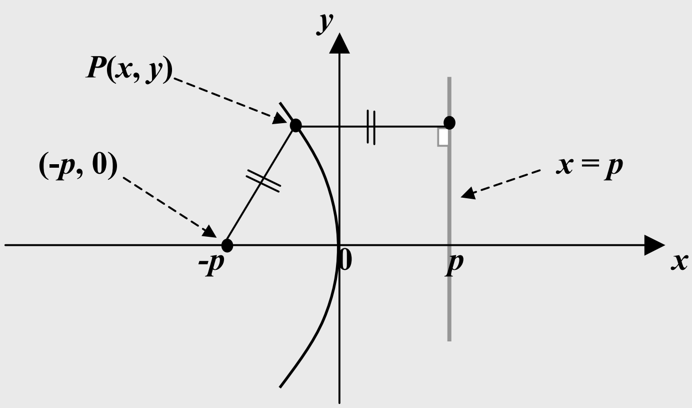
$U^2 = \{x - (-p)\}^2 + (y - 0)^2 = (x + p)^2 + y^2$, and $V^2 = (x - p)^2 + (y - y)^2 = (x - p)^2$.

And next, since both distances are the same, we get

$U^2 = V^2 \Rightarrow (x + p)^2 + y^2 = (x - p)^2 \Rightarrow x^2 + 2px + p^2 + y^2 = x^2 - 2px + p^2 \Rightarrow y^2 = -4px$.

So the connective equation between $\boldsymbol{x}$ and $\boldsymbol{y}$ is $y^2 = -4px$, and $\boldsymbol{x}$ and $\boldsymbol{y}$ are the coordinates of the arbitrary point $\boldsymbol{P(x, y)}$ of the parabola $\boldsymbol{B}$.
Thus, $y^2 = -4px$ is the equation of the parabola $\boldsymbol{B}$, which is in the figure below.

Fig. 4



And of course, in this case, too, we can put $y^2 = -4px$ this way: $x = -\dfrac{1}{4p}y^2$.

Then, setting $a = -\dfrac{1}{4p}$, we get $x = ay^2$.

And next, setting $p = \dfrac{1}{4}$ in $y^2 = -4px$ or in $x = -\dfrac{1}{4p}y^2$, we get $x = -y^2$.

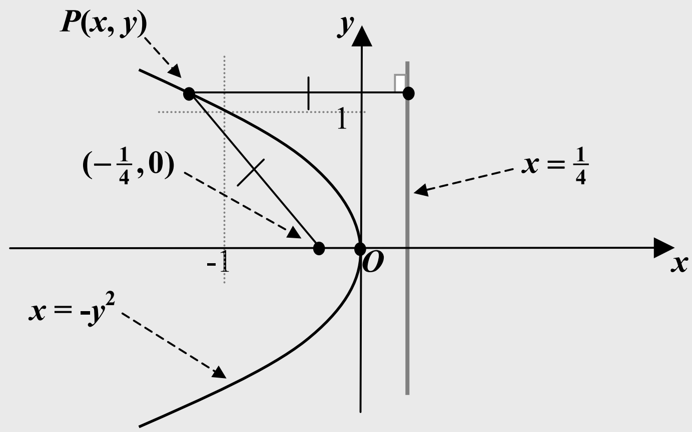
What then is the vertex of the parabola $x = -y^2$?

We know that the vertex is the midpoint between the focus and the directrix.
And we know that the focus of $y^2 = -4px$ is $(-p, 0)$, and the directrix is $x = p$.
So the vertex of $y^2 = -4px$ is at $\boldsymbol{O}(0, 0)$.
And we know that setting $p = \dfrac{1}{4}$ in $y^2 = -4px$, we get $x = -y^2$.

So the focus of $\boldsymbol{x} = \textbf{-}\boldsymbol{y}^2$ is $(\mathbf{0},\ -\frac{\mathbf{1}}{\mathbf{4}})$, and the directrix is a line $\boldsymbol{x} = \frac{\mathbf{1}}{\mathbf{4}}$.
And the midpoint between the focus and the directrix above is (0, 0), the origin.

So as in the case of $\boldsymbol{y}^2 = \mathbf{-4}\boldsymbol{px}$, the vertex of the parabola $\boldsymbol{x} = \textbf{-}\boldsymbol{y}^2$ is at the origin (0, 0), too.

Fig. 5



Also, the parabola above is symmetric about the $\boldsymbol{x}$-axis.

And the axis of symmetry is perpendicular to the directrix, and includes the vertex (0, 0). Thus, the axis of symmetry of $\boldsymbol{x} = \textbf{-}\boldsymbol{y}^2$ is a line $\boldsymbol{y} = \mathbf{0}$, too, which is the $\boldsymbol{x}$-axis. Besides, in this case, the parabola is said to be open to the left, and can be said to be ***concave-left***.

1.6. Vertex Forms 3

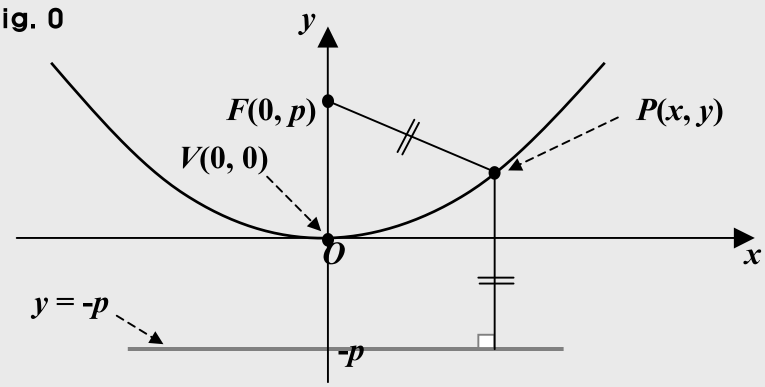
To begin with, putting a parabola in an equation, we can get

$x^2 = 4py$, where p is constant. What parabola then is it?

It is a parabola vertical, the vertex is the origin $O(0, 0)$, the axis of symmetry is $x = 0$, that is, the y-axis, the focus is at $(0, p)$, and the directrix is a line $y = -p$.

And assuming $p > 0$, V is the vertex, F is the focus, and A is the parabola, we can put them all in the x-y plane the way as follows.

Fig. 0



What if its vertex is at a point other than the origin?

For instance, moving (translating) A so that the vertex is at (2, 1), we get a new parabola, so we get a new equation. What then will the new equation be?

Assuming the new parabola is T, we can get T translating the parabola A in the amount of 2 in the direction of the x-axis, and in the amount of 1 in the direction of the y-axis.

Note that the two parabolas themselves are the same, but have different equations.

Fig. 1

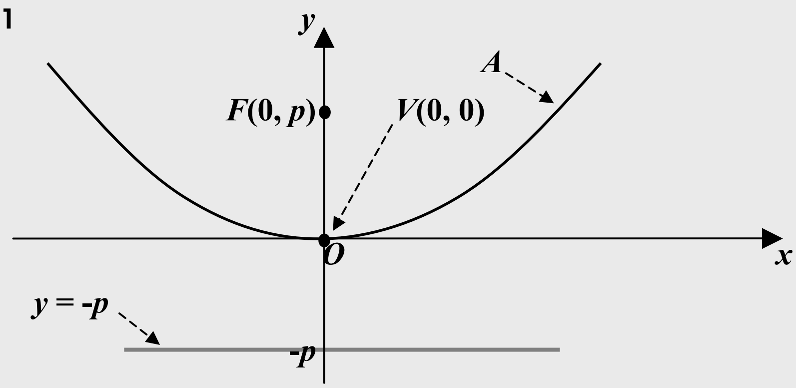
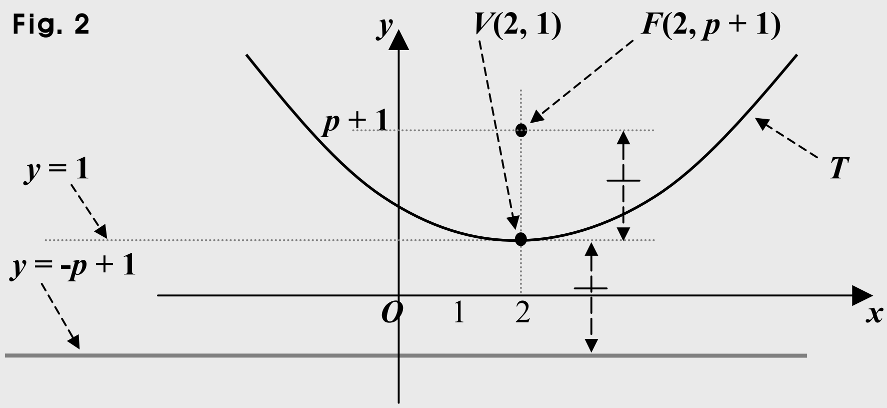


Fig. 2



Then, the equation of the parabola T is $(x-2)^2 = 4p(y-1)$ where $p > 0$.

(Note that we still get $4p$, because the focal distance p did not change. Translations do not change shapes, so the focal distance p does not change either.)

And in general, we can put a vertical parabola in an equation the way as follows.

$(x-u)^2 = 4p(y-v)$ where p, u, and v are constant.

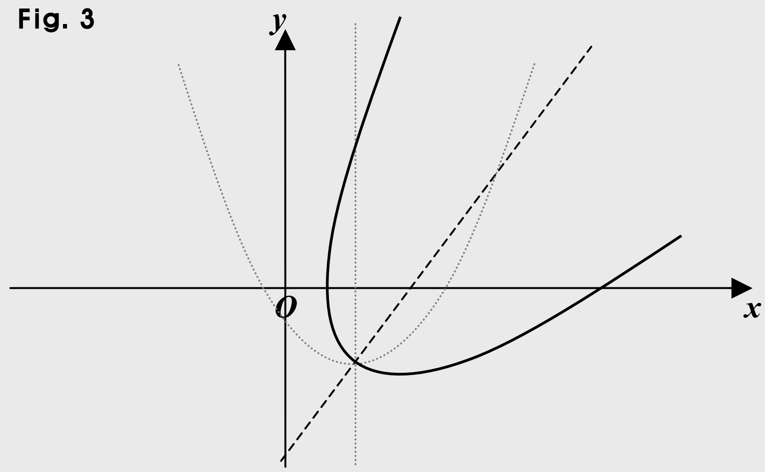
Then, the vertex of the parabola above is (u, v), the axis of symmetry is $x = u$, the focus is $(u, p + v)$, and the directrix is $y = -p + v$.
If $p > 0$, the parabola is concave-up. And if $p < 0$, the parabola is concave-down.

And the equation above is said to be in the vertex form.

And note that the equation above can indicate a vertical parabola only. If a parabola is vertical, the axis of symmetry is perpendicular to the x-axis.

What if a parabola is not perpendicular, and thus, is slanted as the parabola below?

Fig. 3



Given the focus and the directrix, we can derive the equation using the definition for parabolas, or can get the equation applying a transformation to a perpendicular parabola. And the transformation is covered in the section Transformation by Turning in the book **GRAPH OPERATIONS**.

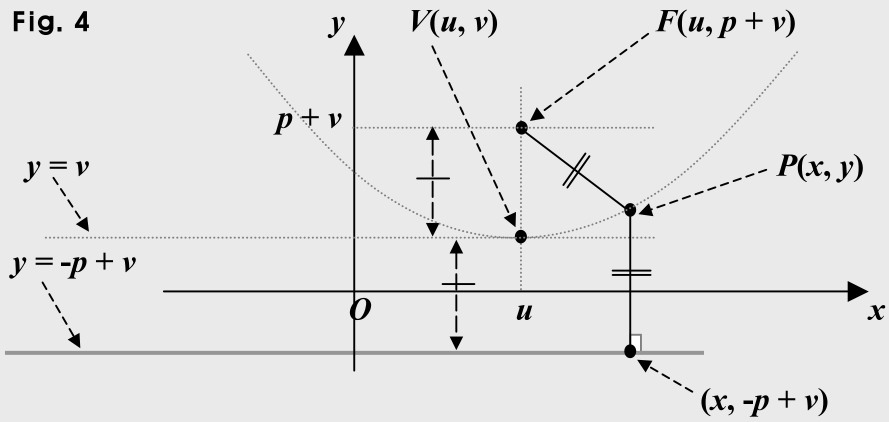
Let's now, for practice purposes, get this equation $(x - u)^2 = 4p(y - v)$. We are going to get it doing the derivation as done in the previous sections.

To begin with, a parabola is a set of points, from each of which, the distance to the focus is the same as the distance to the directrix.
So no matter what point we may choose in a parabola, the distance from the point to the focus is the same as the distance from the point to the directrix.

Using thus, the fact above, we can come up with an equation of a parabola.
So suppose now that the focus is $(u, p + v)$, and the directrix is $y = -p + v$.

Then again, if G is the parabola, and $P(x, y)$ is an arbitrary point in G, assuming P is now moving along the parabola G, we can put the snapshot of P the way as follows.

Fig. 4



Next, assuming again, $\boldsymbol{U}$ is the distance from $\boldsymbol{P(x, y)}$ to the focus $\boldsymbol{(u, p + v)}$, and $\boldsymbol{V}$ is the distance between $\boldsymbol{P(x, y)}$ and the directrix $\boldsymbol{y = -p + v}$, we get

$\boldsymbol{U^2 = (x - u)^2 + \{y - (p + v)\}^2}$, and $\boldsymbol{V^2 = (x - x)^2 + \{y - (-p + v)\}^2}$.

So we get $\boldsymbol{U^2 = (x - u)^2 + (y - p - v)^2}$, and $\boldsymbol{V^2 = (y + p - v)^2}$.

And next, since the two distances are the same, we get

$\boldsymbol{U^2 = V^2 \Rightarrow (x - u)^2 + (y - p - v)^2 = (y + p - v)^2 \Rightarrow (x - u)^2 = (y + p - v)^2 - (y - p - v)^2}$.

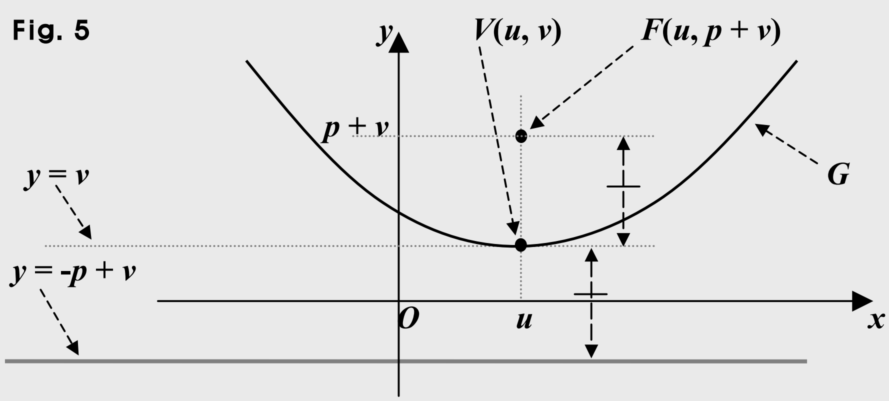
Meanwhile, we have an identity $\boldsymbol{A^2 - B^2 = (A - B)(A + B)}$.

So we get $\boldsymbol{(x - u)^2 = \{(y + p - v) - (y - p - v)\}\{(y + p - v) + (y - p - v)\}}$

$\boldsymbol{= (y + p - v - y + p + v)(y + p - v + y - p - v) = 2p(2y - 2v) = 4p(y - v)}$.

Thus, we get $\boldsymbol{(x - u)^2 = 4p(y - v)}$, which is the equation of the parabola $\boldsymbol{G}$.
And putting the parabola $\boldsymbol{G}$ in a graph, we can put it the way as follows.

Fig. 5



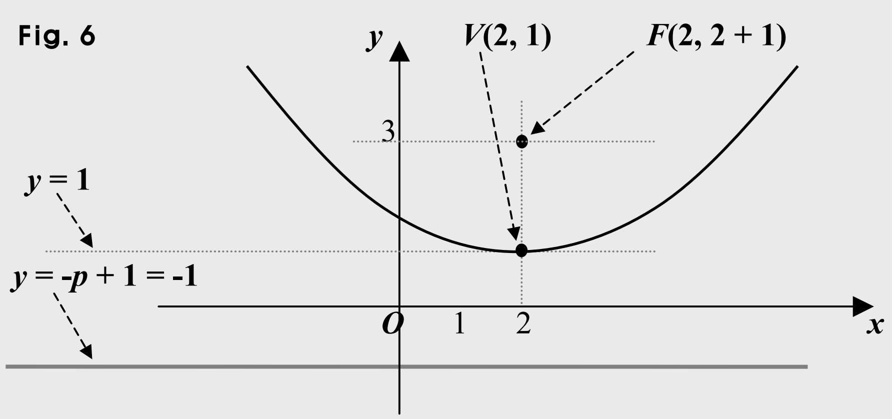
Note that the distance from the focus to the vertex is p, which is of course, the distance form the vertex to the directrix, too.

And of course, we can put $(x - u)^2 = 4p(y - v)$ this way: $y - v = \dfrac{1}{4p}(x - u)^2$.

Then, setting $m = \dfrac{1}{4p}$, we get $y - v = m(x - u)^2 \Rightarrow y = m(x - u)^2 + v$.

The equation above is often used in high school math, and is called the vertex form. Assuming for instance, the vertex is $(1, 2)$, and $p = 2$, we can put the parabola the way as follows.

Fig. 6



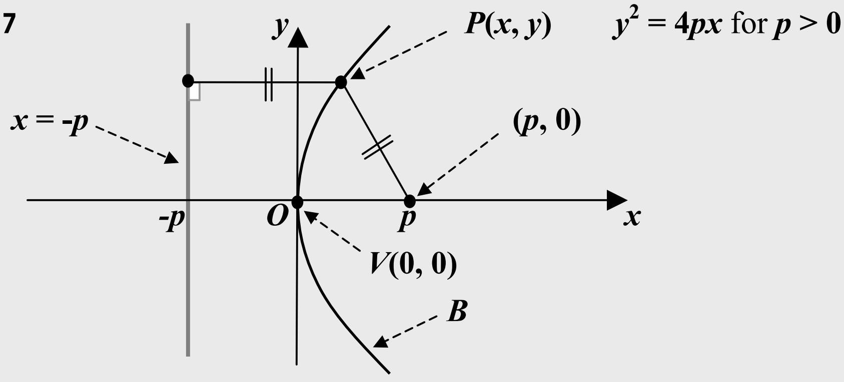
And of course, the equation of the parabola above is $(x - 2)^2 = 8(y - 1)$.
Next, putting another parabola in an equation, we can get

$y^2 = 4px$, where p is constant. What parabola then is it?

It is a parabola horizontal, the vertex is the origin $(0, 0)$, the axis of symmetry is $y = 0$, that is, the x-axis, the focus is at $(p, 0)$, and the directrix is a line $x = -p$.

And assuming $p > 0$, V is the vertex, F is the focus, and B is the parabola, we can put them all in the x-y plane the way as follows.

Fig. 7



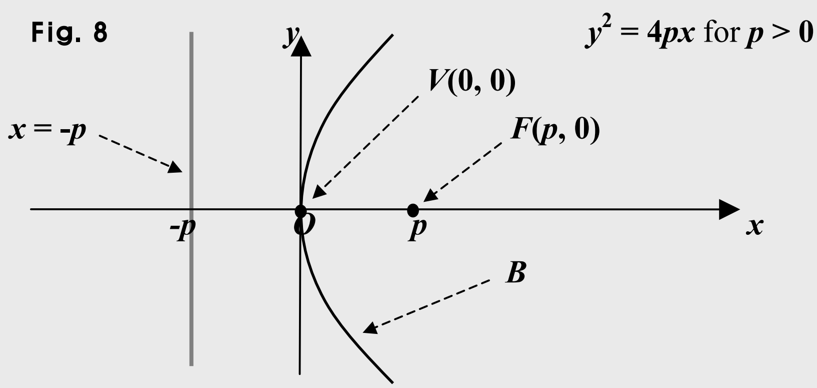
What if again, its vertex is at a point other than the origin?

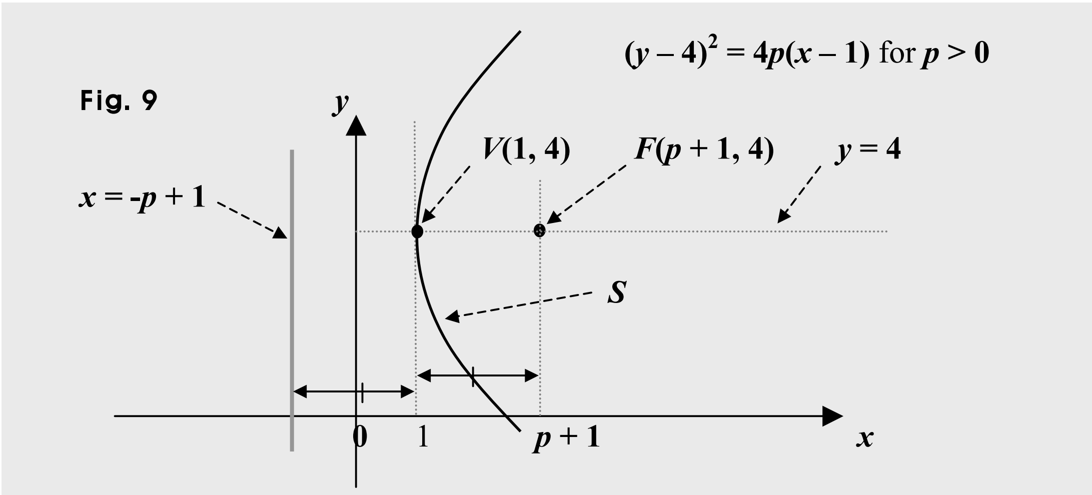
For instance, moving (translating) the parabola ***B*** so that the vertex is at (1, 4), we get a new parabola, so we get a new equation. What then will the new equation be?

Assuming the new parabola is ***S***, we can get ***S*** translating the parabola ***B*** in the amount of 1 in the direction of the ***x***-axis, and in the amount of 4 in the direction of the ***y***-axis.

Then, to get straight to the point, the equation of ***S*** is $(y - 4)^2 = 4p(x - 1)$.
And we can put ***B*** and ***S*** in their graphs the way as follows.

Fig. 8





The equation of the parabola S is $(y-4)^2 = 4p(x-1)$ where $p > 0$.

(Note that $4p$ remains the same, because the focal distance p did not change.
Translations do not change shapes, so the focal distance p does not change either.)
And in general, we can put a horizontal parabola in an equation the way as follows.

$(y-v)^2 = 4p(x-u)$ where p, u, and v are constant.

Then, the vertex of the parabola above is (u, v), the focus is $(p+u, v)$, and the directrix is a line $x = -p + u$.

If $p > 0$, the parabola is concave-right, that is, open to the right. If however, $p < 0$, the parabola is concave-left, that is, open to the left.
And the equation above is said to be in the vertex form, too.
Note however, the equation above can indicate a horizontal parabola only. If a parabola is horizontal, the axis of symmetry is parallel to the x-axis.

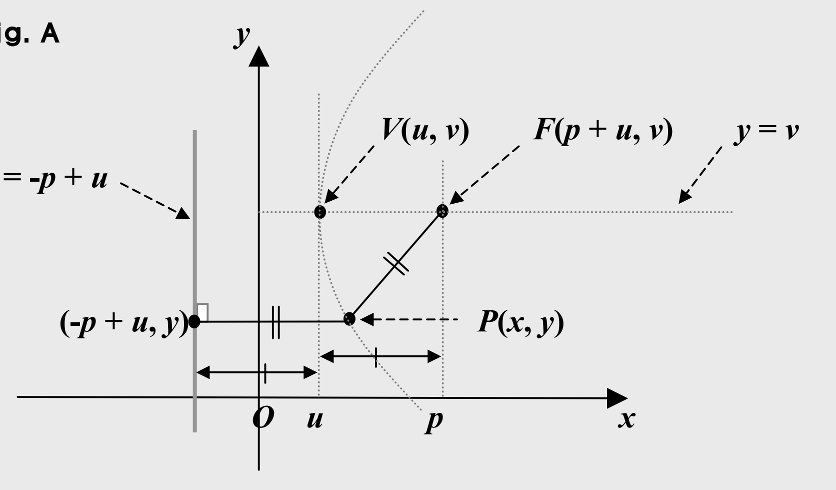
Let's now, for practice purposes, get this equation: $(y-v)^2 = 4p(x-u)$.
We are going to get it doing the derivation as done in the previous sections.

To begin with, a parabola is a set of points, from each of which, the distance to the focus is the same as the distance to the directrix.
So no matter what point we may choose in a parabola, the distance from the point to the focus is the same as the distance from the point to the directrix.

Using thus, the fact above, we can come up with an equation of a parabola.
So suppose now that the focus is $(p + u, v)$, and the directrix is $x = -p + u$.

Then again, if H is the parabola, and $P(x, y)$ is an arbitrary point in H, assuming P is now moving along the parabola H, we can put the snapshot of P the way as follows.

Fig. A



Next, assuming again, U is the distance from $P(x, y)$ to the focus $(p + u, v)$, and V is the distance between $P(x, y)$ and the directrix $x = -p + u$, we get

$U^2 = \{x - (p + u)\}^2 + (y - v)^2$, and $V^2 = \{x - (-p + u)\}^2 + (y - y)^2$.

So we get $U^2 = (x - p - u)^2 + (y - v)^2$, and $V^2 = (x + p - u)^2$.

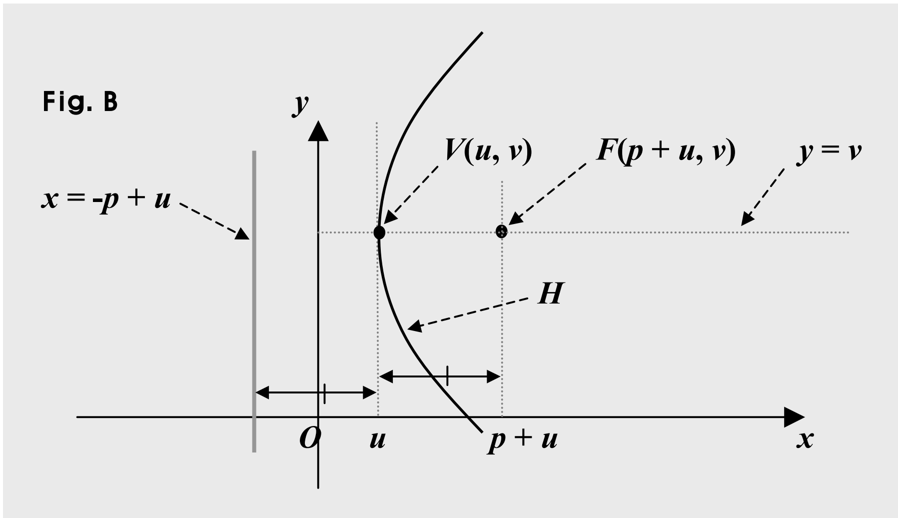
And next, since the two distances are the same, we get
$U^2 = V^2 \Rightarrow (x - p - u)^2 + (y - v)^2 = (x + p - u)^2 \Rightarrow (y - v)^2 = (x + p - u)^2 - (x - p - u)^2$.

Meanwhile, we have an identity $A^2 - B^2 = (A - B)(A + B)$.

So we get $(y - v)^2 = \{(x + p - u) - (x - p - u)\}\{(x + p - u) + (x - p - u)\}$

$= (x + p - u - x + p + u)(x + p - u + x - p - u) = 2p(2x - 2u) = 4p(x - u)$.

Thus, we get $(y - v)^2 = 4p(x - u)$, which is the equation of the parabola H.
And putting the parabola H in a graph, we can put it the way as follows.

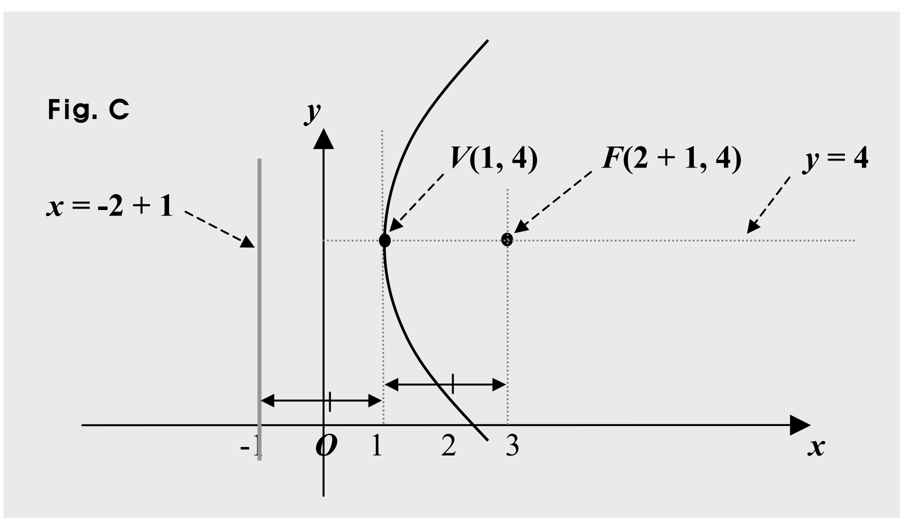


Note that the distance from the focus to the vertex is $\boldsymbol{p}$, which is of course, the distance from the vertex to the directrix, too.

And of course, we can put $(y - v)^2 = 4p(x - u)$ this way: $x - u = \dfrac{1}{4p}(y - v)^2$.

Then, setting $m = \dfrac{1}{4p}$, we get $x - u = m(y - v)^2 \Rightarrow x = m(y - v)^2 + u$.

The equation above is often used in high school math, and is called the vertex form. Assuming for instance, the vertex is **(1, 4)**, and $p = 2$, we can put the parabola the way as follows.



And in $(y - v)^2 = 4p(x - u)$, replacing p with 2, u with 1, and v with 4, we get

$(y - 4)^2 = 8(x - 1)$, which is the equation of the parabola above, and can be put the way, too, of course: $x - 1 = \frac{1}{8}(y - 4)^2$, which is often put this way, too: $x = \frac{1}{8}(y - 4)^2 + 1$.

1.7. General Equations

As mentioned earlier in the previous sections, the equations in the vertex form can indicate parabolas that are perpendicular only. If a parabola is perpendicular, the axis of symmetry is perpendicular to the x-axis or parallel to the x-axis. If the axis of symmetry is vertical (perpendicular to the x-axis), the parabola is vertical, and if the axis of symmetry is horizontal (parallel to the x-axis), the parabola is horizontal.

What then about equations that can indicate parabolas of all kinds?

Expressing a parabola perpendicular or not, we can use an equation as follows.

$$ax^2 + by^2 + cxy + dx + ey + f = 0$$

where a, b, c, d, e, and f are constant, of course.

The equation above can however, indicate other conic, too.
The conic that the equation indicates depends on the values of the constants.
If it indicates a parabola, it has to meet some conditions on the constants as follows.

$$ax^2 + by^2 + cxy + dx + ey + f = 0$$

where $ab \geq 0$ and $a^2 + b^2 \neq 0$ if $c \neq 0$.

What do we mean by $ab \geq 0$ and $a^2 + b^2 \neq 0$ if $c \neq 0$?

First, "$ab \geq 0$ and $a^2 + b^2 \neq 0$" means that a and b have the same sign, and cannot be 0 at the same time.
So "$ab \geq 0$ and $a^2 + b^2 \neq 0$ if $c \neq 0$" means that if c is not 0, one of a and b is nonzero, and if both are nonzero, both have the same sign.

Thus, if $c \neq 0$, then if $ab \geq 0$ and $a^2 + b^2 \neq 0$, the equation indicates a parabola, and the parabola is skewed or slanted.
That is, if $c \neq 0$, and the equation indicates a parabola, the parabola is not perpendicular.
What if $c = 0$?

Then, either a or b is 0.
So if $c = 0$, we get $ax^2 + dx + ey + f = 0$ or $by^2 + dx + ey + f = 0$.
Note: $ae \neq 0$ in the first equation, so none of a and e is 0, and $bd \neq 0$ in the second.

And we can call all those equations above general equations for parabolas, and the equations are in the general form. How then can we get those general equations?

If we find the equation of a parabola slanted, the equation will be in the form as follows.

$ax^2 + by^2 + cxy + dx + ey + f = 0$. How then can we find it?

Given the focus and the directrix, we can derive it using the definition for parabolas, or can find it applying a transformation to an equation of a parabola perpendicular. And the transformation can be called *transformation by turning*, which is covered in the book **Function Transformations**. What then about parabolas perpendicular?

We know if $c = 0$, we get $ax^2 + dx + ey + f = 0$ or $by^2 + dx + ey + f = 0$.

Each equation above indicates in fact, a perpendicular parabola. The first is vertical, and the second is horizontal. And each is in the non-vertex form. How then can we get it?

Simplifying the equations in the vertex form, we get the equations in the non-vertex form.
That is to say that simplifying $(x - u)^2 = m(y - v)$, we get an equation in the non-vertex form as follows. $ax^2 + dx + ey + f = 0$, which indicates a parabola vertical.

$(x - u)^2 = m(y - v) \Rightarrow x^2 - 2ux + u^2 = my - mv \Rightarrow x^2 - 2ux - my + u^2 + mv = 0$.

So setting $a = 1$, $d = -2u$, $e = -m$, and $f = u^2 + mv$, we get $ax^2 + dx + ey + f = 0$, because all the letters other than x and y are all constants.

And next, simplifying $(y-v)^2 = m(x-u)$, we get an equation in the non-vertex form as follows. $by^2 + dx + ey + f = 0$, which indicates a parabola horizontal.
So simplifying it, we get this:
$(y-v)^2 = m(x-u) \Rightarrow y^2 - 2vy + v^2 = mx - mu \Rightarrow y^2 - 2vy - mx + v^2 + mu = 0$.

So setting $b = 1$, $d = -2v$, $e = -m$, and $f = v^2 + mu$, we get $by^2 + dx + ey + f = 0$.

And usually, we use several different vertex forms when producing equations. So for instance, all the equations below are in the vertex form, and the parabolas are vertical.

$(x-a)^2 = m(y-b)$, $a(y-b) = m(x-c)^2$, $y-b = m(x-c)^2$, $y = m(x-a)^2 + b$,

$(x-1)^2 = 3(y-2)$, $2(y-1) = (x-3)^2$, $y-5 = 2(x+3)^2$, $y = 2(x+1)^2 + 3$, etc.

What then about the equations as follows?

$x^2 = ay$, $y = cx^2$, $bx^2 = ay$, $by = cx^2$, $y = 2x^2$, $3y = 2x^2$, etc.

They are in the vertex form, and all indicate vertical parabolas.
And all their vertices are at the origin.
Next, if the parabolas are horizontal, we can use the vertex forms the way as follows.

$(y-a)^2 = m(x-b)$, $a(x-b) = m(y-c)^2$, $x-b = m(y-c)^2$, $x = m(y-a)^2 + b$,

$(y-1)^2 = 3(x-2)$, $2(x-1) = (y-3)^2$, $x-5 = 2(y+3)^2$, $x = 2(y+1)^2 + 3$, etc.

And the equations below are in the vertex form, and indicate horizontal parabolas.

$y^2 = ax$, $x = cy^2$, $by^2 = ax$, $bx = cy^2$, $x = 2y^2$, $3x = 2y^2$, etc.

What then about the non-vertex form?

If the parabola is vertical, we often use this form: $ax^2 + bx + c + dy = 0$ where $ad \neq 0$.
Note: $ad \neq 0$ means that neither of a and d is 0.

Usually though, we put it this way: $y = ax^2 + bx + c$, where $a \neq 0$, of course.
So for instance, we can have $y = x^2 + 2x + 3$, $y = x^2 - 5x + 2$, $y = -x^2 + 3x - 1$, etc.

Or sometimes, we can have one this way, too: $\boldsymbol{y = 1 - 5x - 3x^2}$.
And the same is true for the non-vertex forms for horizontal parabolas, too.

So if it is horizontal, we often use this form: $\boldsymbol{ay^2 + by + c + dx = 0}$ where $\boldsymbol{ad \neq 0}$.

Usually though, we put it this way: $\boldsymbol{x = ay^2 + by + c}$, where $\boldsymbol{a \neq 0}$, of course.

So for instance, we can have $\boldsymbol{x = y^2 + 2y + 3}$, $\boldsymbol{x = y^2 - 5y + 2}$, $\boldsymbol{x = -y^2 + 3y - 1}$, etc.
Or sometimes, we can have one this way, too: $\boldsymbol{x = 1 - 5y - 3y^2}$.
How then can we put into the standard form an equation in the non-vertex form?

You can put it in this form: $\boldsymbol{y = m(x - a)^2 + b}$, $\boldsymbol{y = k(x - u)^2 + v}$, $\boldsymbol{y = a(x - p)^2 + q}$, etc.

They are all the same though. It all depends on your teacher's preference of letters. And the form is called the vertex form, because $\boldsymbol{(p, q)}$ is the vertex if we use $\boldsymbol{y = a(x - p)^2 + q}$.

So converting for instance, $\boldsymbol{y = ax^2 + bx + c}$ to the vertex form equivalent, we can do it the way as follows. (Many people call this method the complete square method.)

To begin with, we get $\boldsymbol{ax^2 + bx + c = a(x^2 + \frac{b}{a}x) + c}$

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

Meanwhile, we have $\boldsymbol{-a\left(\frac{b}{2a}\right)^2 + c = -\frac{b^2}{4a} + \frac{4ac}{4a} = -\frac{b^2 - 4ac}{4a}}$.

So we get $\boldsymbol{y = ax^2 + bx + c = a\left(x + \frac{b}{2a}\right)^2 - \frac{b^2 - 4ac}{4a}}$.

Thus, we get $\boldsymbol{y = a\left(x + \frac{b}{2a}\right)^2 - \frac{b^2 - 4ac}{4a}}$.

Now, setting $\boldsymbol{p = -\frac{b}{2a}}$, and $\boldsymbol{q = -\frac{b^2 - 4ac}{4a}}$, we get $\boldsymbol{y = a(x - p)^2 + q}$.
And the vertex is $\boldsymbol{(p, q)}$.

And notice that if a parabola is $y = ax^2 + bx + c$, the axis of symmetry is $x = -\dfrac{b}{2a}$.

For instance, assuming $y = 2x^2 - 4x + 5$, we can convert it to the vertex form equivalent, that is, we can put it in the vertex form the way as follows.

$y = 2x^2 - 4x + 5 = 2(x^2 - 2x) + 5 = 2(x^2 - 2x + 1 - 1) + 5 = 2(x^2 - 2x + 1) - 2 + 5$

$= 2(x - 1)^2 + 3 \Rightarrow y = 2(x - 1)^2 + 3$, which is now, in the vertex form. The vertex is (1, 3).

And notice that the axis of symmetry is $x = 1$.

For another instance, assuming $y = -2x^2 - 3x + 5$, we can put it in the vertex form the way as follows.

$$y = -2x^2 - 3x + 5 = -2\{x^2 + \frac{3}{2}x + \left(\frac{3}{4}\right)^2 - \left(\frac{3}{4}\right)^2\} + 5 = -2\left(x + \frac{3}{4}\right)^2 + 2\left(\frac{3}{4}\right)^2 + 5$$

$$= -2\left(x + \frac{3}{4}\right)^2 + \frac{9}{8} + 5 \Rightarrow y = -2\left(x + \frac{3}{4}\right)^2 + \frac{49}{8}$$, which is now, in the vertex form.

And the vertex is $(-\dfrac{3}{4}, \dfrac{49}{8})$. And notice that the axis of symmetry is $x = -\dfrac{3}{4}$.

And the same is true for converting $x = ay^2 + by + c$ to the standard equivalent, too.

So to begin with, we get $ay^2 + by + c = a(y^2 + \dfrac{b}{a}y) + c$

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

Meanwhile, we have $-a\left(\dfrac{b}{2a}\right)^2 + c = -\dfrac{b^2}{4a} + \dfrac{4ac}{4a} = -\dfrac{b^2 - 4ac}{4a}$.

So we get $x = ay^2 + by + c = a\left(y + \dfrac{b}{2a}\right)^2 - \dfrac{b^2 - 4ac}{4a}$.

Thus, we get $x = a\left(y + \dfrac{b}{2a}\right)^2 - \dfrac{b^2 - 4ac}{4a}$.

Now, setting $\boldsymbol{q} = -\dfrac{\boldsymbol{b}}{\boldsymbol{2a}}$, and $\boldsymbol{p} = -\dfrac{\boldsymbol{b^2 - 4ac}}{\boldsymbol{4a}}$ we get $\boldsymbol{x = a(y - q)^2 + p}$.

Note that the settings for $\boldsymbol{p}$ and $\boldsymbol{q}$ are reversed in this case.
So in this case, too, the vertex is $(\boldsymbol{p}, \boldsymbol{q})$.

And notice that if a parabola is $\boldsymbol{x = ay^2 + by + c}$, the axis of symmetry is $\boldsymbol{y = -\dfrac{b}{2a}}$.

For instance, assuming $\boldsymbol{x = 2y^2 - 4y + 5}$, we can convert it to the vertex form equivalent, that is, we can put it in the vertex form the way as follows.

$\boldsymbol{x = 2y^2 - 4y + 5 = 2(y^2 - 2y) + 5 = 2(y^2 - 2y + 1 - 1) + 5 = 2(y^2 - 2y + 1) - 2 + 5}$

$\boldsymbol{= 2(y - 1)^2 + 3 \Rightarrow x = 2(y - 1)^2 + 3}$, which is now, in the vertex form. The vertex is (3, 1).

And notice that the axis of symmetry is $\boldsymbol{y = 1}$.

For another instance, assuming $\boldsymbol{x = -2y^2 - 3y + 5}$, we can put it in the vertex form the way as follows.

$$\boldsymbol{x = -2y^2 - 3y + 5 = -2\{y^2 + \frac{3}{2}y + \left(\frac{3}{4}\right)^2 - \left(\frac{3}{4}\right)^2\} + 5 = -2\left(y + \frac{3}{4}\right)^2 + 2\left(\frac{3}{4}\right)^2 + 5}$$

$$\boldsymbol{= -2\left(y + \frac{3}{4}\right)^2 + \frac{9}{8} + 5 \Rightarrow x = -2\left(y + \frac{3}{4}\right)^2 + \frac{49}{8}}$$, which is now in the vertex form.

And the vertex is $\boldsymbol{(\frac{49}{8}, -\frac{3}{4})}$.

And notice that the axis of symmetry is $\boldsymbol{y = -\dfrac{3}{4}}$.

1.8. Half Parabolas

Sometimes, we get to use an equation as follows.

$$y = m\sqrt{x-s} + t$$

In the equation above, $\boldsymbol{m}$, $\boldsymbol{s}$, and $\boldsymbol{t}$ are constant, of course, and $\boldsymbol{m \neq 0}$.
What equation then is it?

It indicate a half parabola, so it is an equation of a half parabola.
There can be however, two different halves for a parabola.

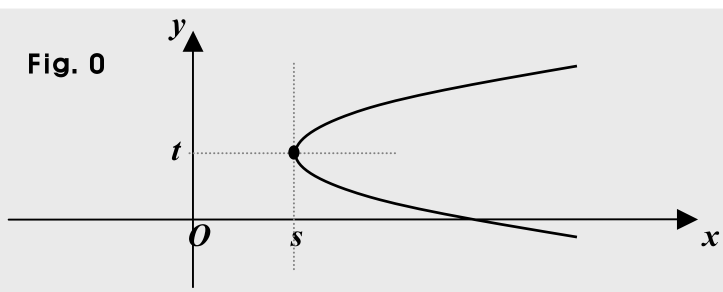
If the parabola is vertical, it can be a left half or a right half.
If the parabola is horizontal, it can an upper half or a lower half.
So what half does the equation indicate?

Finding the equation of the parabola first, we get

$$y = m\sqrt{x-s} + t \Rightarrow y - t = m\sqrt{x-s} \Rightarrow (y-t)^2 = m^2(x-s).$$

So we can now see that the parabola is horizontal, and is concave-right, because $\boldsymbol{m^2 > 0}$.
And the vertex is at $\boldsymbol{(s, t)}$, and the axis of symmetry is $\boldsymbol{y = t}$.

Fig. 0



What half then does the equation $y = m\sqrt{x-s}+t$ indicate?

First, s and t are constant, and $\sqrt{x-s}$ can be defined for $x \geq s$, because what's inside the square root has to be greater than or equal to 0, that is, $x - s \geq 0$, so $x \geq s$.

And next, we know $\sqrt{x-s} \geq 0$ if $x \geq s$.
So if $m > 0$, we get $m\sqrt{x-s} \geq 0$, which means that the value of $(m\sqrt{x-s}+t)$ increases as x increases from s. If therefore, $m > 0$, y increases as x increases from s.

So if $m > 0$, the equation $y = m\sqrt{x-s}+t$ indicates the upper half.

Next, if $m < 0$, we get $m\sqrt{x-s} \leq 0$, which means that as x increases from s, the value of $(m\sqrt{x-s}+t)$ decreases. If therefore, $m < 0$, y decreases as x increases from s.

So if $m < 0$, the equation $y = m\sqrt{x-s}+t$ indicates the lower half.

By the way, translating a parabola $y^2 = m^2x$ by s in the direction of the x-axis, and also, by t in the direction of the y-axis, we get the parabola $(y - t)^2 = m^2(x - s)$.
Let's for instance, put in a graph the curve of $y = \frac{1}{2}\sqrt{x-2}+1$.
To begin with, we get

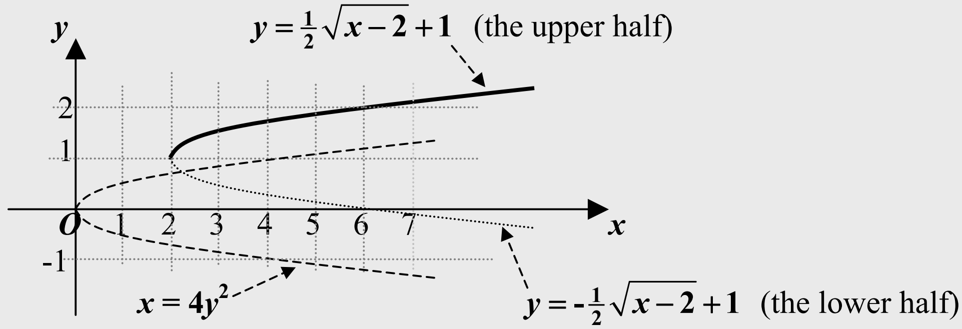
$$y = \frac{\sqrt{x-2}}{2}+1 \Rightarrow y-1 = \frac{\sqrt{x-2}}{2} \Rightarrow (y-1)^2 = \frac{\sqrt{x-2}}{4} \Rightarrow x-2 = 4(y-1)^2.$$

What parabola then is it?

Translating a parabola $x = 4y^2$ by 2 along the x-axis, and by 1 along the y-axis, we get the parabola $x - 2 = 4(y - 1)^2$.
So it is a horizontal parabola, the vertex is $(2, 1)$, and the axis of symmetry is $y = 1$.

Thus, $y = \frac{1}{2}\sqrt{x-2}+1$ indicates the upper half of the horizontal parabola above, which is $x - 2 = 4(y - 1)^2$. And putting them now in a graph, we get this:

Fig. 1



What then about this equation: $x = m\sqrt{y-t} + s$?

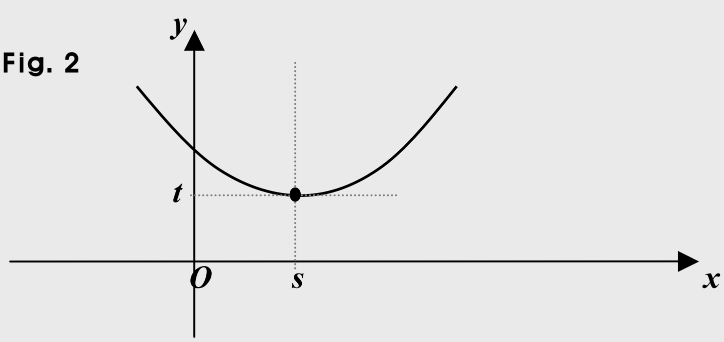
It indicates the right half or the left half of a parabola vertical.
What half then is it?

Finding the equation of the parabola first, we get

$$x = m\sqrt{y-t} + s \Rightarrow x - s = m\sqrt{y-t} \Rightarrow (x-s)^2 = m^2(y-t).$$

So we can now see that the parabola is vertical, and is concave-up, because $m^2 > 0$. And the vertex is at (s, t), and the axis of symmetry is $x = s$.

Fig. 2



What half then, does the equation $x = m\sqrt{y-t} + s$ indicate?

First, s and t are constant, and $\sqrt{y-t}$ can be defined for $y \geq t$, because what's inside the square root has to be greater than or equal to 0, that is, $y - t \geq 0$, so $y \geq t$.

And next, we know $\sqrt{y-t} \geq 0$ if $y \geq t$.

So if $m > 0$, we get $m\sqrt{y-t} \geq 0$, which means that the value of $m\sqrt{y-t}+s$ increases as y increases from t. If therefore, $m > 0$, x increases as y increases from y.

So if $m > 0$, the equation $x = m\sqrt{y-t}+s$ indicates the right half.

Next, if $m < 0$, we get $m\sqrt{y-t} \leq 0$, which means that as y increases from t, the value of $m\sqrt{y-t}+s$ decreases. If therefore, $m < 0$, x decreases as y increases from t.

So if $m < 0$, the equation $x = m\sqrt{y-t}+s$ indicates the left half.

And translating a parabola $y^2 = m^2x$ by s in the direction of the x-axis, and also, by t in the direction of the y-axis, we get the parabola $(y-t)^2 = m^2(x-s)$.

Let's for instance, put in a graph the curve of $x = \frac{1}{2}\sqrt{y-1}+2$.

To begin with, we get

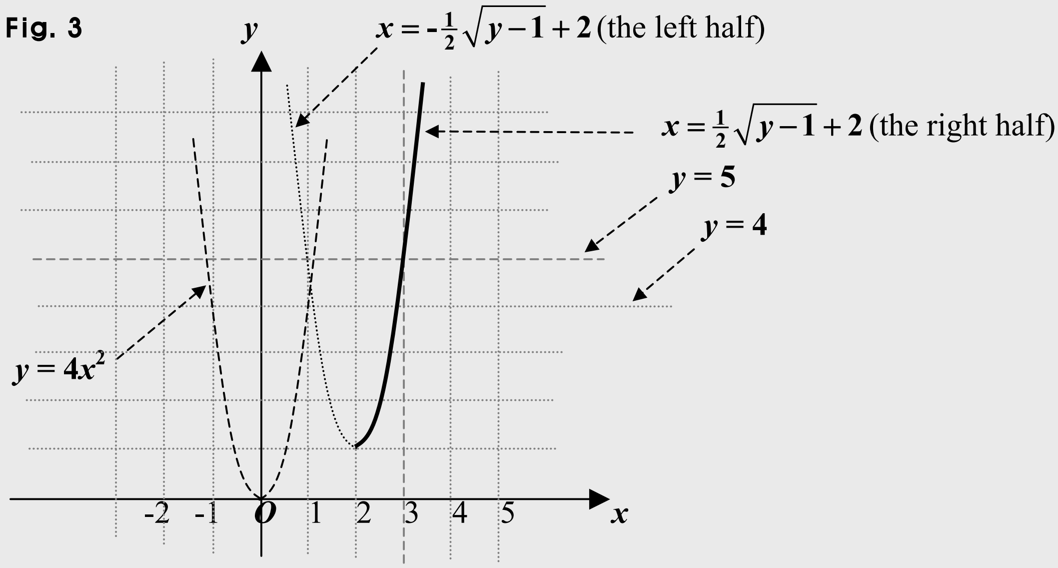
$$x = \frac{\sqrt{y-1}}{2}+2 \Rightarrow x-2 = \frac{\sqrt{y-1}}{2} \Rightarrow (x-2)^2 = \frac{y-1}{4} \Rightarrow y-1 = 4(x-2)^2.$$

What parabola then is it?

Translating a parabola $y = 4x^2$ by 2 along the x-axis, and by 1 along the y-axis, we get the parabola $y-1 = 4(x-2)^2$.
So it is a vertical parabola, the vertex is $(2, 1)$, and the axis of symmetry is $x = 2$.

And thus, $x = \frac{1}{2}\sqrt{y-1}+2$ indicates the right half of the vertical parabola above, which is $y-1 = 4(x-2)^2$. And putting them now in a graph, we get this:

Fig. 3



And the left half of the parabola is $x = -\frac{1}{2}\sqrt{y-1}+2$.

1.9. Equations for Parabolas 2

Doing problems, we often need to find a curve such as a line or parabola.
Finding a curve, we find the equation of it.
We can't just find it, of course. We want to know the information on the curve to be found. So we want to get the information first. What information then do we need?

Let's begin with necessary information to find a line.
What information then do we need?

Finding a line, we need the slope and a point in the line. What if no slope is given?

Then, we need two points in the line. So finding a line, we need either of the two as follows

> One point in the line and the slope of the line
> Two points in the line

Suppose for instance, the slope is a, and a point (s, t) is in the line.
Then, the line is $y-t=a(x-s)$.

Suppose next, two points (s, t) and (u, v) are in the line.
Then, the slope is $\dfrac{v-t}{u-s}$, and the line is $y-t=\dfrac{v-t}{u-s}(x-s)$.

What then about this: $y-v=\dfrac{v-t}{u-s}(x-u)$?
Both are the same equations.

$$y - v = \frac{v-t}{u-s}(x-u) \Rightarrow (y-v)(u-s) = (v-t)(x-u)$$
$$\Rightarrow yu - ys - vu + vs = vx - vu - tx + tu \Rightarrow yu - ys + vs = vx - tx + tu$$
$$\Rightarrow x(v-t) + y(s-u) = vs - tu$$

$$y - t = \frac{v-t}{u-s}(x-s) \Rightarrow (y-t)(u-s) = (v-t)(x-s)$$
$$\Rightarrow yu - ys - tu + ts = vx - vs - tx + ts \Rightarrow yu - ys - tu = vx - vs - tx$$
$$\Rightarrow x(v-t) + y(s-u) = vs - tu$$

Let's now, move on to parabolas.

If the parabola is vertical, we can use an equation in the non-vertex form as follows.

$y = ax^2 + bx + c$, where a, b, and c are constant.

And if the parabola is horizontal, an equation in the non-vertex form is $x = ay^2 + by + c$.

So if using each of the equation above finding a particular parabola, we need to find a, b, and c, which are three unknowns. So we need to solve a system of three equations.

Given a point in the parabola, we can use the coordinates of the point, and set up an equation for the three unknowns, which are, of course, a, b, and c.

So given three points in the parabola, we can get a system of three equations.
Then, solving the system, we can get the parabola.
That's not the only way though.

We can be given information many different ways.
If using the non-vertex form above to find the parabola, what we have to do is to find the values of a, b, and c.

Knowing more about the equation in the non-vertex form above, we can often find the parabola quickly and easily.
So let's now take a close look at the non-vertex form, and see what we can use.

If the parabola is vertical, we can use the non-vertex form: $y = ax^2 + bx + c$.

Then, first, assuming p is the focal distance, we get $|a| = \dfrac{1}{4p}$.

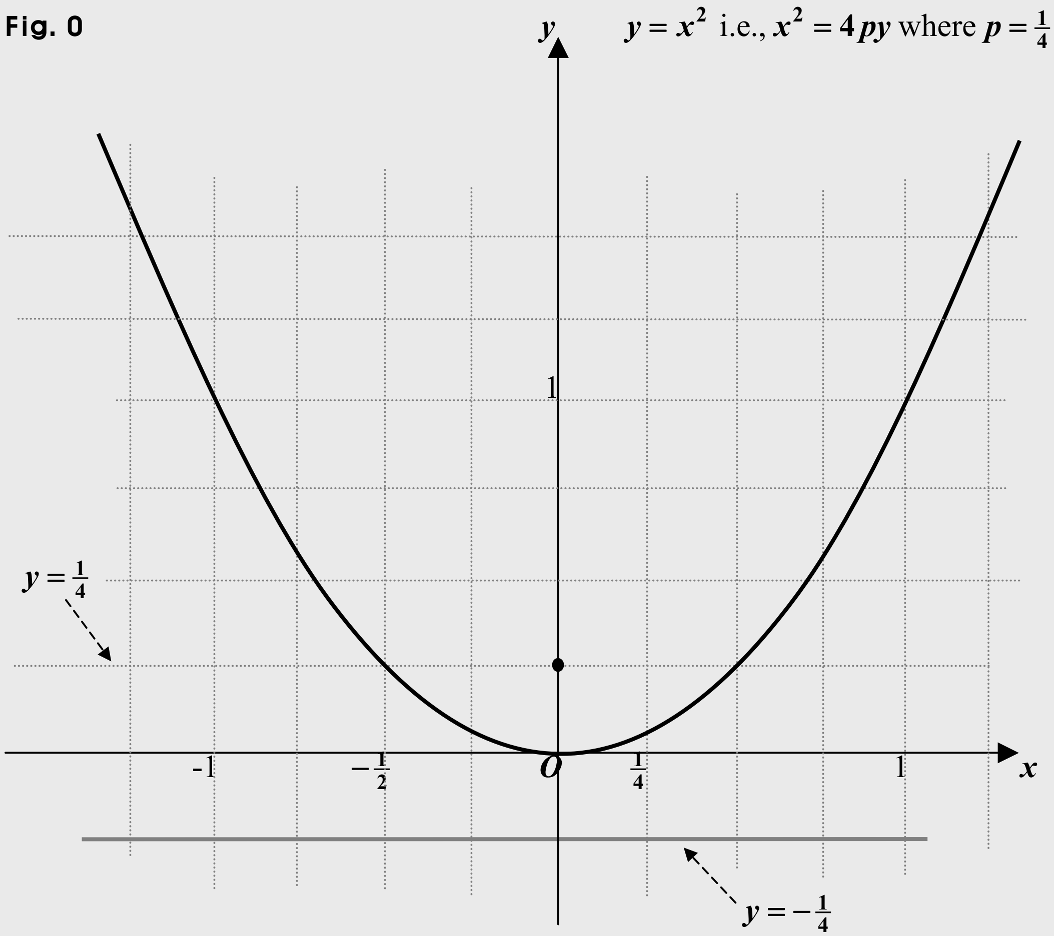
Next, if concave-up, $a > 0$, and if concave-down, $a < 0$.
So if we know the focal distance and the concavity, we can get a.
Isn't there anything else then about the value of a?

It can be said that $|a|$ indicates the slope of the parabola.

The bigger the value of $|a|$, the steeper the parabola.

For instance, in $y = ax^2 + bx + c$, we can see that y-value changes as x-value changes.
And the bigger the value of $|a|$, the faster the y-value changes as the x-value changes.

Fig. 0



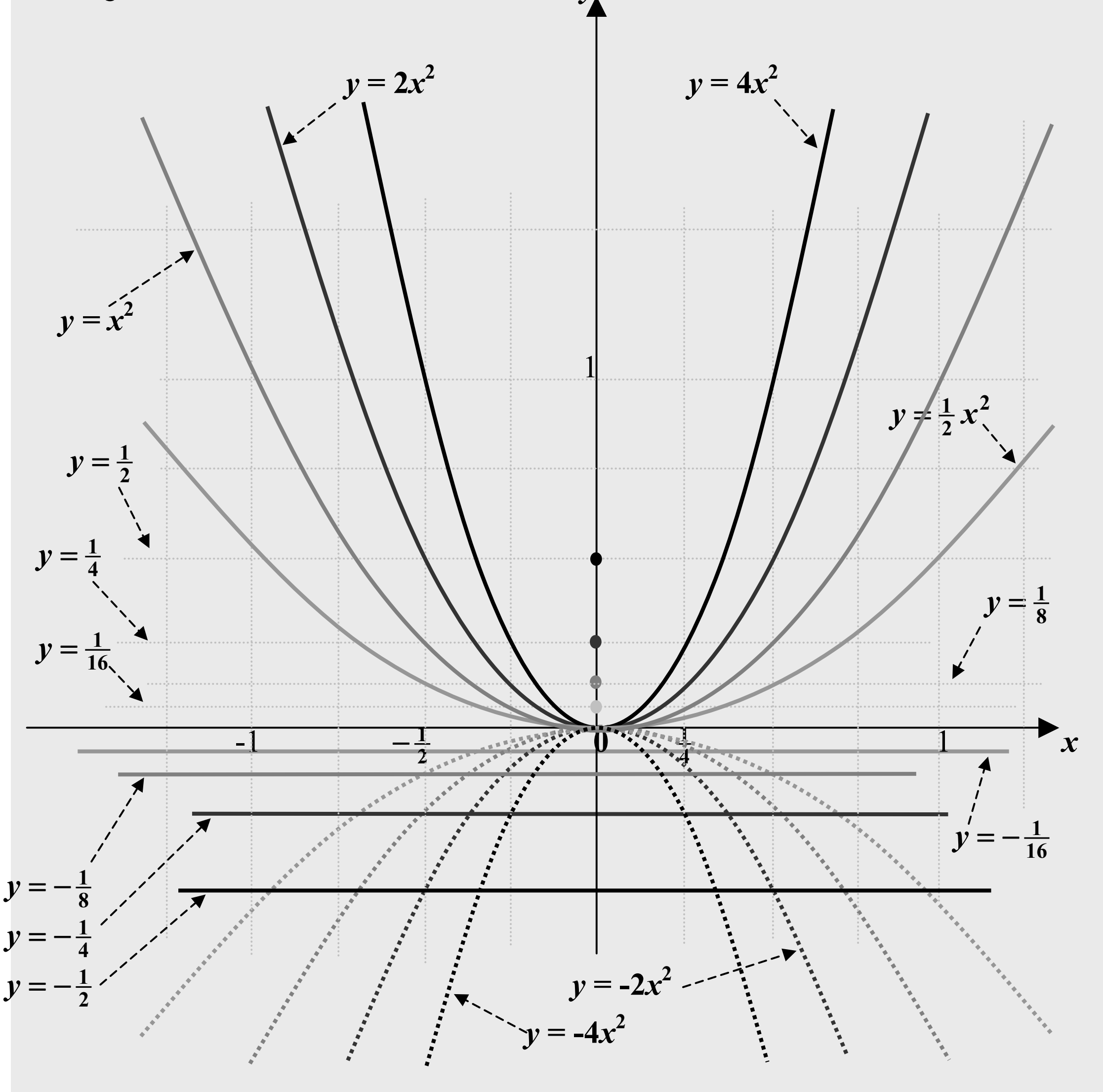
Next, let's have a look at some parabolas in the form of $y = ax^2$.
For instance, consider the parabolas as follows.

$$y = \frac{1}{2}x^2 \Rightarrow x^2 = 2y = 4 \cdot \frac{1}{2} y \Rightarrow \text{focus} = (0,\ \frac{1}{2}),\ \text{directrix}: y = -\frac{1}{2},\ \text{and vertex} = (0,0).$$

$$y = 2x^2 \Rightarrow x^2 = \frac{1}{2} y = 4 \cdot \frac{1}{8} y \Rightarrow \text{focus} = (0,\ \frac{1}{8}),\ \text{directrix}: y = -\frac{1}{8},\ \text{and vertex} = (0,0).$$

$$y = 4x^2 \Rightarrow x^2 = \frac{1}{4} y = 4 \cdot \frac{1}{16} y \Rightarrow \text{focus} = (0,\ \frac{1}{16}),\ \text{directrix}: y = -\frac{1}{16},\ \text{and vertex} = (0,0).$$

Fig. 1



So we can see that as the value of ***a*** gets bigger ***in magnitude***, the focus and directrix get closer to the vertex. Also, the parabola gets more bent, that is, steeper or narrower.

Suppose now that we move a parabola, $y = ax^2$ by ***m*** along the ***x***-axis and by ***n*** along the ***y***-axis. That is, we translate the parabola.

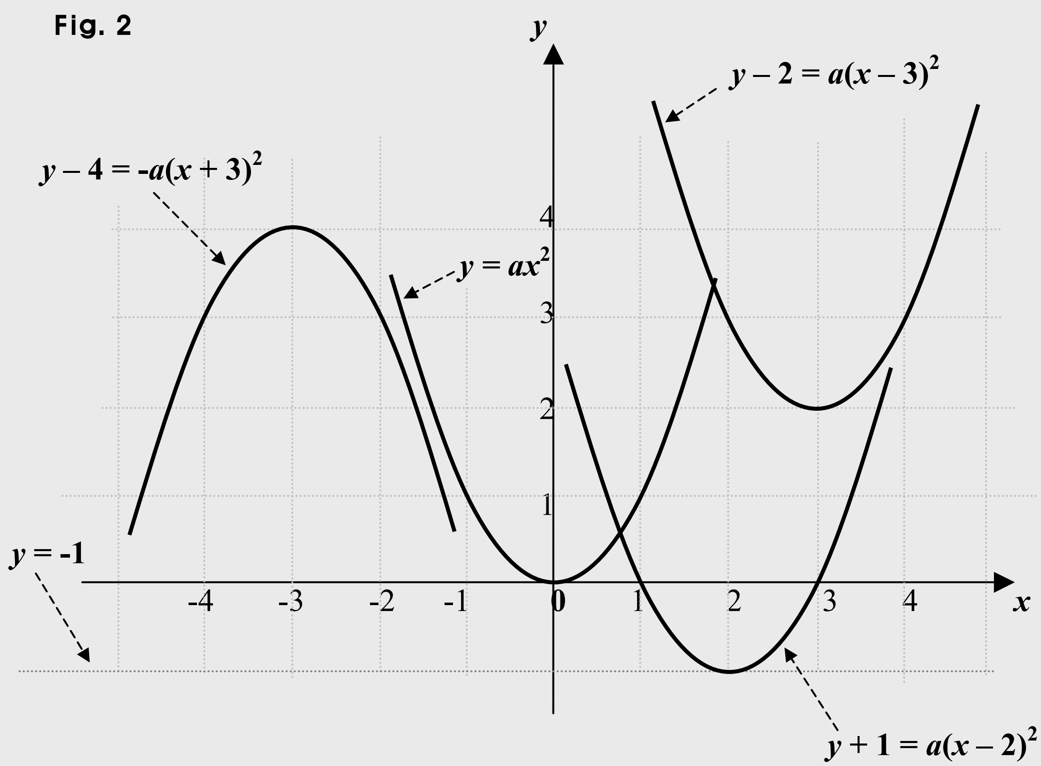
Then, the vertex, focus, etc. get translated in the same manner. So we get a new parabola. The new one keeps the same shape as the original, but takes different position.

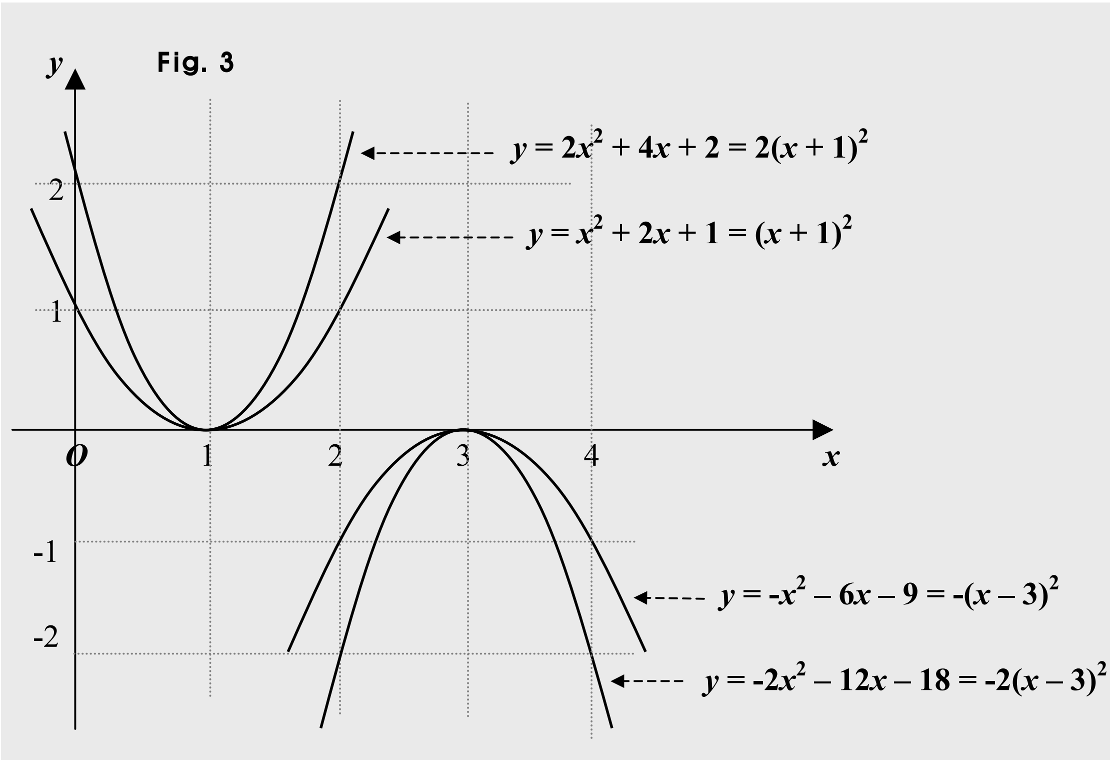
The equation of the new one is as follows. $y - n = a(x - m)^2$.

So all the parabolas themselves in the graph in Fig. 2 are the same, but placed at different locations respectively.

That's because they all have the same value of ***a in magnitude***. Mathematically however, they are all different parabolas not only because each has a different equation, but also because each has a different position in the coordinate plane.

Fig. 2





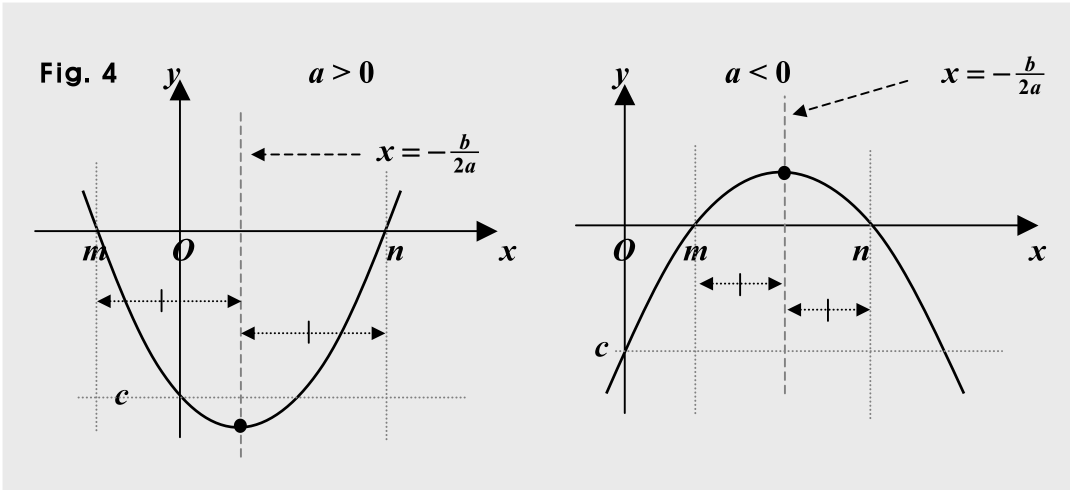
Next, if putting $y = ax^2 + bx + c$ in a graph, we can notice some useful facts about the constants.

The y-intercept is c, which is the y-value when $x = 0$.

The axis of symmetry is $x = \frac{-b}{2a}$, which is also the average of the two roots (x-intercepts).

The sum of the two roots (x-intercepts) is $-\frac{b}{a}$, and the product of the two roots is $\frac{c}{a}$.

That is, if m and n are the two roots, we get $m + n = -\frac{b}{a}$, $mn = \frac{c}{a}$, and $\frac{m+n}{2} = -\frac{b}{2a}$.



So assuming the parabola we want to find is $y = ax^2 + bx + c$, if using the facts above setting up a system of three equations for a, b, and c, and solving the system, we find the parabola. And setting up the system, we need one of the sets of information as follows.

1. Two points in the parabola, the focal distance, and the concavity

2. Two points and the y-intercept (The two points are in the parabola, of course.)

3. Two points and the axis of symmetry

4. Two points and the sum of the two roots

5. Two points and the product of the two roots

6. One point and the sum of and the product of the two roots.

Etc. So there can be many other cases we can consider.

Let's see now what we can do in each of the cases above.

1. Two points and the focal distance, along with the concavity

First, assuming p is the focal distance, we get $|a| = \dfrac{1}{4p}$.

Next, if concave-up, $a > 0$, and if concave-down, $a < 0$.
Next, using the two points, we can set up two equations for b and c.
So solving the system of the two equations, we can get b and c, and then the parabola.
For instance, the focal distance is $\frac{1}{4}$, and the parabola is concave-up, and has (1, 2) and (2, 1).

Then first, assuming the parabola is $y = ax^2 + bx + c$, we get $a = 1$.

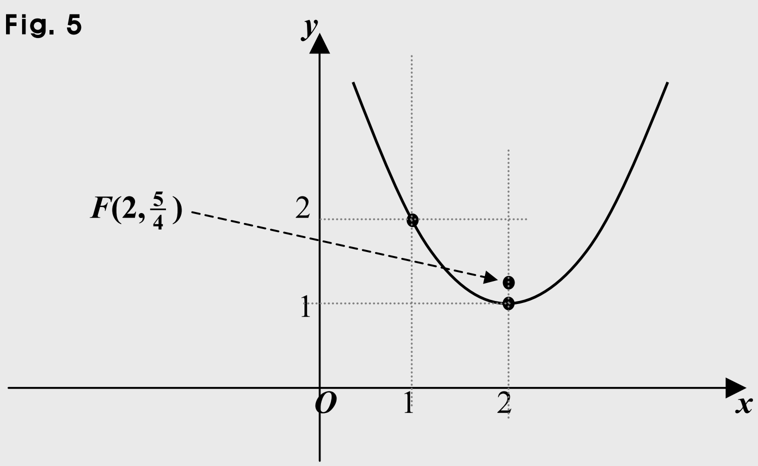
Next, since (1, 2) is in the parabola, we get $2 = a + b + c = 1 + b + c \Rightarrow b + c = 1$.

Next, since (2, 1) is in the parabola, we get $1 = 4a + 2b + c = 4 + 2b + c \Rightarrow 2b + c = -3$.

So next, solving the system, we get $b + c = 1 \Rightarrow c = 1 - b \Rightarrow 2b + c = 2b + 1 - b = b + 1 = -3 \Rightarrow b = -4 \Rightarrow c = 1 - b = 1 + 4 = 5$.

Thus, the parabola is $y = x^2 - 4x + 5$.

Fig. 5



2. Two points and the y-intercept

First, given the y-intercept, we are given the value of c, so we have only to set up two equations, each of which is for a and b, of course.
So next, using the two points given, we can set up the two equations.
Then, solving the system of the two, we can get a and b, then the parabola.

For instance, the y-intercept is 4, and the parabola has (1, 2) and (2, 1).

Then first, assuming the parabola is $y = ax^2 + bx + c$, we get $c = 4$.

Next, since (1, 2) is in the parabola, we get $2 = a + b + c = a + b + 4 \Rightarrow a + b = -2$.

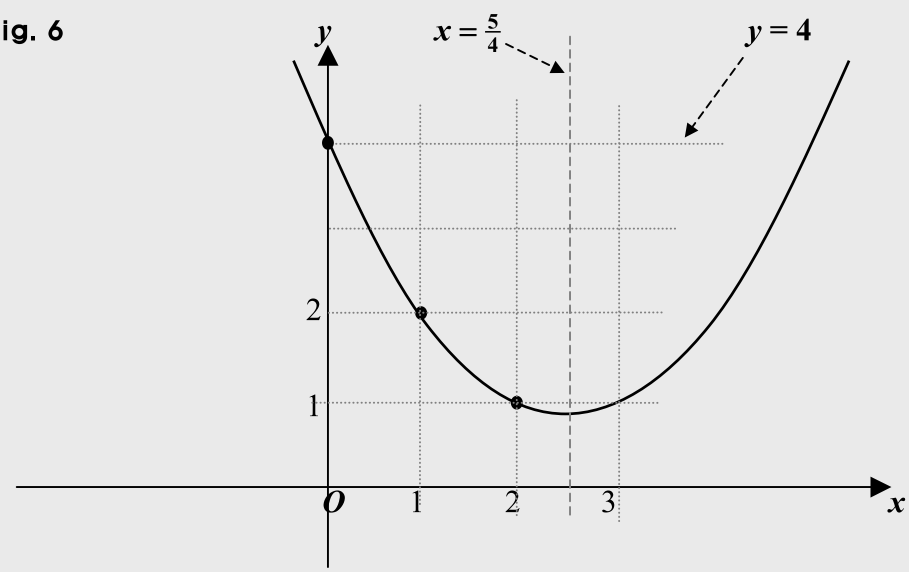
Next, using (2, 1), we get $1 = 4a + 2b + c = 4a + 2b + 4 \Rightarrow 4a + 2b = -3$.

So next, solving the system, we get $a + b = -2 \Rightarrow b = -2 - a$

$\Rightarrow 4a + 2b = 4a - 4 - 2a = 2a - 4 = -3 \Rightarrow 2a = 1 \Rightarrow a = \frac{1}{2}$.

$\Rightarrow b = -2 - a = -2 - \frac{1}{2} = -\frac{5}{2}$. Thus, the parabola is $y = \frac{1}{2}x^2 - \frac{5}{2}x + 4$.

Fig. 6



3. Two points and the axis of symmetry

First, given the axis of symmetry, we are given an equation for $\boldsymbol{a}$ and $\boldsymbol{b}$.

That's because assuming for instance, the axis of symmetry is $\boldsymbol{x = 2}$, we get $\boldsymbol{2 = -\frac{b}{2a}}$.

So we have only to set up two more equations.
So next, using the two points given, we can set up the two equations, each of which is for $\boldsymbol{a}$, $\boldsymbol{b}$, and $\boldsymbol{c}$.
And solving the system of the three equations, we can get $\boldsymbol{a}$, $\boldsymbol{b}$, and $\boldsymbol{c}$, then the parabola.

For instance, the axis of symmetry is $\boldsymbol{x = 3}$, and the parabola has (1, 2) and (2, 1).

Then first, assuming the parabola is $\boldsymbol{y = ax^2 + bx + c}$, we get $\boldsymbol{3 = -\frac{b}{2a} \Rightarrow b = -6a}$.

Next, using (1, 2), we get $\boldsymbol{2 = a + b + c}$, and using (2, 1), we get $\boldsymbol{1 = 4a + 2b + c}$.
So next, solving the system, we can get first $\boldsymbol{(4a + 2b + c) - (a + b + c) = 1 - 2 = -1}$

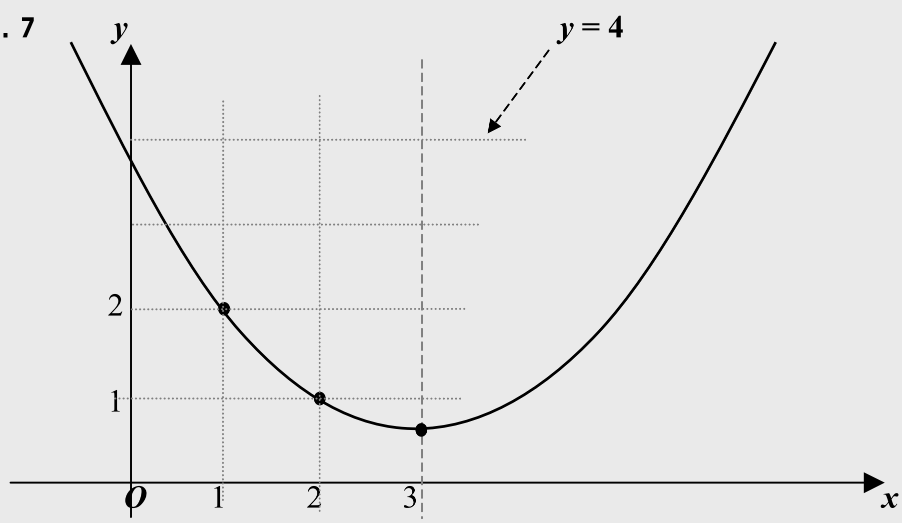
$\boldsymbol{\Rightarrow 3a + b = -1 \Rightarrow b = -3a - 1 \Rightarrow -6a = -3a - 1}$, because we have $\boldsymbol{b = -6a}$.

So we get $\boldsymbol{3a = 1 \Rightarrow a =}\frac{1}{3}\boldsymbol{\Rightarrow b = -6a = -2}$. And we have $\boldsymbol{2 = a + b + c}$.

So we get $\boldsymbol{2 =}\frac{1}{3}\boldsymbol{- 2 + c \Rightarrow c =} \frac{11}{3}$. Thus, the parabola is $\boldsymbol{y = \frac{1}{3}x^2 - 2x + \frac{11}{3}}$.

Notice that there are infinitely many parabolas that can share the same set of two roots.

Fig. 7



4. Two points and the sum of the two roots

Given the sum of the two roots, we are given an equation for $\boldsymbol{a}$ and $\boldsymbol{b}$.

That's because the sum of the two roots is $-\frac{b}{a}$ if the parabola is $y = ax^2 + bx + c$.

So assuming for instance, the sum is 3, we get $\mathbf{3} = -\frac{b}{a}$.

So we have only to set up two more equations.
So next, using the two points given, we can set up the two equations, each of which is for a, b, and c.
And solving the system of the three equations, we can get a, b, and c, then the parabola.
For instance, the sum of the two roots is 10, and the parabola has (1, 2) and (2, 1).

Then first, assuming the parabola is $y = ax^2 + bx + c$, we get $\mathbf{10} = -\frac{b}{a} \Rightarrow b = \mathbf{-10}a$.

Next, using (1, 2), we get $\mathbf{2} = a + b + c$, and using (2, 1), we get $\mathbf{1} = \mathbf{4}a + \mathbf{2}b + c$.

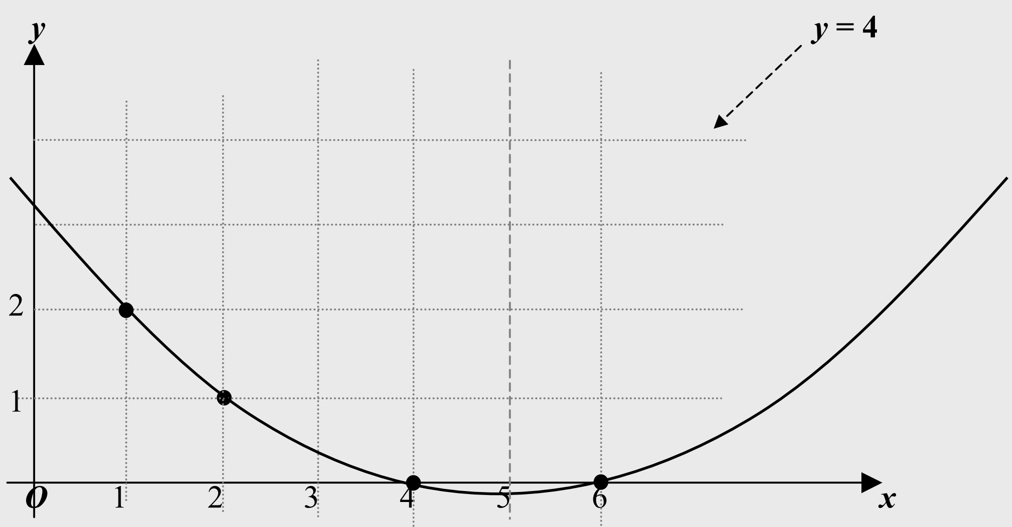
So next, solving the system, we can get first $(\mathbf{4}a + \mathbf{2}b + c) - (a + b + c) = \mathbf{1} - \mathbf{2} = \mathbf{-1}$

$\Rightarrow \mathbf{3}a + b = \mathbf{-1} \Rightarrow b = \mathbf{-3}a - \mathbf{1} \Rightarrow \mathbf{-10}a = \mathbf{-3}a - \mathbf{1}$, because we have $b = \mathbf{-10}a$.

So we get $\mathbf{7}a = \mathbf{1} \Rightarrow a = \frac{1}{7} \Rightarrow b = \mathbf{-10}a = -\frac{10}{7}$. And we have $\mathbf{2} = a + b + c$.

Thus, we get $\mathbf{2} = \frac{1}{7} - \frac{10}{7} + c \Rightarrow c = \frac{23}{7}$. So the parabola is $y = \frac{\mathbf{1}}{\mathbf{7}}x^2 - \frac{\mathbf{10}}{\mathbf{7}}x + \frac{\mathbf{23}}{\mathbf{7}}$.

Fig. 8



5. Two points and the product of the two roots

Given the product of the two roots, we are given an equation for $\boldsymbol{a}$ and $\boldsymbol{c}$.

That's because the product of the two roots is $\frac{c}{a}$ if the parabola is $y = ax^2 + bx + c$.

So assuming for instance, the product is 3, we get $3 = \frac{c}{a}$.

So we have only to set up two more equations.
So next, using the two points given, we can set up the two equations, each of which is for $\boldsymbol{a}$, $\boldsymbol{b}$, and $\boldsymbol{c}$.
And solving the system of the three equations, we can get $\boldsymbol{a}$, $\boldsymbol{b}$, and $\boldsymbol{c}$, then the parabola.
For instance, the product of the two roots is 10, and the parabola has (1, 2) and (2, 1).

Then first, assuming the parabola is $y = ax^2 + bx + c$, we get $10 = \frac{c}{a} \Rightarrow c = 10a$.

Next, using (1, 2), we get $2 = a + b + c$, and using (2, 1), we get $1 = 4a + 2b + c$.

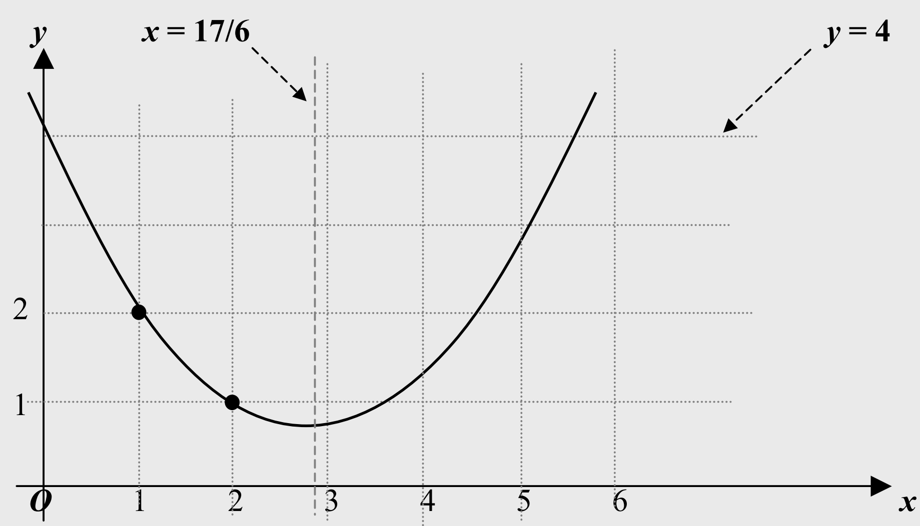
So next, solving the system, we can get first $(4a + 2b + c) - 2(a + b + c) = 1 - 4 = -3$

$\Rightarrow 2a - c = -3 \Rightarrow c = 2a + 3 \Rightarrow 10a = 2a + 3$, because we have $c = 10a$.

So we get $8a = 3 \Rightarrow a = \frac{3}{8} \Rightarrow c = 10a = \frac{15}{4}$. And we have $2 = a + b + c$.

Thus, we get $2 = \frac{3}{8} + b + \frac{15}{4} \Rightarrow b = -\frac{17}{8}$. So the parabola is $y = \frac{3}{8}x^2 - \frac{17}{8}x + \frac{15}{4}$.

Fig. 9



How though, does the parabola above have no root?

It has no real root, but has two complex roots, which are not real numbers, of course.
The problem says just the product of the roots, and the product is 10.
And in fact, the product of the two complex roots is 10.
We can get the two complex roots using the quadratic formula. So using it, we can get $\boldsymbol{x = \frac{17 \pm \sqrt{-71}}{6}}$, which are not real numbers, since what's inside the root sign is negative.
Taking however, the product of the two roots, we get

$$\frac{17+\sqrt{-71}}{6} \cdot \frac{17-\sqrt{-71}}{6} = \frac{17^2-(-71)}{36} = \frac{360}{36} = 10.$$

6. One point and the sum of and the product of the two roots.

Assuming for instance, m and n are the two roots, the sum is p, and the product is q, we can set $m + n = p$, and $mn = q$.

Then, the values of m and n are the two solutions to an equation $a(x^2 - px + q) = 0$.
That's because we have $m + n = p$, and $mn = q$, so we can get

$$a(x^2 - px + q) = 0 \Rightarrow x^2 - px + q = 0 \Rightarrow x^2 - (m + n)x + mn = 0$$

$$\Rightarrow (x - m)(x - n) = 0 \Rightarrow x = m \text{ or } n.$$

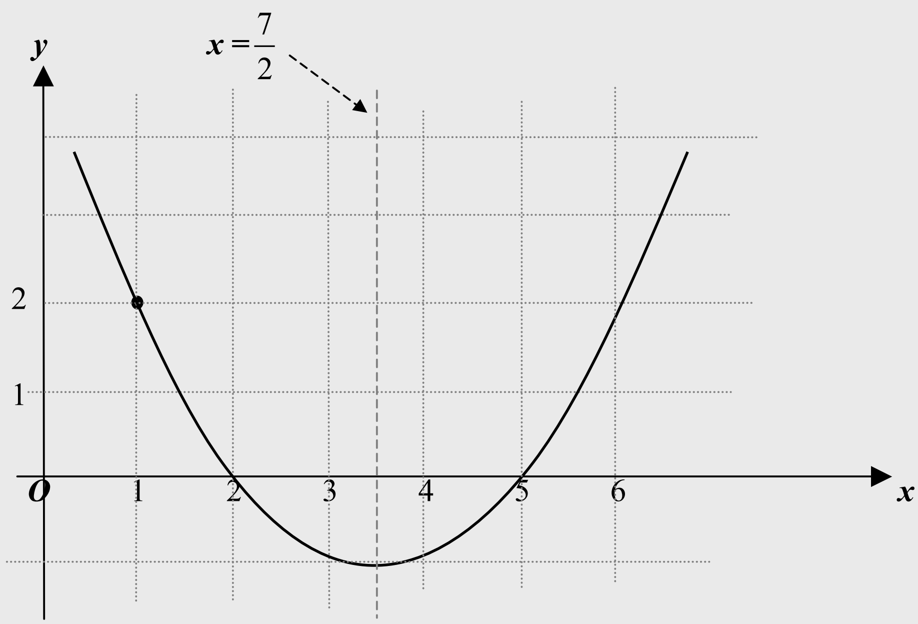
And we know m and n are the two roots of the parabola we want.
So $y = a(x^2 - px + q)$ is the parabola, because m and n are the two roots of it.

And we know p and q, because both are given. So using the one point given, we can get an equation for a, and solving the equation, we can get the value of a, then the parabola.

For instance, the sum of the two roots is 7, the product of the two is 10, and the parabola has (1, 2). Then first, we can assume that the parabola is $y = a(x^2 - 7x + 10)$.
Next, using (1, 2), we get $2 = a(1 - 7 + 10) = 4a \Rightarrow a = \frac{1}{2}$.

So the parabola is $y = \frac{1}{2}x^2 - \frac{7}{2}x + 5$.

Fig. A



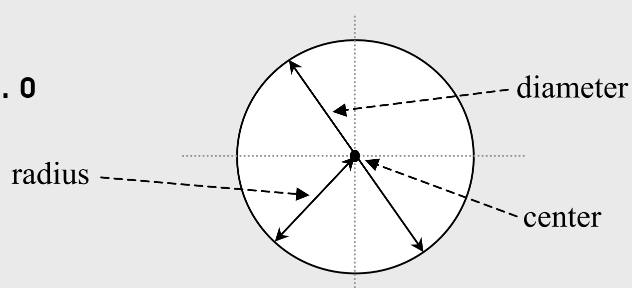
2.0. What is a circle?

A circle is in a plane, and is a set of points, from each of which, the distance to a particular point called the center is the same. And the same distance is called the radius.

Like a line and a parabola, a circle is one of the basic curves called conic sections, often called conics for short. So a circle is a conic.

And a circle is a simply closed curved-line segment that is symmetric and has infinitely many axes of symmetry, each of which is called a diameter.

Fig. 0



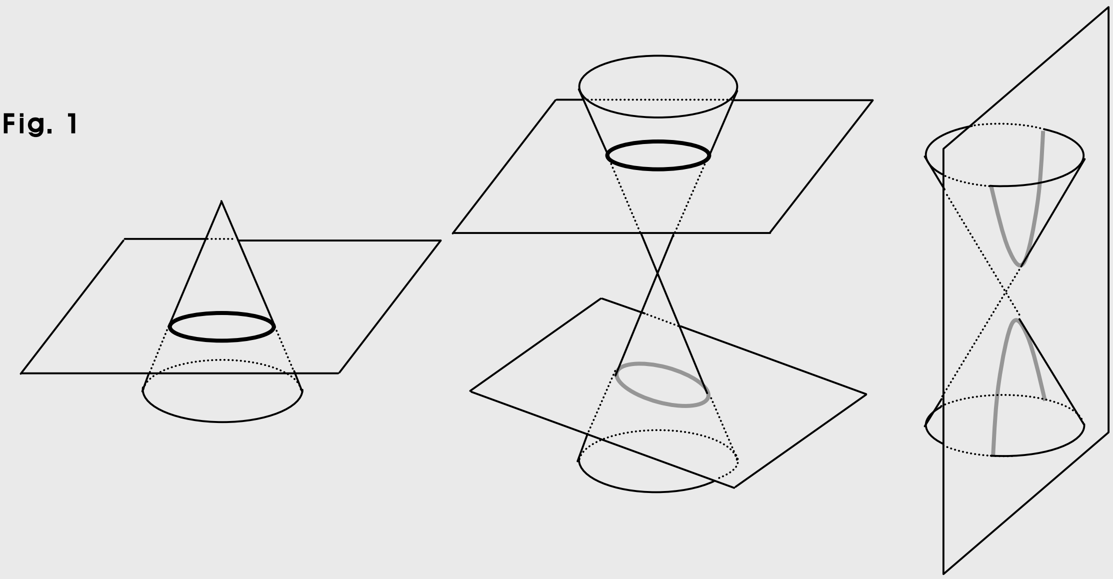
How then, can we get a circle?

As shown in the figure below, if cutting a right cone with a plane, we can get a cross section called a circle. So it's called a conic section, just called a conic for short.

And if a cone is a right cone, the line connecting the vertex of the cone and the center of the base is perpendicular to the base, and thus, makes a right angle (90^o) with the base.

The cross section in black below is a curve closed, and is a circle. So if the plane is parallel to the base, but does not include the vertex, the cross section is a circle.

Fig. 1



By the way, we have an object similar to a circle. Squeezing a circle, we get such an object, if you will. Such an object is called an ellipse, which is a conic, too, and is the closed curved line segment in gray in the figure above. So if the plane is neither parallel to the base, nor passing through the base, and does not include the vertex, the cross section is an ellipse.

If however, the plane is perpendicular to the base, and doesn't include the vertex, or if the plane is not parallel to the side of the cone (called the generator of the cone, too), and passes through the base, we get a cross section called a hyperbola.

And we can also explain a circle the way as follows.

Suppose that a point is moving along a curve in the ***x-y*** plane, and that the distance from the moving point to a particular point is constant. Then, the particular point is called the center, the distance is called the radius, and the curve is a circle.

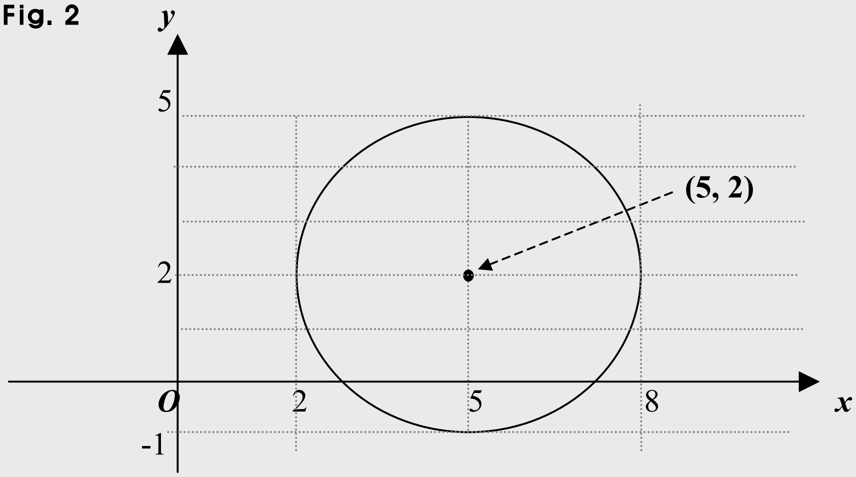
So if a point **(*x*, *y*)** is in a circle, no matter where the point **(*x*, *y*)** may be, the distance from the point **(*x*, *y*)** to the point called the center is constant, that is, the same.

How then can we explain or describe a particular circle?

Normally, in math, we put a circle in the ***x-y*** plain, and describe the circle specifying the center as well as the radius.

So for instance, if a circle is centered at a point (5, 2) in the ***x-y*** plane, and its radius is 3, we put the circle in a graph the way as follows.

Fig. 2



And in math, making a circle, we define it. So defining a circle particular, we make the particular circle. How then can we define a particular circle?

As in the cases of other conics as a parabola, we can define a circle using an equation that indicates the circle. So we can define a particular circle producing the equation of it.

For instance, assuming ***C*** is the circle described above, we can define ***C*** the way below.

$$(x-5)^2 + (y-2)^2 = 3^2$$

And the equation above is in the standard form, and is a standard equation of a circle.

Putting the standard equation above in the general form, we just expand (or simplify) it. Then, we get $x^2 + y^2 - 10x - 4y + 20 = 0$, which is a general equation of a circle.

And in general, putting a circle in the standard form, we can put it the way as follows.

$$(x - u)^2 + (y - v)^2 = r^2, \text{ where } r \neq 0$$

What then do we mean by u, v, and r?

The center is at (u, v), and the radius is r.

So just describing a particular circle, we often describe it specifying the radius. If however, we define a particular circle placed in the x-y plane, we want to specify the center and the radius.
And putting a circle in an equation in general, we can put it the way as follows.

$$x^2 + y^2 + ax + by + c = 0, \text{ where } a^2 + b^2 > 4c$$

And the equation above is in the general form, and is called a general equation of a circle.
What do we mean by, though, this: $a^2 + b^2 > 4c$?

It means that if the equation makes a circle, we get this: $a^2 + b^2 - 4c > 0$.
So if we get $a^2 + b^2 \leq 4c$, the equation does ***not*** make a circle. Usually though, we just put it this way: $x^2 + y^2 + ax + by + c = 0$, assuming implicitly, $a^2 + b^2 - 4c > 0$.

For instance, $x^2 + y^2 - 2x - 4y + 6 = 0$ seems indicating a circle, but does not indicate it.

That's because we get $a^2 + b^2 - 4c = (-2)^2 + (4)^2 - 4 \cdot 6 = 4 + 16 - 24 = -4 < 0$.
So we get $a^2 + b^2 - 4c < 0 \Rightarrow a^2 + b^2 < 4c$.

And putting in fact, the equation in the standard form, we get
$x^2 + y^2 - 2x - 4y + 6 = x^2 - 2x + 1 - 1 + y^2 - 4y + 4 - 4 + 6 = (x - 1)^2 + (y - 2)^2 + 1 = 0$

$\Rightarrow (x - 1)^2 + (y - 2)^2 = -1$, which is impossible, and does not indicate a circle.

And we can give a name to an equation, too. For instance, assuming C is the equation of the circle above, we can put C the way as follows.

$\boldsymbol{C(x, y) = x^2 + y^2 + ax + by + c = 0}$.

What then about the standard equation?

Assuming S is the equation $(x - u)^2 + (y - v)^2 = r^2$, we can put S the way below, too.

$\boldsymbol{S(x, y) = (x - u)^2 + (y - v)^2 - r^2 = 0}$, where $r \neq 0$.

What if however, $r = 0$?

Then, the equation does not indicate a circle but just a point. What point then is it?

It is a point at (u, v). It's because if $(x - u)^2 + (y - v)^2 = 0$, $(x - u)$ and $(y - v)$ both have to be 0 at the same time. So we get $x = u$ and $y = v$, and thus, we just get this: (u, v).

And we know that there is no circle of radius 0. So if S indicates a circle, we get $r \neq 0$.

And it is ***not*** the case where the equation $x^2 + y^2 = r^2$ is defined for x real and y real. ('x real' means every real number can be x.)

The equation is in fact, defined for $-r \leq x \leq r$, and $-r \leq y \leq r$.

That is, the domain is $-r \leq x \leq r$, and the range is $-r \leq y \leq r$.

That's simply because the circle is of radius r, and is centered at the origin $O(0, 0)$.

And moving the circle above in the amount of u along the x-axis, and in the amount of v along the y-axis, we get another circle as follows. $(x - u)^2 + (y - v)^2 = r^2$.
What circle then is it?

If we move the center of a circle, the circle moves the way the center gets moved.

So moving a circle, we move its center, keeping the radius intact, of course.
And in the case above, the center moves from the origin to the point (u, v).

So we can notice that the circle is $(x - u)^2 + (y - v)^2 = r^2$, since the center is (u, v) and the radius is r. And it is not the case either the equation is defined for x real and y real.

What then, are the domain and the range?

It is defined for $-r + u \le x \le a + r$, and $-r + v \le y \le r + v$.

That is, the domain is $u - r \le x \le u + r$, and the range is $v - r \le y \le v + r$.

So if a circle is of radius r and is centered at (u, v), we can put it in the standard equation more specifically the way as follows.

$$(x - u)^2 + (y - v)^2 = r^2 \text{ where } r \neq 0,\ u - r \le x \le u + r, \text{ and } v - r \le y \le v + r$$

Normally though, we just put it this way: $(x - u)^2 + (y - v)^2 = r^2$, assuming the others implicitly, of course.

And a circle has some physical properties as follows.

To begin with, a circle can be said to be the simplest line segments curved and closed.

Suppose next, we have designated a particular point in the x-y plane.
Then, a circle is a set of all points from each of which the distance to the particular point is constant. In other words, all the individual distances from all the points in the circle to the particular point are the same.   What then, is the particular point?

It is the center of the circle, of course.   What then is the distance stated above?

It is the radius, of course. So a radius is a line segment connecting the center and a point in the circle.   What then is a line segment twice the radius?

It is a diameter, of course, and is passing through the center, and connecting two facing points in the circle.
How then can we specify a circle? That is, what determines a particular circle?

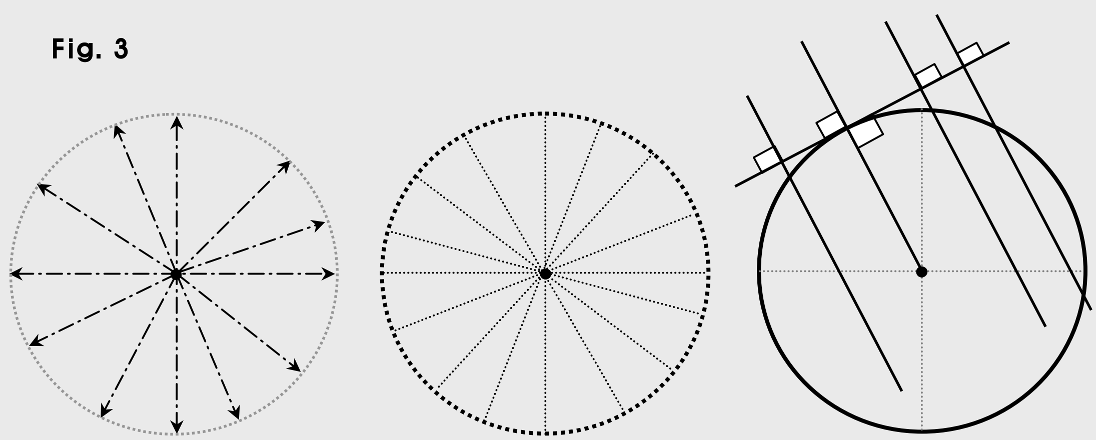
The center and radius do so. Therefore, we can specify a circle by the center and radius.
A radius is the distance from the center of a circle to a point in the circle.
There are quite a few radii in a circle, then.

They all have the same length, of course, and there are in fact, infinitely many of those.
Radii radiate from the center of a circle, if you will.
The radius radiates in every direction.
That is, the radii radiate from the center of a circle in all directions.

Besides, any line tangent to a circle is perpendicular to a line passing through the tangent point and the center of the circle. Thus, a line tangent to the circle is perpendicular to the line passing through the tangent point and the center of the circle.

So a line tangent to a circle is perpendicular to the radius that includes the tangent point. Thus, a line tangent to a circle is perpendicular to any line parallel to that radius, too.

Fig. 3



2.1. Equations for Circles 1

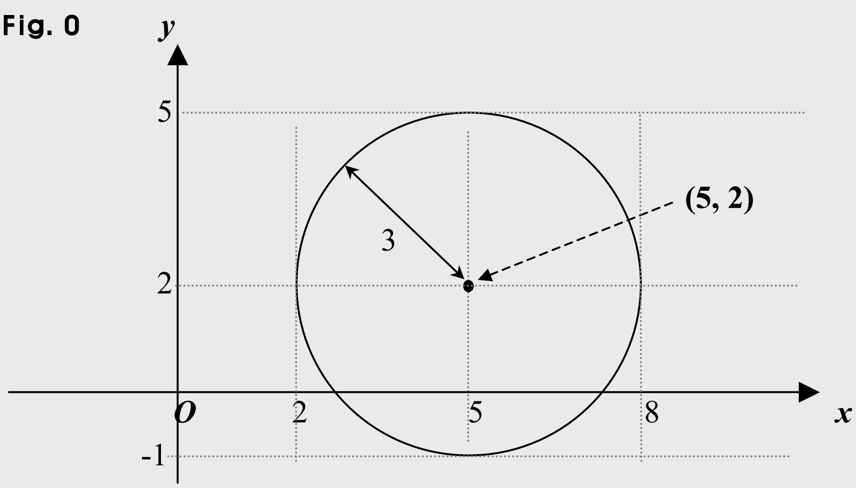
To begin with, putting a circle in an equation, we can get this:

$$(x - 5)^2 + (y - 2)^2 = 9$$

What circle then is it?

We can put it this way, too: $(x - 5)^2 + (y - 2)^2 = 3^2$.
And it is a circle centered at (5, 2), and its radius is 3.
So we can put the circle in the ***x*-*y*** plane the way as follows.

Fig. 0



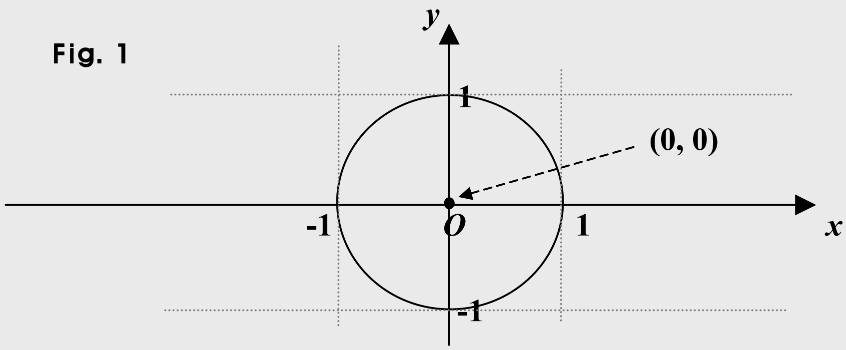
What circle then is the circle as follows?

$$x^2 + y^2 = 1$$

We can put it this way, too: $x^2 + y^2 = 1^2$, or $(x - 0)^2 + (y - 0)^2 = 1^2$.

So the circle is centered at (0, 0), that is, at the origin, and the radius is 1.
And we call such a circle a unit circle, because the radius is 1, which is often called a unit, too. And we can put the unit circle in the ***x-y*** plane the way as follows.

Fig. 1

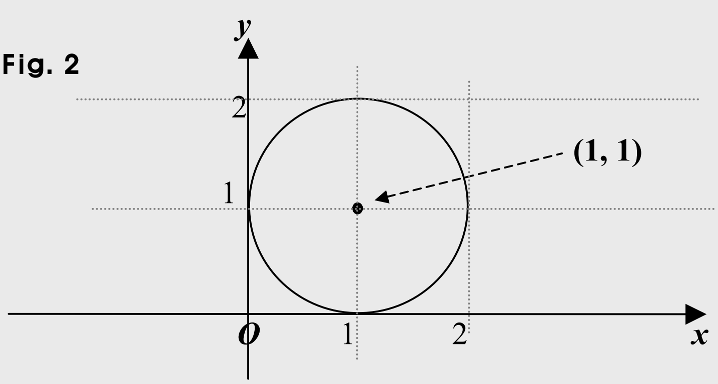


What if the center is at (1, 1), and the radius is 1?

Then, the circle will be as follows. $(x-1)^2 + (y-1)^2 = 1^2$

And we can put the circle above in a graph the way as follows.

Fig. 2



How then can we get such equations as above?

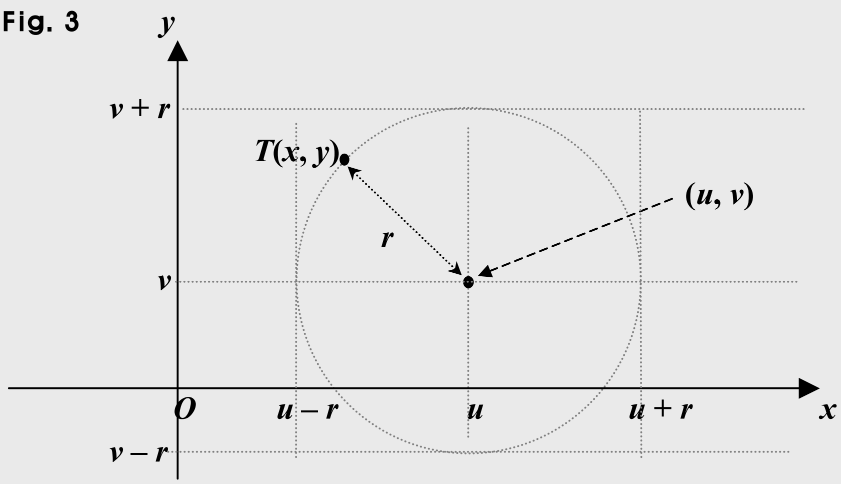
We know the fact that a circle is a set of points, from each of which, the distance to the point called the center is constant, that is, the same. So?

So using the fact above, we can get the equation of a circle.

Thus, to begin with, suppose (u, v) is the center of a circle called C, a point $T(x, y)$ is an arbitrary point in the circle C, that is, a random point representing all the points in C, and the distance from T to the center is r.

Then, the radius is the distance, that is, r, because every point in the center is the distance away from the center. So we can put the point T in the x-y plane the way as follows.

Fig. 3



Suppose next, the point T is moving along the circle C.
Then, the distance from $T(x, y)$ to the center is r.   How?

No matter where the point T may be in the circle C, the distance from the point T to the center (u, v) is always the same, and is r.   Using thus, the fact above, we can get the equation of E.   How then can we use the fact?

We can use the distance formula, often called Pythagorean theorem.

What then do we get?

We get an equation expressed in terms of the coordinates of the point T, that is, we get an equation in terms of x and y. And we call such an equation a connective equation.

So in this case, the connective equation is the equation that connects the two variables used as the coordinates of the arbitrary point $\boldsymbol{T(x, y)}$ in the curve called the circle $\boldsymbol{C}$. How then do we call the equation?

The equation explains every point in the circle $\boldsymbol{C}$, that is, it indicates the circle $\boldsymbol{C}$.
So the equation is called the equation of the circle $\boldsymbol{C}$.
So let's now get the equation. How then can we apply the distance formula?

Assuming $\boldsymbol{d}$ is the distance between two points, $\boldsymbol{\Delta x}$ is the difference in $\boldsymbol{x}$-coordinates, and $\boldsymbol{\Delta y}$ is the difference in $\boldsymbol{y}$-coordinates, we can put the formula the way as follows.

$$d^2 = (\Delta x)^2 + (\Delta y)^2$$

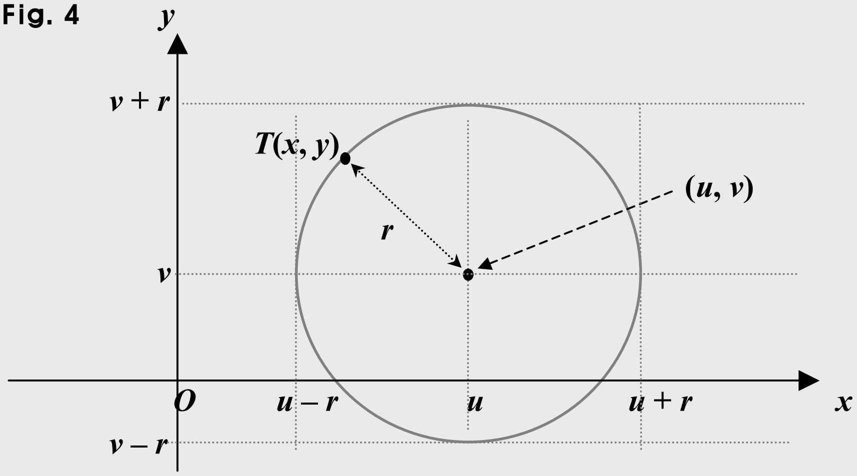
So using the distance formula, since the distance is $\boldsymbol{r}$, we can get $\boldsymbol{r^2 = (x - u)^2 + (y - v)^2}$.

And the equation above is expressed in terms of the coordinates of the point $\boldsymbol{T(x, y)}$, so it is an equation in terms of $\boldsymbol{x}$ and $\boldsymbol{y}$. And we call such an equation a connective equation.

So in this case, the connective equation is the equation that connects the two variables used as the coordinates of the arbitrary point $\boldsymbol{T(x, y)}$ in the curve called the circle $\boldsymbol{C}$.

Thus, the equation of the circle $\boldsymbol{C}$ is $\boldsymbol{(x - u)^2 + (y - v)^2 = r^2}$, where $\boldsymbol{r \neq 0}$.
And putting the circle $\boldsymbol{C}$ in the $\boldsymbol{x}$-$\boldsymbol{y}$ plane, we get this:

Fig. 4



So we can now say that the circle C is centered at the point (u, v), and that the radius is r, that is, the diameter is $2r$.
And putting the circle C in its equation, we get $(x - u)^2 + (y - v)^2 = r^2$, where $r \neq 0$.
What equation then do we get if the center is at (-1, 3), and the radius is 2?

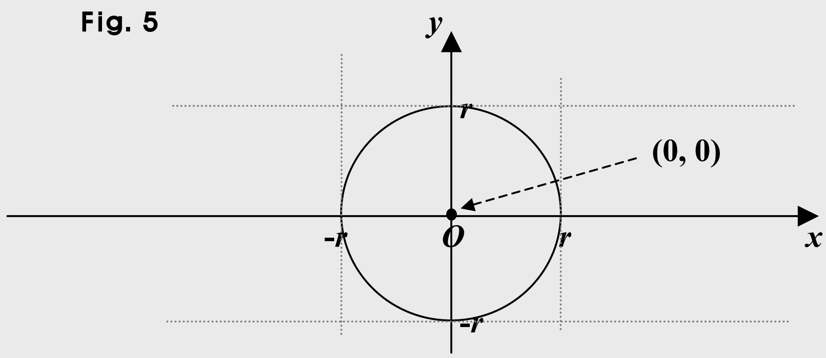
It is $(x + 1)^2 + (y - 3)^2 = 2^2$. So for another instance, if the center is at the origin, and the radius is r, the equation is $x^2 + y^2 = r^2$

And if a circle is a unit circle centered at (2, -1), the equation is $(x - 2)^2 + (y + 1)^2 = 1^2$.
What then, about the domain and the range?

If the equation of a circle is $(x - u)^2 + (y - v)^2 = r^2$, it is defined for $-r + u \leq x \leq a + r$, and $-r + v \leq y \leq r + v$.
That is, the domain is $u - r \leq x \leq u + r$, and the range is $v - r \leq y \leq v + r$.
What if the circle is $x^2 + y^2 = r^2$?

Then, the domain is $-r \leq x \leq r$, that is, $|x| \leq r$, and also, the range is $-r \leq y \leq r$ or $|y| \leq r$.

Fig. 5



Now, putting threads together, we have four equations as follows.

The circle of radius 1 centered at the origin is $x^2 + y^2 = 1$.

A unit circle, a circle of radius 1 centered at (a, b) is $(x - a)^2 + (y - b)^2 = 1$.

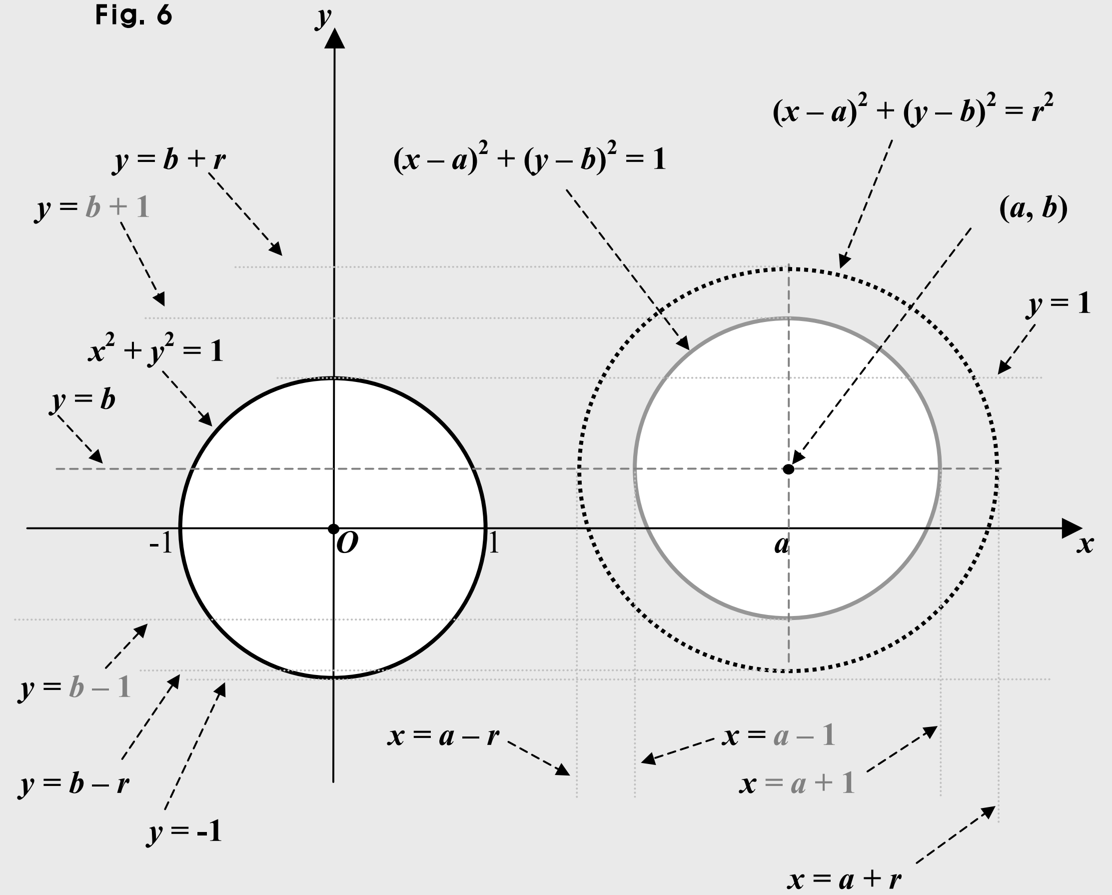
A circle of radius r centered at (a, b) is $(x - a)^2 + (y - b)^2 = r^2$ where a, b, and r are constant and $r > 0$.

The equation above is in the standard form.
And putting a circle in an equation in the general from, we get this:

$$x^2 + y^2 + ax + by + c = 0 \text{ where } a^2 + b^2 > 4c$$

Let's now, put some circles in one graph.
In Fig. 6 below, a circle centered at (a, b) with the radius of r is dotted in black.
A unit circle centered at (a, b) is in gray.
And the unit circle centered at the origin is in black.

Fig. 6



And if C is the unit circle centered at the origin $O(0, 0)$, moving the circle C by a along the x-axis, and by b along the y-axis, we get a unit circle centered at (a, b).

And the two circles on the right are called concentric circles, since both have the same center.

2.2. Equations for Circles 2

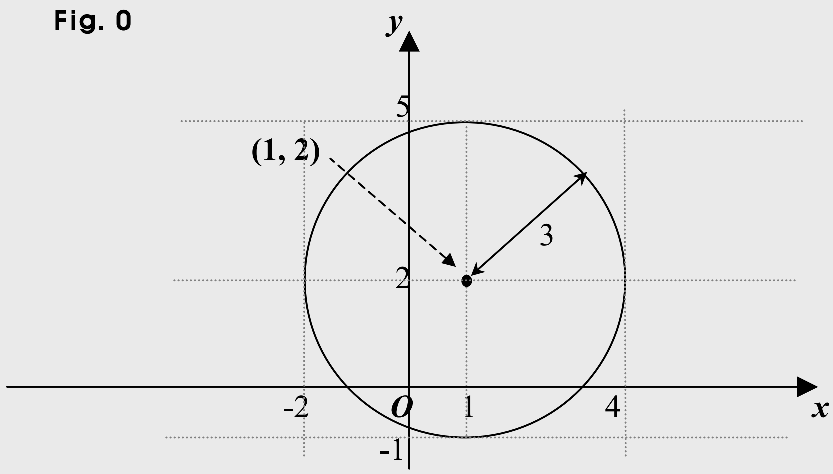
To begin with, putting a circle in an equation, we can get this:

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

What circle then is it?

It is a circle of radius 3 centered at (1, 2).
So we can put the circle in the ***x-y*** plane the way as follows.

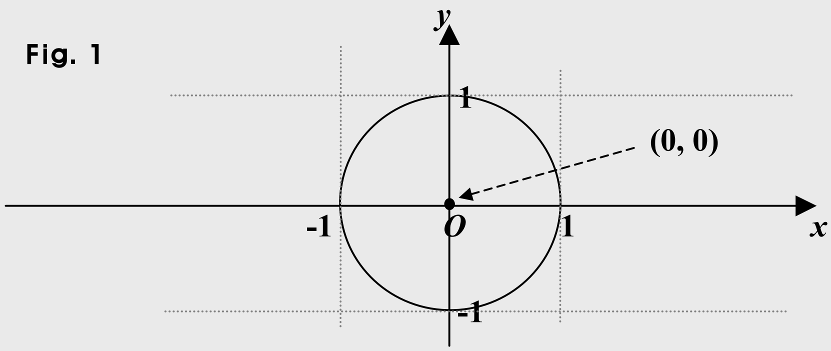
Fig. 0



What circle then is this? $x^2 + y^2 = 1$

It is a circle of radius 1 centered at (0, 0), that is, at the origin, and is called a unit circle.
And we can put the unit circle in the ***x-y*** plane the way as follows.

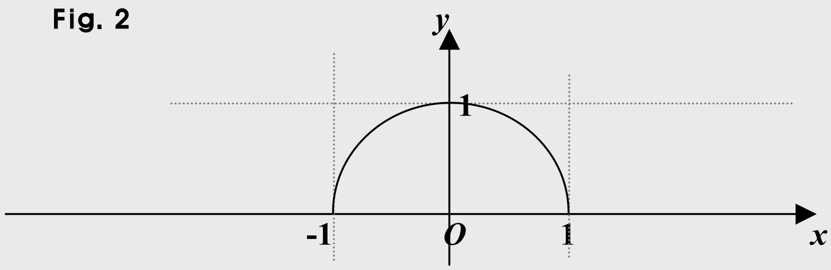
Fig. 1



What then about the equation as follows? $y = \sqrt{1 - x^2}$

It is not a circle but a half circle, and is the upper half of a circle where the radius is 1, and the center is at the origin. So we can put it in a graph the way as follows.

Fig. 2

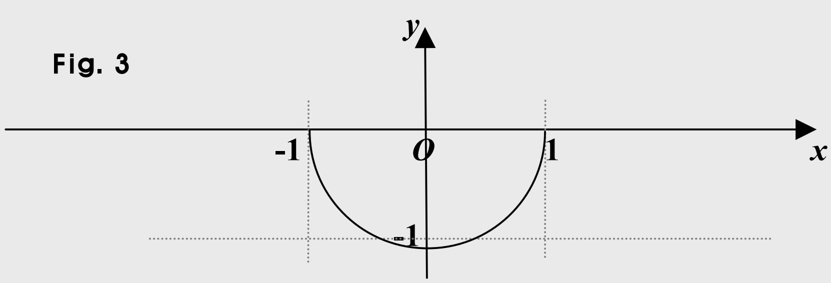


And in fact, we can get $y = \sqrt{1 - x^2} \Rightarrow y^2 = 1 - x^2 \Rightarrow x^2 + y^2 = 1$.

What then about this: $y = -\sqrt{1 - x^2}$?

It is a half circle, too, but is, this time, the lower half of the circle where the radius is 1, and the center is at the origin. So we can put it in a graph the way as follows.

Fig. 3

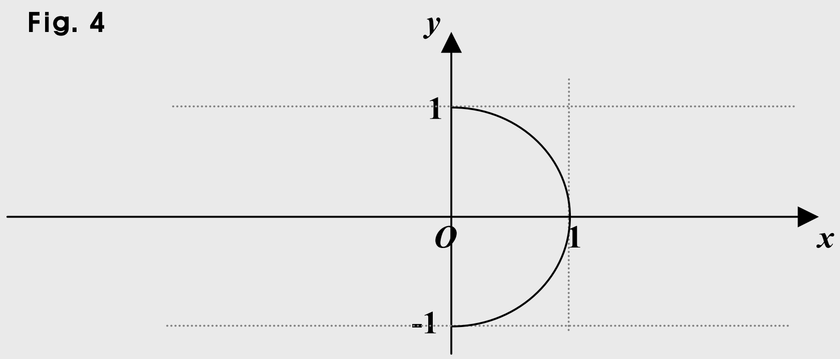


And we can get this, too, of course: $\boldsymbol{y = -\sqrt{1-x^2} \Rightarrow y^2 = 1 - x^2 \Rightarrow x^2 + y^2 = 1}$.

What then, about this equation: $\boldsymbol{x = \sqrt{1-y^2}}$?

It is a half circle, too, but is, this time, the right half of the circle of radius 1 centered at the origin. So we can put it in a graph the way as follows.

Fig. 4



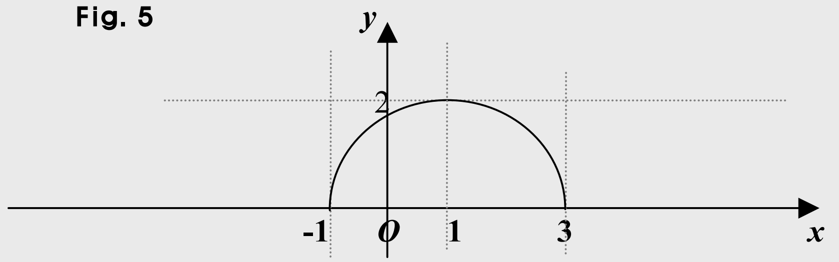
And we get this, too, of course: $\boldsymbol{x = \sqrt{1-y^2} \Rightarrow x^2 = 1 - y^2 \Rightarrow x^2 + y^2 = 1}$.

And the left half is $\boldsymbol{x = -\sqrt{1-y^2}}$.

What then about the equation as follows? $\boldsymbol{y = \sqrt{4-(x-1)^2}}$

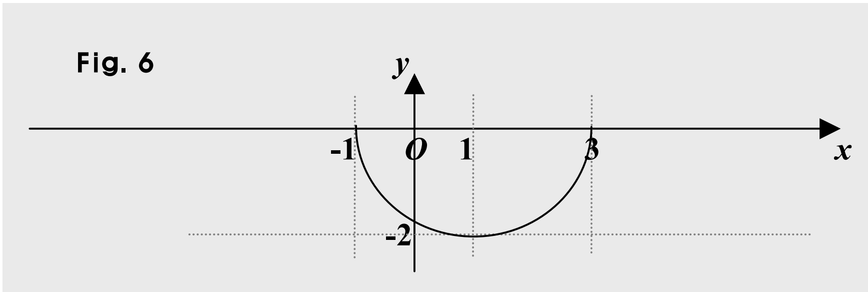
It is a half circle, too, but is, this time, the upper half of a circle where the radius is 2, and the center is at (1, 0). So the graph is as follows.

Fig. 5



And we get this, of course: $\boldsymbol{y = \sqrt{4-(x-1)^2} \Rightarrow y^2 = 4 - (x-1)^2 \Rightarrow (x-1)^2 + y^2 = 4 = 2^2}$.

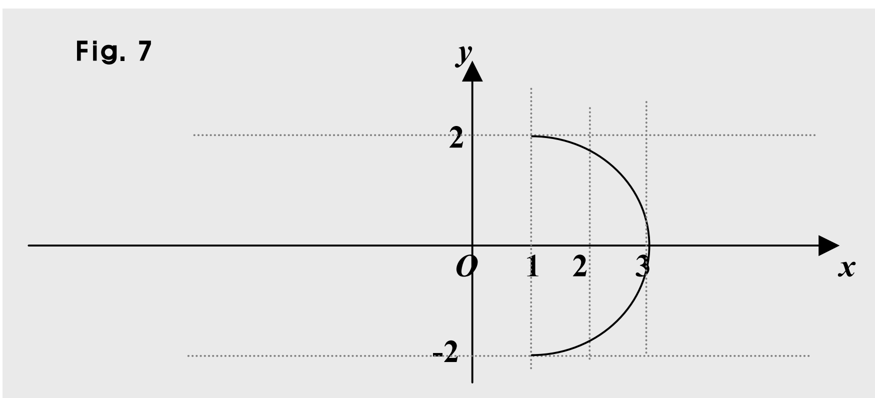
Similarly, the lower half is $\boldsymbol{y = -\sqrt{4-(x-1)^2}}$. And the graph is this:



What then about the right half?

Its equation is not this: $x = \sqrt{4-(y-1)^2}$, but this: $x - 1 = \sqrt{4-y^2}$.

In other words, it is $x = \sqrt{4-y^2} + 1$.   And the graph is as follows.



And we get $x - 1 = \sqrt{4-y^2} \Rightarrow (x-1)^2 = 4 - y^2 \Rightarrow (x-1)^2 + y^2 = 4 = 2^2$.

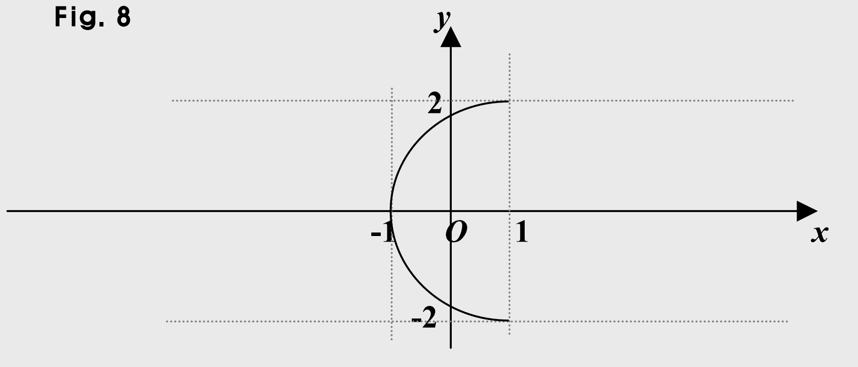
What then about the left half?

Its equation is this: $x - 1 = -\sqrt{4-y^2}$.

We can put it this way, too: $x = 1 - \sqrt{4-y^2}$.

And the graph is as follows.

Fig. 8

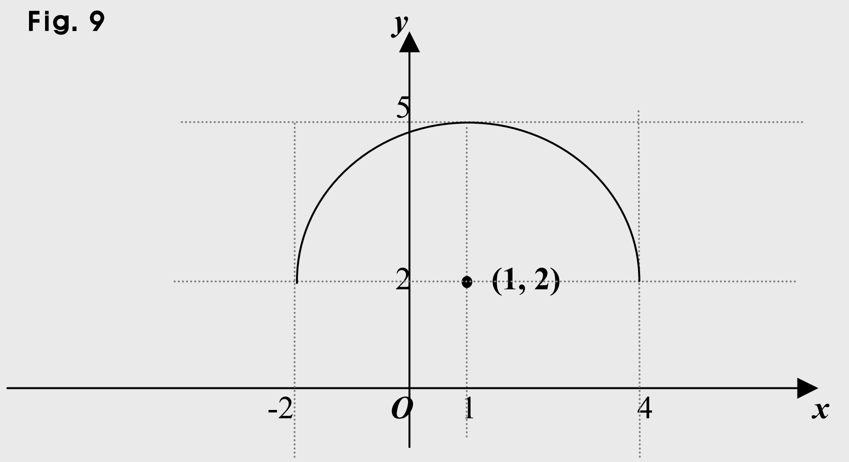


What about the upper half of a circle where the radius is 3, and the center is at (1, 2)?

Its equation is $\boldsymbol{y-2=\sqrt{9-(x-1)^2}}$.

And we can put it this way, too: $\boldsymbol{y=\sqrt{9-(x-1)^2}+2}$. And the graph is as follows.

Fig. 9



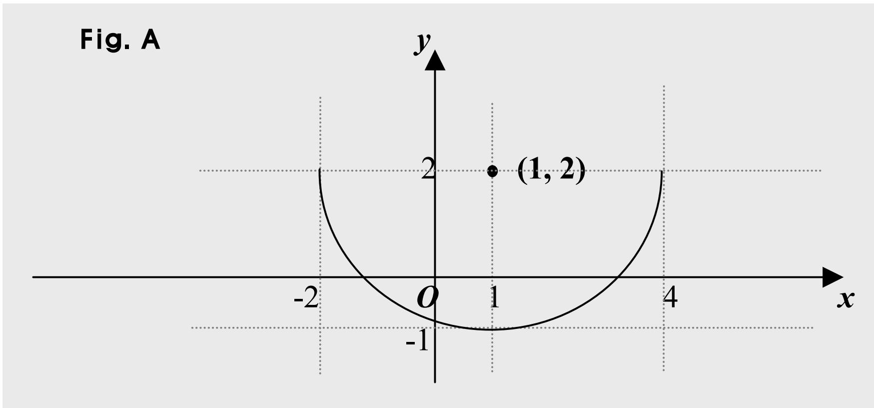
What then about the lower half?

Its equation is $\boldsymbol{y-2=-\sqrt{9-(x-1)^2}}$.

Also, we can put it this way: $\boldsymbol{y=-\sqrt{9-(x-1)^2}+2}$.

It can be put this way, too: $\boldsymbol{y=2-\sqrt{8-x^2+2x}}$ if what's inside the root is simplified.

And the graph is this:



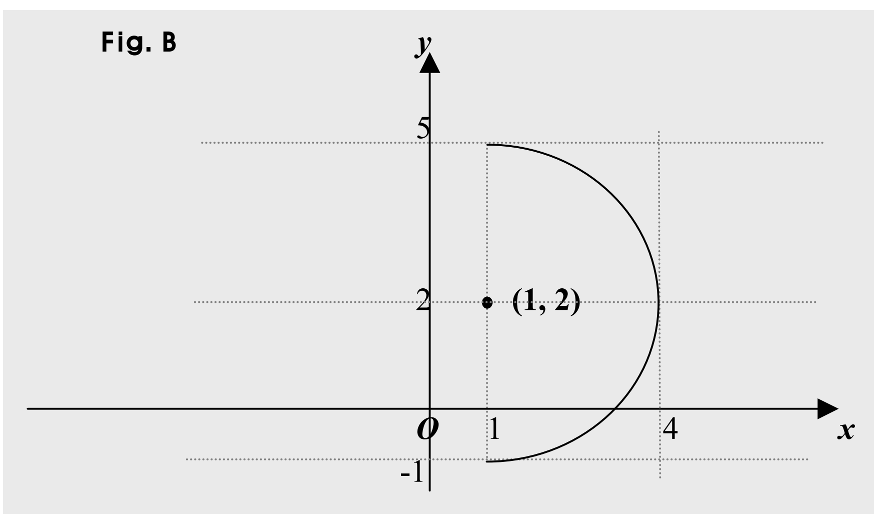
What about the right half?

Its equation is $\boldsymbol{x-1=\sqrt{9-(y-2)^2}}$.

Also, we can put it this way: $\boldsymbol{x=\sqrt{9-(y-2)^2}+1}$.

Or it can be put this way, too: $\boldsymbol{x=1+\sqrt{5-y^2+4y}}$ if what's inside the root is simplified.

And the graph is as follows.



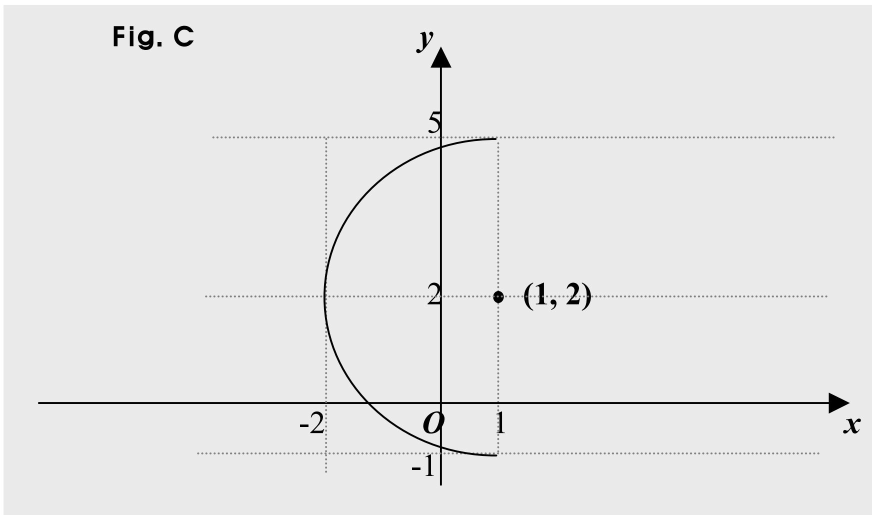
What about the left half?

Its equation is this: $\boldsymbol{x-1=-\sqrt{9-(y-2)^2}}$.

Also, we can put it this way, too: $\boldsymbol{x=1-\sqrt{9-(y-2)^2}}$.

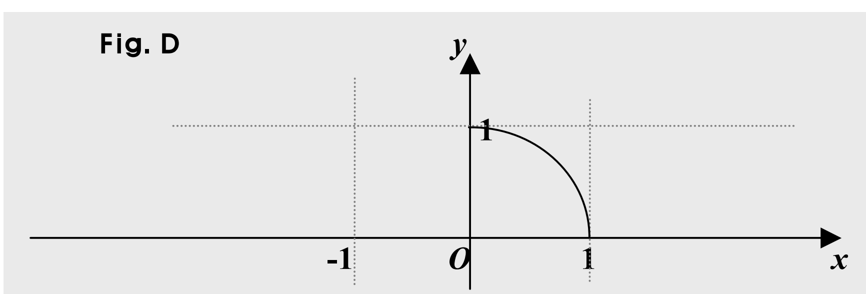
It can be put this way, too: $\boldsymbol{x=1-\sqrt{5-y^2+4y}}$ if what's inside the root is simplified.

And the graph is this:



What then about a quarter circle, that is, a quarter of a circle?

If it is the right upper quarter, and the whole circle is of radius 1, and is centered at the origin, the graph is as follows.



Then, we can see that the quarter is defined for $\boldsymbol{0 \le x \le 1}$.

And the equation of the upper half is $\boldsymbol{y=\sqrt{1-x^2}}$.

So the upper quarter is $\boldsymbol{y=\sqrt{1-x^2}}$ for $\boldsymbol{0 \le x \le 1}$, which is the domain.

And just producing this: $y = \sqrt{1 - x^2}$ without the domain, we mean the largest possible domain, so the domain is **-1** $\le x \le$ **1**.

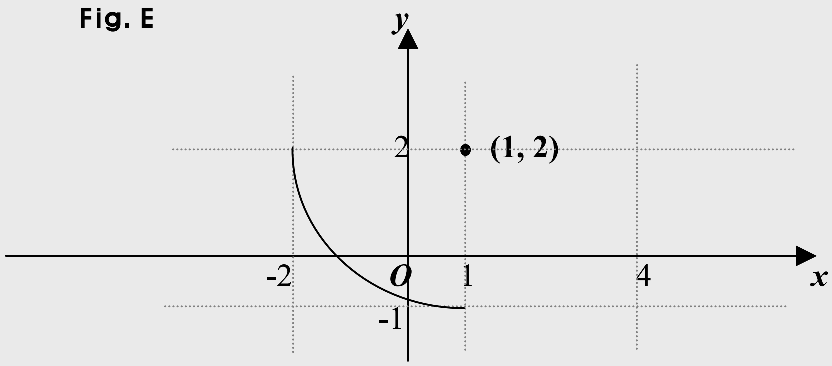
So just showing this: $y = \sqrt{1 - x^2}$, we mean this: $y = \sqrt{1 - x^2}$ for **-1** $\le x \le$ **1**.

And by the same token, the lower quarter is $y = -\sqrt{1 - x^2}$ for $0 \le x \le 1$.

What then about the lower left quarter of a circle where the center is at (1, 2) and the radius is 3?

First, the lower half is $y - 2 = -\sqrt{9 - (x-1)^2}$. And the graph is as follows.

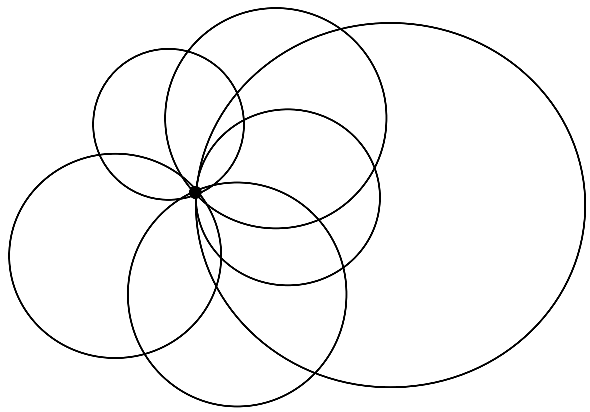
Fig. E



So the lower left quarter is $y - 2 = -\sqrt{9 - (x-1)^2}$ for **-2** $\le x \le$ **1**.

And with a similar manner, we can define many other parts of a circle.
So you may want to try some yourself.

2.3. Equations for Circles 3



In math, making a circle or defining a circle, we put it in an equation, or produce the equation of the circle. What then do we need making or defining a circle?

We need the radius and the center. What then do we do with those?

We can get the equation of the circle putting them into a template. And the template is called the standard equation for circles. So we can call it a formula, too. And the template is

$$(x - a)^2 + (y - b)^2 = r^2,$$

where (a, b) is the center, and r is the radius.

So for instance, if (-2, 3) is the center, and 2 is the radius, we can define the circle, and the circle is $(x + 2)^2 + (y - 3)^2 = 2^2$.

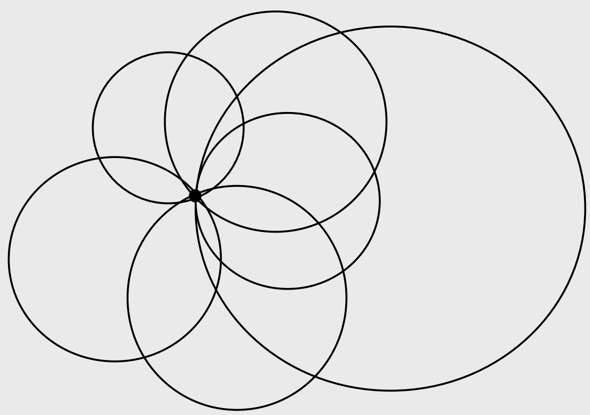
Thus, given the center and the radius, we can quickly define the circle, that is, produce the equation of it using the template above.
What if we are given only either of the two?

Given the center only, we can come up with infinitely many circles that share the center. And we call those circles concentric circles.

So given the center only, we cannot find the circle.
Next, given the radius only, we can come up with infinitely many circles, too, and each circle has a different center. So given the radius only, we cannot find the circle, either. What if this time, given a particular point, can we find a particular or unique circle passing through the particular point?

No, we can't. That's because we can have more than one circle that can pass through the particular point. We can have in fact, infinitely many of those.

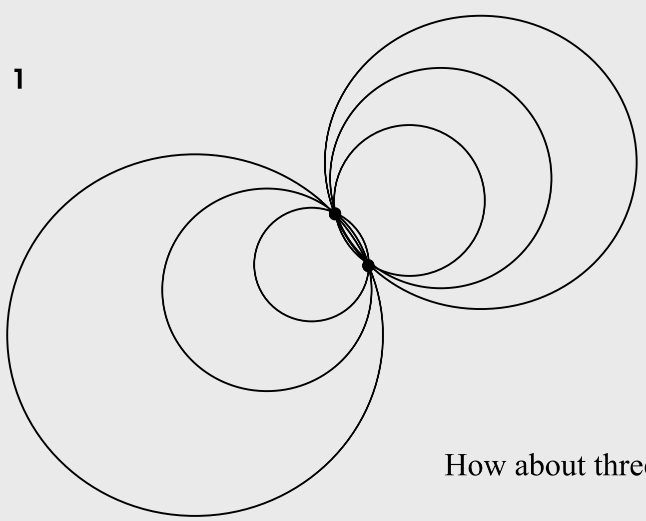
Fig. 0



What if this time, we are given two particular points?

Then again, we can construct infinitely many circles that can share the two points.

Fig. 1

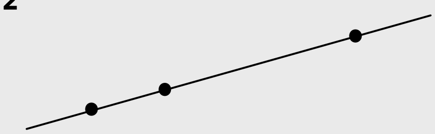


How about three particular points?

Then, we are likely to construct a particular circle. Why not definitely but likely?

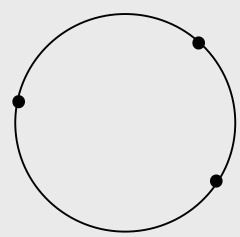
That's because if the three points are in a line, we are not able to make a circle.

Fig. 2



So the three points are not in a line if a circle includes the three.

Fig. 3



Having three points not in a line, we can make (define) a particular circle.
That is, only one circle passes though the three points. How then can we get the circle?

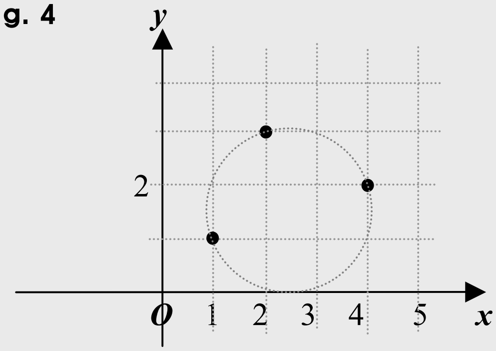
We can easily get it using the equation as follows. $\boldsymbol{x^2 + y^2 + ax + by + c = 0}$. How?

Putting each of the three points into the equation, we get an equation for $\boldsymbol{a}$, $\boldsymbol{b}$, and $\boldsymbol{c}$.
So we get a system of three equations for $\boldsymbol{a}$, $\boldsymbol{b}$, and $\boldsymbol{c}$.
And solving the system, we can get the equation of the circle.

What then, about the standard equation as follows? $\boldsymbol{(x - a)^2 + (y - b)^2 = c^2}$

We can still use it in that case, too, where three points not in a line are given.
So let's now find, for instance, a circle that has three points **(1, 1)**, **(2, 3)**, and **(4, 2)** using the standard equation. Putting the three points in a graph, we get this:

Fig. 4



We'll find the center first, and then, get the radius.

The three points are in a circle, so each point is the same distance away from the center. We can use the distance formula to find the center, and the formula is $\boldsymbol{d^2 = (\Delta x)^2 + (\Delta y)^2}$.

In the formula above, $\boldsymbol{d}$ is the distance between two points, $\boldsymbol{\Delta x}$ is the difference between $\boldsymbol{x}$-coordinates, and $\boldsymbol{\Delta y}$ is the difference between $\boldsymbol{y}$-coordinates.

So assuming first, the center is $\boldsymbol{(s, t)}$, since the distance from each point to the center is the same, we can get two equations the way as follows.

One is $\boldsymbol{(a - 1)^2 + (b - 1)^2 = (a - 2)^2 + (b - 3)^2}$ if we use two points **(1, 1)** and **(2, 3)**.

And the other is $\boldsymbol{(s - 1)^2 + (t - 1)^2 = (s - 4)^2 + (t - 2)^2}$ if we use **(1, 1)** and **(4, 2)**.

So we now have a system of two equations for $\boldsymbol{s}$ and $\boldsymbol{t}$.
The equations above look complicated, but can be simplified by some factorizations.

We have a factorization identity, which is $\boldsymbol{u^2 - v^2 = (u + v)(u - v)}$. So we get

$$\boldsymbol{(s - 1)^2 + (t - 1)^2 = (s - 2)^2 + (t - 3)^2 \Rightarrow (s - 1)^2 - (s - 2)^2 = (t - 3)^2 - (t - 1)^2}$$

$$\boldsymbol{\Rightarrow \{(s - 1) + (s - 2)\}\{(s - 1) - (s - 2)\} = \{(t - 3) + (t - 1)\}\{(t - 3) - (t - 1)\}}$$

$$\boldsymbol{\Rightarrow (2s - 3)\cdot 1 = (2t - 4)\cdot(-2) \Rightarrow 2s - 3 = -4t + 8 \Rightarrow 2s + 4t = 11}.$$

$$(s-1)^2+(t-1)^2=(s-4)^2+(t-2)^2 \Rightarrow (s-1)^2-(s-4)^2=(t-2)^2-(t-1)^2$$

$$\Rightarrow \{(s-1)+(s-4)\}\{(s-1)-(s-4)\}=\{(t-2)+(t-1)\}\{(t-2)-(t-1)\}$$

$$\Rightarrow (2s-5)\cdot 3=(2t-3)\cdot(-1) \Rightarrow 6s-15=-2t+3 \Rightarrow 6s+2t=18 \Rightarrow 3s+t=9.$$

Thus, we get a new system simplified as follows. $2s + 4t = 11$, and $3s + t = 9$.
Solving the system above, we can find the center (s, t).
Then, we can directly put the center into the standard form $(x - s)^2 + (y - t)^2 = c^2$.

So next, we want to get the radius, which is c in this case.
We don't really have to find c, though. Why not?

We have only to find c^2, since the equation $(x - a)^2 + (y - b)^2 = c^2$ has c^2 and not c.

And the actual solution to the system is $s = 2.5$, and $t = 1.5$.

So we get $c^2 = (s-1)^2 + (t-1)^2 = (2.5-1)^2 + (1.5-1)^2 = 1.5^2 + 0.5^2$

$$=\left(\frac{3}{2}\right)^2+\left(\frac{1}{2}\right)^2=\frac{9+1}{4}=\frac{10}{4}=\frac{5}{2}$$

Therefore, the circle is $(x-\frac{3}{2})^2+(y-\frac{1}{2})^2=\frac{5}{2}$

Anyway, finding c, we get $c^2=\frac{10}{4} \Rightarrow c=\pm\frac{\sqrt{10}}{2}$, and c is a radius, so we get $c=\frac{\sqrt{10}}{2}$.

Suppose this time, two points are given, one is the center, and the other is in a circle.
Then, the center is given, so we can take advantage of the standard equation as follows.

$(x - a)^2 + (y - b)^2 = c^2$, where (a, b) is the center, and c is the radius.

First, putting the center into the form, we get an equation where c is the only unknown.
So next, to find c, we can put the other point into the equation.

Thus, for instance, if the center is $(1, 2)$, putting it into the standard equation, we get
$(x - a)^2 + (y - b)^2 = c^2 \Rightarrow (x - 1)^2 + (y - 2)^2 = c^2$.

And next, if the point given is **(2, 2)**, putting it into the equation above, we get

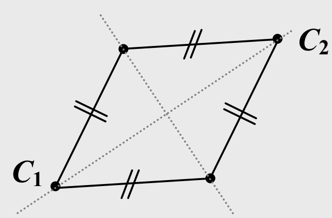
$(x-1)^2 + (y-2)^2 = c^2 \Rightarrow (2-1)^2 + (2-2)^2 = c^2 \Rightarrow 1 = c^2$.

So the circle is $(x-1)^2 + (y-2)^2 = 1$, which is a unit circle centered at **(1, 2)**.

Suppose this time, we are given two points and a radius, but the center is not given. Then, we get to find not one circle particular but two particular circles. Why?

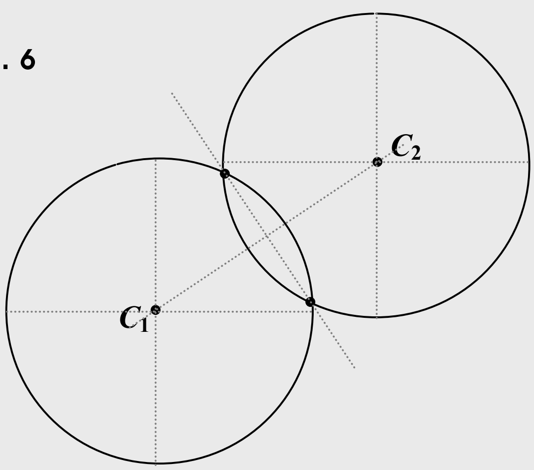
That is because in a plane, we can have not one but two particular points, each of which is the same distance away from the two points given, and the same distance is the radius. For instance, we can have two points C_1 and C_2 in the same plane, and each of C_1 and C_2 is the same distance away from the two points given as in Fig. 5 below.

Fig. 5



So C_1 and C_2 are the centers of two particular circles, which have the same radii.

Fig. 6



Then, we may want to begin with finding the two centers.
How do we find the two centers, though?

We can use the idea of the center of a circle. What's the idea?

The idea is that the center is the same distance away from every point in the circle, and that the same distance is the radius.

The two centers can be thus, found by the distance formula.

So assuming for instance, the two points are **(1, 2)** and **(3, 4)**, and the radius is **5**, we get a system of two equations for $\boldsymbol{a}$ and $\boldsymbol{b}$ as follows.

One is $(a-1)^2 + (b-2)^2 = 5^2$, and the other is $(a-3)^2 + (b-4)^2 = 5^2$.

And we can get a simpler version of the system using a factorization.
We have a factorization identity, $u^2 - v^2 = (u+v)(u-v)$. So we get

$$(a-1)^2 + (b-2)^2 = (a-3)^2 + (b-4)^2 \Rightarrow (a-1)^2 - (a-3)^2 = (b-4)^2 - (b-2)^2$$

$$\Rightarrow \{(a-1) + (a-3)\}\{(a-1) - (a-3)\} = \{(b-4) + (b-2)\}\{(b-4) - (b-2)\}$$

$$\Rightarrow (2a-4)2 = (2b-6)(-2) \Rightarrow 4a - 8 = -4b + 12 \Rightarrow 4a + 4b = 20 \Rightarrow a + b = 5.$$

What about the other of the two equations in the system?

We can use either of $(a-1)^2 + (b-2)^2 = 5^2$ and $(a-3)^2 + (b-4)^2 = 5^2$.

Choosing for instance, the first of the two above, we get a simpler system where $a + b = 5$, and $(a-1)^2 + (b-2)^2 = 5^2$.

Then, to begin with, we get $a + b = 5 \Rightarrow b = 5 - a$.

So we get $(a-1)^2 + (b-2)^2 = 5^2 \Rightarrow (a-1)^2 + (5-a-2)^2 = 5^2$.

Meanwhile, $(a-1)^2 + (5-a-2)^2 = (a-1)^2 + (3-a)^2 = (a-1)^2 + (a-3)^2$. How?

We can get $(x-c)^2 = (-c+x)^2 = \{(-1)(c-x)\}^2 = (-1)^2(c-x)^2 = 1(c-x)^2 = (c-x)^2$.

In other words, $(x - c)^2 = \{-(x - c)\}^2 = (-x + c)^2 = (c - x)^2$. For instance, $7^2 = (-7)^2 = 7^2$.

So next, we get $(a - 1)^2 + (a - 3)^2 = a^2 - 2a + 1 + a^2 - 6a + 9 = 2a^2 - 8a + 10$.

Since we began with $(a - 1)^2 + (b - 2)^2 = 5^2$, we now get $2a^2 - 8a + 10 = 5^2$.

Thus, we get $2a^2 - 8a + 10 = 5^2 \Rightarrow 2a^2 - 8a - 15 = 0$, which cannot be factored in the integer level. So?

So using the quadratic formula, we can get the values of a.
Don't remember the formula?

Assuming $ax^2 + bx + c = 0$, we get $x = \dfrac{-b \pm \sqrt{b^2 - 4ac}}{2a}$, which is the formula.
So using the quadratic formula, we get

$$a = \frac{8 \pm \sqrt{64 - 4 \cdot 2 \cdot (-15)}}{2 \cdot 2} = \frac{8 \pm \sqrt{184}}{4} = \frac{8 \pm \sqrt{4 \cdot 46}}{4} = \frac{8 \pm 2\sqrt{46}}{4} = \frac{4 \pm \sqrt{46}}{2}.$$

Such a long calculation, isn't it?
We have a simpler version, which is $x = \dfrac{-k \pm \sqrt{k^2 - ac}}{a}$, where $b = 2k$ in the formula.

So using the simpler version, we get $a = \dfrac{4 \pm \sqrt{16 - 2 \cdot (-15)}}{2} = \dfrac{4 \pm \sqrt{46}}{2}$.

What if we can't remember the formula at all at the exam?

Then, put the equation into the complete square. How?

We have $2a^2 - 8a - 15 = 0$. So we get

$$2a^2 - 8a - 15 = 2(a^2 - 4a) - 15 = 2(a^2 - 4a + 4 - 4) - 15 = 2\{(a-2)^2 - 4\} - 15$$

$$= 2(a-2)^2 - 8 - 15 = 2(a-2)^2 - 23 = 0$$

$\Rightarrow (a-2)^2 = \frac{23}{2} \Rightarrow a - 2 = \pm\sqrt{\frac{23}{2}} \Rightarrow a = 2 \pm \sqrt{\frac{23}{2}}$ How do we get $2 \pm \sqrt{\frac{23}{2}} = \frac{4 \pm \sqrt{46}}{2}$?

We can get $2 \pm \sqrt{\frac{23}{2}} = \frac{4}{2} \pm \sqrt{\frac{46}{4}} = \frac{4}{2} \pm \frac{\sqrt{46}}{2} = \frac{4 \pm \sqrt{46}}{2}$.

Next, we have $b = 5 - a$. So we get $b = 5 - a = 5 - \frac{4 \pm \sqrt{46}}{2} = \frac{6 \mp \sqrt{46}}{2}$. Why $\mp$?

Doing it separately,

when $a = \frac{4+\sqrt{46}}{2}$, we get $b = 5 - a = 5 - \frac{4+\sqrt{46}}{2} = \frac{6-\sqrt{46}}{2}$,

and when $a = \frac{4-\sqrt{46}}{2}$, we get $b = 5 - a = 5 - \frac{4-\sqrt{46}}{2} = \frac{6+\sqrt{46}}{2}$.

Now, putting threads together, we get $a = \frac{4 \pm \sqrt{46}}{2}$, and $b = \frac{6 \mp \sqrt{46}}{2}$.

So one center is $(\frac{4+\sqrt{46}}{2}, \frac{6-\sqrt{46}}{2})$, and the other is $(\frac{4-\sqrt{46}}{2}, \frac{6+\sqrt{46}}{2})$.

For reference, the centers are approximately **(5.39, -0.39)** and **(-1.39, 6.39)**.

Next, putting each center into the standard form directly, and setting

$(a_1, b_1) = (\frac{4+\sqrt{46}}{2}, \frac{6-\sqrt{46}}{2})$, and $(a_2, b_2) = (\frac{4-\sqrt{46}}{2}, \frac{6+\sqrt{46}}{2})$, we get

$(x - a_1)^2 + (y - b_1)^2 = 5^2$, and $(x - a_2)^2 + (y - b_2)^2 = 5^2$, since the radius is 5.

Now, what if we are given information not enough to determine a particular circle?

For instance, we are given only two points in the particular circle, and of course, neither of the two is the center. And for another instance, we are given the center only.

Then, we'll have to find a group of circles. That is, the group of circles is the solution. What do we mean by a group of circles, though?

Suppose we are given the center only, and for instance, the center is **(1, -2)**.
Then, using the standard form $(x-a)^2 + (y-b)^2 = c^2$, we can get $(x-1)^2 + (y+2)^2 = c^2$.

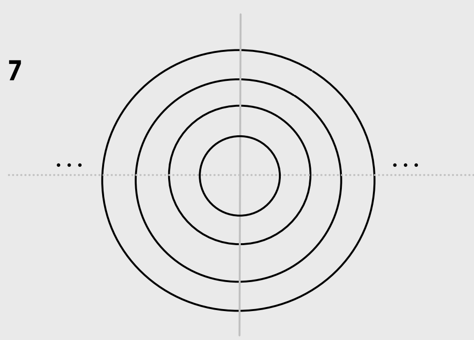
So the solution can be such a group of concentric circles as follows.

$(x-1)^2 + (y-2)^2 = c^2$ where c is a positive integer.

Concentric circles share the same center.

The equation above indicates each of all circles where the centers are **(1, 2)**, but the radii are different from each other. That is, the radii are 0.1, 0.15, 1, 1.9, 2, 3, 4, etc.

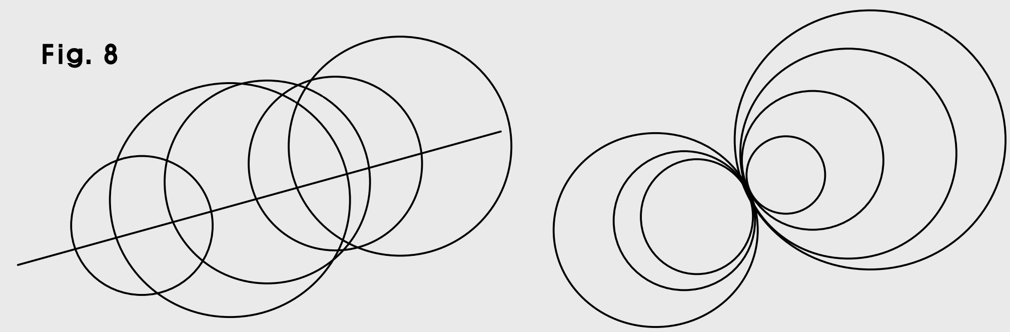
Fig. 7



That's not it, of course.

So for another instance, the solution can also be a group of circles where the centers are in a line, or a group of circles tangent to each other.

Fig. 8



Doing a problem with finding a particular circle, we should be able to find information that can determine the particular circle.
Sometimes, the problem is so straightforward that all necessary information is given to us directly.

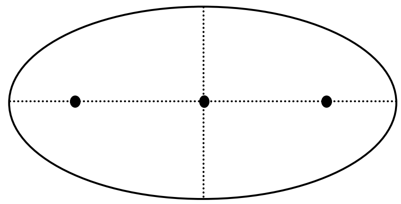
For instance, three points in a particular circle are specified as in the first example above.

Some other times, all the information is given in an implicit manner.
In such cases, we should be able to extract or deduce such information from the problem description. That is, we need to track down the information on the center and radius.

In fact, if the problem is well defined, the problem is the solution. Why?

It all depends on how we put the problem.
Solving a problem, we put it many different ways.
By the way, math begins with definitions. So whatever the problem it may be, we'd better define it properly in the first place. Then, the solution will be with us.

3.0. What is an ellipse?



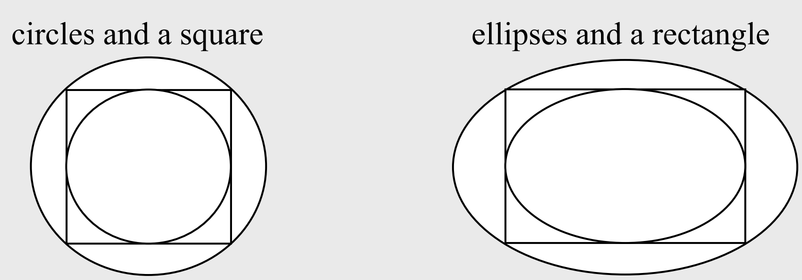
An ellipse is in a plane, and is a set of points, from each of which,
the sum of the two distances to two particular points is constant, that is, the same.
And the two particular points are called the foci of the ellipse.

An ellipse is one of the basic curves called conic sections, often called conics for short. So it is a conic, and we call it an oval, too, since it looks like an egg. Also, it can be called a long circle, too. What do we mean by a long circle?

An ellipse can be quite close to a circle, but can be quite different from a circle, too. So to begin with, comparing an ellipse with a circle, we can say that a circle can be an incircle or a circumcircle of a ***square***, and that an ellipse can be inscribed in or can be circumscribed about a ***rectangle***.

As in Fig. 0, an incircle of a square is tangent to all the four sides of the square, and a circumcircle of a square includes all the four vertices of the square. So an ellipse inscribed in a rectangle is tangent to all the four sides of the rectangle, and an ellipse circumscribed about a rectangle is passing through all the four vertices of the rectangle.

Fig. 0

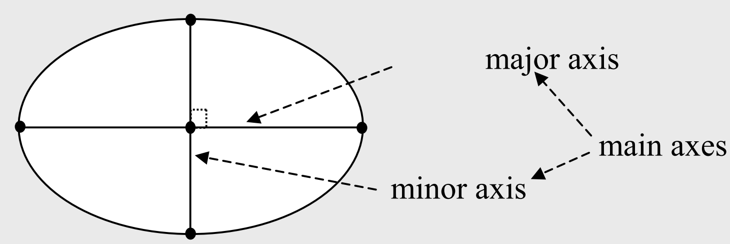


Next, a circle is a simply closed curved-line segment that is symmetric and has infinitely many axes of symmetry, each of which is a diameter.

An ellipse, too, is a simply closed curved-line segment that is symmetric.

It has however, only ***two axes of symmetry***, one is a ***major axis***, and the other is a ***minor axis***, called a conjugate axis, too. And we call the two axes of symmetry ***main axes***, too, and can put the main axes in an ellipse the way as follows.

Fig. 1



As shown above, the two main axes are perpendicular to each other.

And we can notice that the major axis connects two points the farthest away from the center of an ellipse, and that the minor axis connects two points the closest to the center.

In other words, the major axis connects two antipodal points the farthest away from each other, and the minor axis connects two antipodal points the closest to each other.

And of course, the two main axes meet at the center.
How then can we get an ellipse?

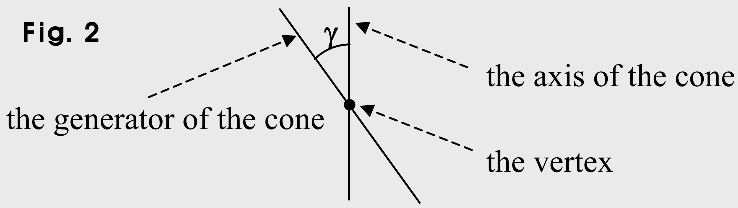
If cutting a right cone with a plane, we can get a cross section called an ellipse. So it's called a conic section, just called a conic for short. What is a right cone though?

If a cone is a right cone, the line connecting the vertex of the right cone and the center of the base is perpendicular to the base, and thus, makes a right angle (90^o) with the base.

And the line stated above is called the axis of the cone. And in the study of conics, such a right cone is made or generated the way as follows.

First, take a line as the axis of the cone to be made.
Next, make another line meet the axis of the cone at a point at an angle γ, which is an acute angle, that is, **$0 < \gamma < 90^o$**.

Fig. 2



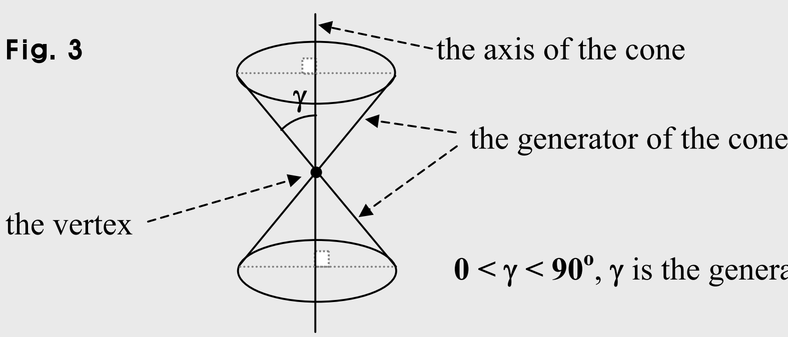
Then, we call the other line the *generator of the cone*, and the point where the other line, that is, the generator of the cone meets the axis of the cone is the vertex of the cone.

So the vertex of the cone is the point where the generator meets the axis of the cone.

And in the figure above, the angle γ is called the *generator angle*.

Then, rotating the generator around the axis fixing the generator at the vertex, we get a double-cone as in Fig. 3.

Fig. 3



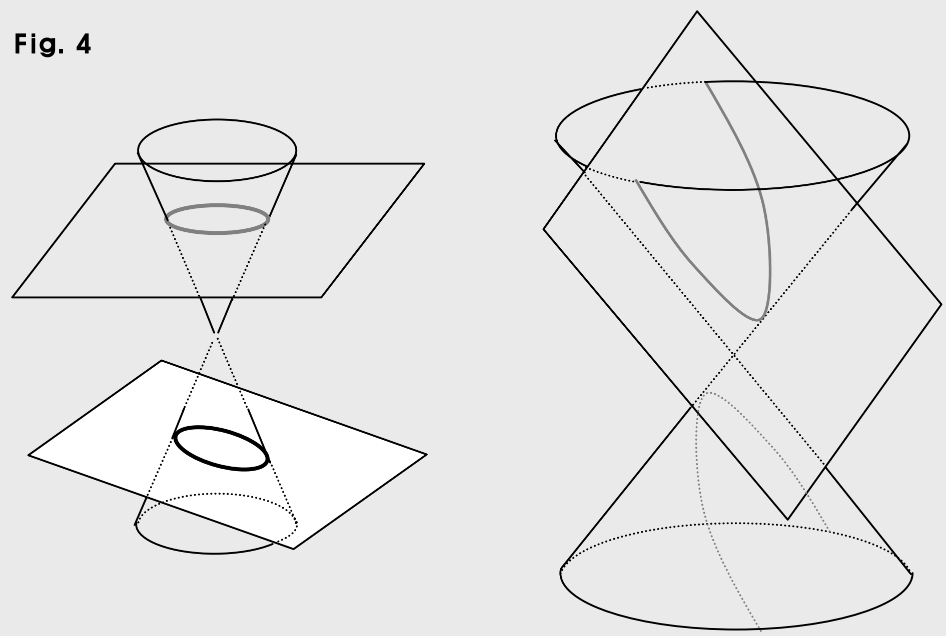
$0 < \gamma < 90^o$, γ is the generator angle

And note that the generator and the axis are lines, and lines have infinite lengths, so the lengths of the cones are infinite, too. So we cannot show them all. Showing thus, such a cone or a curve in math, we just show some part of it.

How then can we get an ellipse cutting the cone with a plane?

As in Fig. 4, the cross section in black is a curve closed, and is an ellipse. So if α is the angle between the plane and the axis of the cone, and $\gamma < \alpha < 90^\circ$, and of course, if the plane does not include the vertex, the cross section is an ellipse.

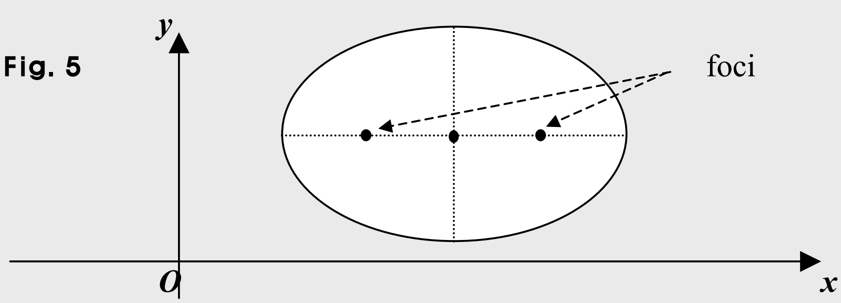
Fig. 4



If however, $\alpha = \gamma$, the cross section is a parabola, and if $\alpha = 90^\circ$, the cross section is a circle. And we can explain an ellipse the way as follows.

An ellipse is a collection of points as in the case of a circle, line, or parabola. Suppose in the ***x-y*** plane, as in Fig. 5, we designate two particular points called the foci of an ellipse. Then, the ellipse is a set of points, from each of which, the sum of the two distances to the two foci is constant, that is, the same.

Fig. 5

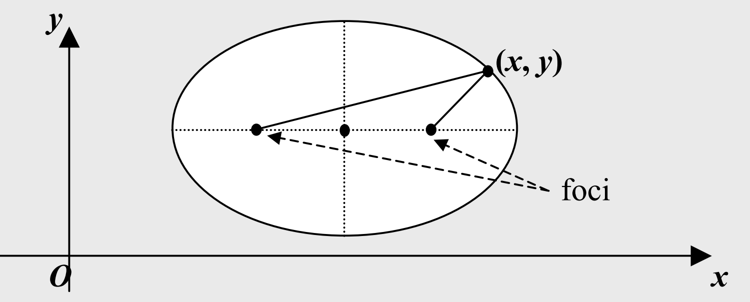


And we can also explain an ellipse the way as follows.

Suppose that a point is moving along a curve in the ***x*-*y*** plane, and that the sum of the two distances from the moving point to two particular points is constant.
Then, the two particular points are called the foci, and the curve is an ellipse.

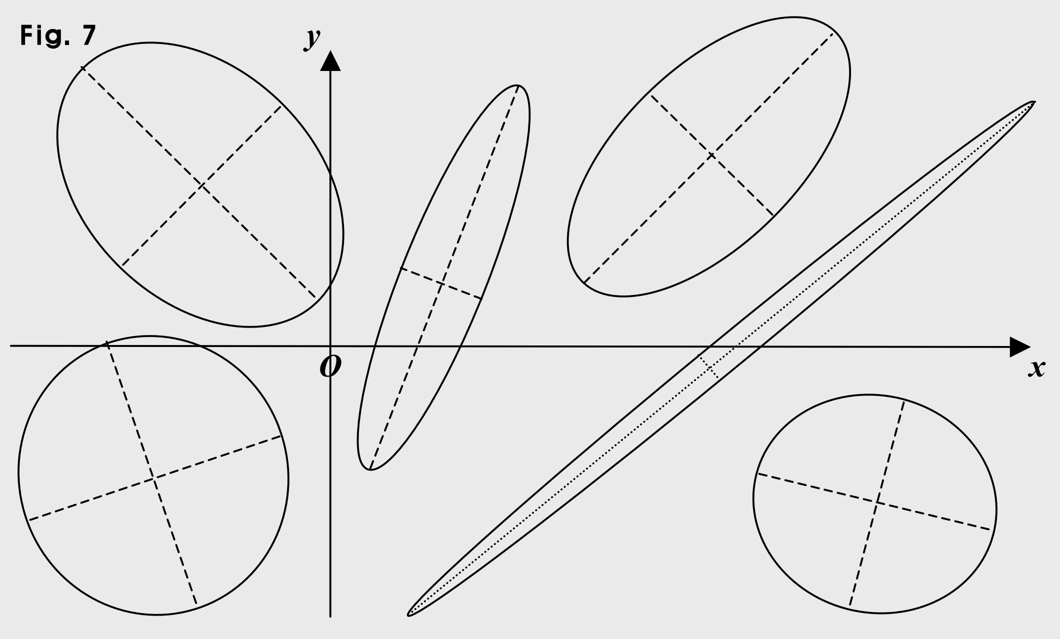
So if a point **(x, y)** is in an ellipse, no matter where the point **(x, y)** may be, the sum of the two distances from the point **(x, y)** to the two points called the foci is constant.

Fig. 6



And we can notice that one of the two main axes is perpendicular to the ***x***-axis. In the figure above, the minor axis is perpendicular to the ***y***-axis. And in that case, of course, the major axis is parallel to the ***x***-axis.
It can be the case, of course, no main axis is perpendicular to the ***x***-axis. Such an ellipse can be said to be skewed or tilted, and some examples can be as follows.

Fig. 7

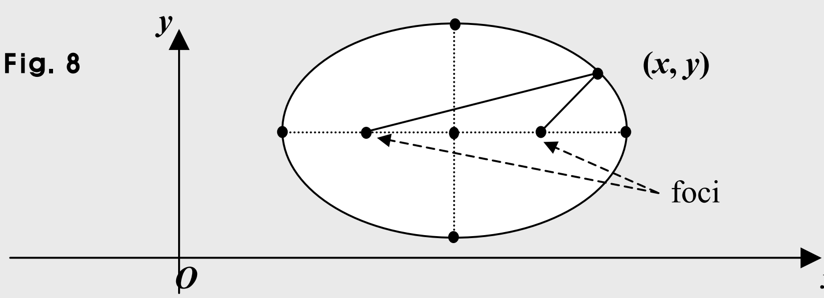


And some ellipses are quite close to circles, and some others look like bars.

Usually though, in high school math or basic courses in college math, if we use an ellipse, one of the two main axes is perpendicular to a coordinate axis. So such an ellipse can be called a perpendicular ellipse.

And in fact, in this book, such an ellipse is said to be perpendicular, and thus, is called a perpendicular ellipse. (Note however, in other books, it may not be called a perpendicular ellipse, and can be called differently.)
Anyway, putting a perpendicular ellipse in the ***x-y*** plane, we can get this:

Fig. 8



And looking at the figure above, we can describe the two main axes the way below, too.

The longer of the two is called the major axis. And the major axis connects two points, from each of which, the difference between the distances to the foci is the largest.

The shorter main axis is called the minor axis. And the minor axis connects two points, from each of which, the distances to the foci are the same.

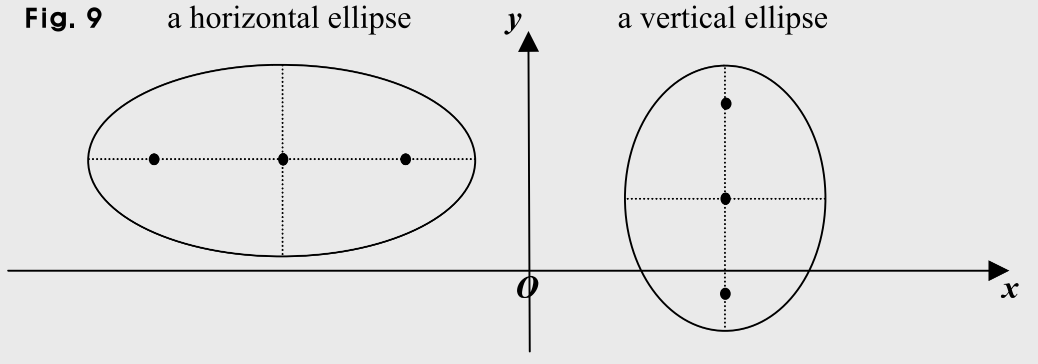
What then about the two points called the foci?

The two foci are in the major axis. So the major axis passes through the foci.
Each focus is an equal distance away from the center of the ellipse.
And we call the equal distance a focal distance.

So a focal distance is the distance from a focus to the center of the ellipse.

And in this book, we have two kinds in ellipses perpendicular.
One is horizontal, and the other is vertical.

Fig. 9



So in an ellipse horizontal, the major axis is horizontal, that is, parallel to the $\boldsymbol{x}$-axis, and in an ellipse vertical, the major axis is vertical, that is, perpendicular to the $\boldsymbol{x}$-axis.
In short, if the major is horizontal, the ellipse is horizontal, and if the major is vertical, the ellipse is vertical.

(Note that though, it is the case in this book, and it may not be the case in other books. That is to say that in other books, ellipses may not be classified the way above, and thus, may not be said to be horizontal or vertical.)

Either way, the foci are in the major axis, and the focal distance is the distance from a focus to the center of an ellipse.

And usually, a ***half major*** axis is called a ***semi major*** axis or a ***major radius***, so a half minor axis is called a semi minor axis or a minor radius.

Where can we find an ellipse though?

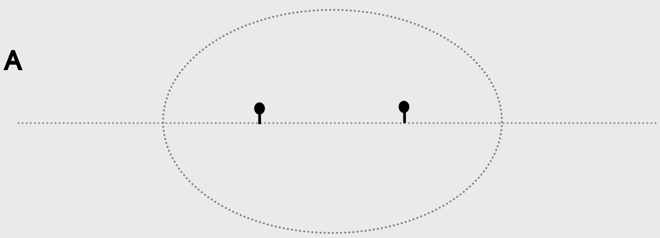
We know that the earth is orbiting around the sun, and the orbit is an ellipse.
Many orbits are ellipses. Probably, every orbit is an ellipse. And also, if an object looks circular, it is quite likely to be an ellipse. How can we actually make an ellipse though?

We know the fact that an ellipse is a set of points, from each of which, the sum of the two distances to the foci is constant.   So?

So using the fact above, we can construct an ellipse the way as follows.

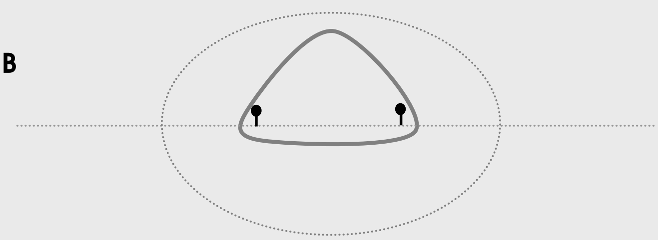
First, set the two foci by fastening, for instance, two nails or thumb tacks to a panel.

Fig. A



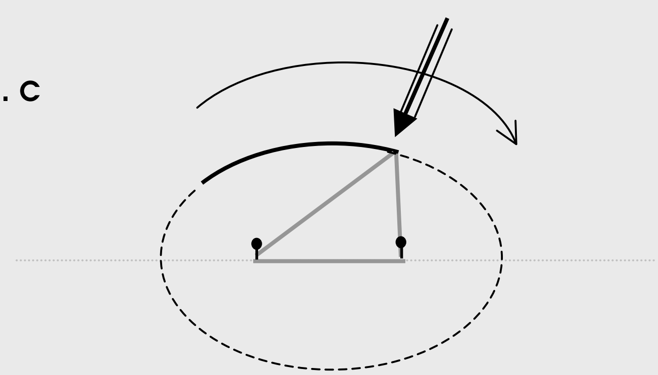
Next, make a loop using a string, and put the loop on the panel so that it wraps around the nails.

Fig. B



Then, while holding the loop taught with a pencil tip, turn the pencil around the nails the way we do drawing a circle. After a complete turn, we get an ellipse.

Fig. C



So we can see that if a point is in the ellipse, the sum of the two distances from the point to the foci is constant, that is, the same, because the length of the loop is constant.
How then can we explain or describe a particular ellipse?

Normally, we put an ellipse in the ***x-y*** plain, and the ellipse is perpendicular, so one of the two main axes is perpendicular to the ***x***-axis. And we can describe such an ellipse specifying the information as follows.

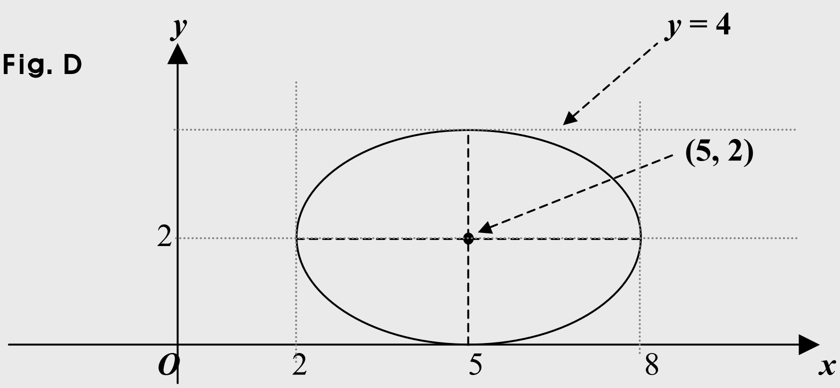
The center of the ellipse and the lengths of the two main axes, along with the orientation, which indicates if the ellipse is horizontal or vertical

If the major axis is horizontal, thus, parallel to the ***x***-axis, the ellipse is horizontal, and if the major axis is vertical, thus, perpendicular to the ***x***-axis, the ellipse is vertical.

(Note that though, it is the case in this book, and it may not be the case in other books. That is to say that in other books, ellipses may not be classified the way above, and thus, may not be said to be horizontal or vertical.)

Anyway, for instance, if an ellipse is horizontal, and is centered at (5, 2), and its main axes are 6 and 4, the major axis is parallel to the ***x***-axis, and the ellipse is as follows.

Fig. D



In math though, making an ellipse, we define it. So defining an ellipse particular, we make the particular ellipse. How then can we define a particular ellipse?

As in the cases of other conics as a circle, we can define an ellipse using an equation that indicates the ellipse. So we can define the particular ellipse producing the equation of it. For instance, assuming ***E*** is the ellipse described above, we can define ***E*** the way below.

$$\frac{(x-5)^2}{3^2}+\frac{(y-2)^2}{2^2}=1$$

And the equation above is in the standard form, and is a standard equation of an ellipse. Putting the standard equation above in the general form, we just expand (or simplify) it.

Then, we get $\mathbf{4x^2 + 9y^2 - 40x - 36y + 100 = 0}$, which is a general equation of an ellipse.

And in general, putting an ellipse in the standard form, we can put it the way as follows.

$$\frac{(x-u)^2}{a^2} + \frac{(y-v)^2}{b^2} = 1, \text{ where } a \neq b, \text{ and } a \text{ and } b > 0$$

What do we mean by though, a and b?

First, if $a > b$, $2a$ is the length of the major axis. And if we just call the length the major, the major is $2a$. Then, $2b$ is the minor, and the ellipse is horizontal.
And in that case, we call a the semi major or the major radius, and b is called the semi minor or the minor radius.

Next, if $a < b$, the major is $2b$, the minor is $2a$, and the ellipse is vertical.
And in that case, we call b the semi major or the major radius, and a is called the semi minor or the minor radius.

So just describing a particular ellipse, we can describe it specifying the main axes (or the two semis or the two radii).

If however, we define a particular ellipse placed in the x-y plane, we want to specify the center and the orientation, together with the two main axes (or the semis or the radii).
What then is the center?

If an ellipse is $\frac{(x-u)^2}{a^2} + \frac{(y-v)^2}{b^2} = 1$, the center of the ellipse is a point (u, v).

Note that though, the standard equation above can indicate an ellipse perpendicular only.
Is there any equation then, that can indicate an ellipse perpendicular or not?

Putting an ellipse in such an equation, we can put it the way as follows.

$$\boldsymbol{ax^2 + by^2 + cxy + ux + vy + w = 0}, \text{ where } \boldsymbol{4ab > c^2}$$

Note that $\boldsymbol{a}$ and $\boldsymbol{b}$ can be the same, and that $\boldsymbol{4ab > c^2}$ can mean this, too: $\boldsymbol{ab > 0}$, which is saying that $\boldsymbol{a}$ and $\boldsymbol{b}$ have the same sign. And of course, $\boldsymbol{a}$, $\boldsymbol{b}$, $\boldsymbol{c}$, $\boldsymbol{u}$, $\boldsymbol{v}$, and $\boldsymbol{w}$ are constant. Note however, for some values of the constants, the equation does not indicate an ellipse, even if $\boldsymbol{4ab > c^2}$.

What then about an equation that is in the general form and indicates an ellipse perpendicular only?

Putting an ellipse in such an equation, we can put it the way as follows.

$$\boldsymbol{ax^2 + by^2 + ux + vy + w = 0}, \text{ where } \boldsymbol{a \neq b}, \text{ and } \boldsymbol{ab > 0}$$

So the equation above indicates an ellipse with a main axis perpendicular to the $\boldsymbol{x}$-axis. Notice that it does not have the $\boldsymbol{xy}$-term, that is, $\boldsymbol{cxy}$ in the equation shown earlier.
Note however, for some values of the constants, the equation does not indicate an ellipse, even if $\boldsymbol{a \neq b}$, and $\boldsymbol{ab > 0}$.

And we can give a name to an equation, too. For instance, assuming $\boldsymbol{E}$ is the equation of an ellipse perpendicular above, we can put $\boldsymbol{E}$ the way as follows.

$$\boldsymbol{E(x, y) = ax^2 + by^2 + ux + vy + w = 0}, \text{ where } \boldsymbol{a \neq b}, \text{ and } \boldsymbol{ab > 0}$$

What then about the standard equation?

Assuming $\boldsymbol{S}$ is the equation $\boldsymbol{\dfrac{(x-u)^2}{a^2} + \dfrac{(y-v)^2}{b^2} = 1}$, we can put $\boldsymbol{S}$ the way as follows

$$S(x,y) = \frac{(x-u)^2}{a^2} + \frac{(y-v)^2}{b^2} - 1 = 0, \text{ where } a \neq b, \text{ and } a \text{ and } b > 0.$$

What if $a = b$ though, in the equation above?

Then, the equation indicates not an ellipse but a circle. What circle then is it?

It is the circle where the radius is a (or b), and the center is at (u, v).

Setting $b = a$, we get $\frac{(x-u)^2}{a^2} + \frac{(y-v)^2}{b^2} - 1 = 0 \Rightarrow \frac{(x-u)^2}{a^2} + \frac{(y-v)^2}{a^2} = 1$

$\Rightarrow (x - u)^2 + (y - v)^2 = a^2$, which indicates a circle of radius a centered at (u, v).

And we know it is ***not*** the case the equation of the circle $(x - u)^2 + (y - v)^2 = a^2$ is defined for x real and y real. ('x real' means every real number can be x.)

It is in fact, defined for $-a + u \leq x \leq a + u$, and $-a + v \leq y \leq a + v$.

That is, the domain is $-a + u \leq x \leq a + u$, and the range is $-a + v \leq y \leq a + v$.

So we can notice that it is ***not*** the case either, the standard equation of an ellipse above is defined for x real and y real. What then are the domain and the range?

The domain is $-a + u \leq x \leq a + u$, and the range is $-b + v \leq y \leq b + v$.

So putting more specifically the standard equation of an ellipse, we can put it this way:

$$\frac{(x-u)^2}{a^2} + \frac{(y-v)^2}{b^2} = 1, \text{ where } a \neq b,\ a\ \&\ b > 0,\ -a + u \leq x \leq a + u, \text{ and } -b + v \leq y \leq b + v.$$

Next, we know an ellipse can be very close to a circle, and can be very flat. So is there any way we can specify how round or flat an ellipse is?

An ellipse has a special number called an ***eccentricity***, which is a ratio of a focal distance to a major radius, that is, the focal distance over the major radius.

So of an ellipse, the eccentricity specifies the degree, to which the ellipse is out of round. Thus, it can tell us how flat or round an ellipse is.

An eccentricity is denoted by e, and we have $\mathbf{0 < e < 1}$ for an ellipse.
And if the eccentricity e is 0, it is a circle.

So the bigger the eccentricity, the flatter the ellipse.

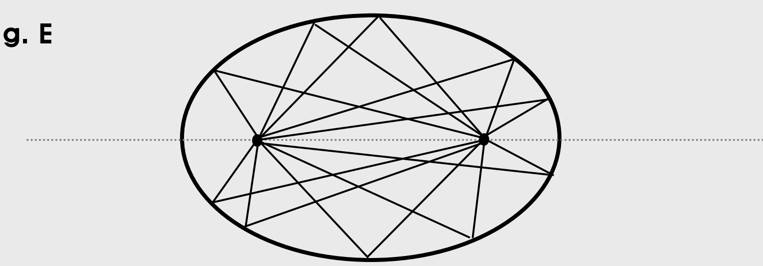
If therefore, e is close to 1, the ellipse looks like a bar.

And an ellipse has a physical property as follows.

If for instance, a light beam comes from one focus, and reflects off the ellipse, it goes to the other focus. So all the light beams coming from one focus and reflecting off the ellipse gather at the other focus.

So we can use such a property constructing a reflecting telescope.

Fig. E



How then can we get the foci and the eccentricity?

Suppose an ellipse is $\dfrac{(x-u)^2}{a^2}+\dfrac{(y-v)^2}{b^2}=1$.

Then, if $a > b > 0$, the foci are $(u - c, v)$ and $(u + c, v)$, where c is the focal distance, and we have $c^2 = a^2 - b^2$.

Then, the eccentricity $\boldsymbol{e} = \dfrac{\boldsymbol{c}}{\boldsymbol{a}}$. And the ellipse is horizontal.

If however, $\boldsymbol{b} > \boldsymbol{a} > \mathbf{0}$, the foci are $(\boldsymbol{u}, \boldsymbol{v} + \boldsymbol{c})$ and $(\boldsymbol{u}, \boldsymbol{v} - \boldsymbol{c})$, where $\boldsymbol{c}$ is the focal distance, and we have $\boldsymbol{c}^2 = \boldsymbol{b}^2 - \boldsymbol{a}^2$.

Then, the eccentricity $\boldsymbol{e} = \dfrac{\boldsymbol{c}}{\boldsymbol{b}}$. And the ellipse is vertical.

3.1. Equations for Ellipses 1

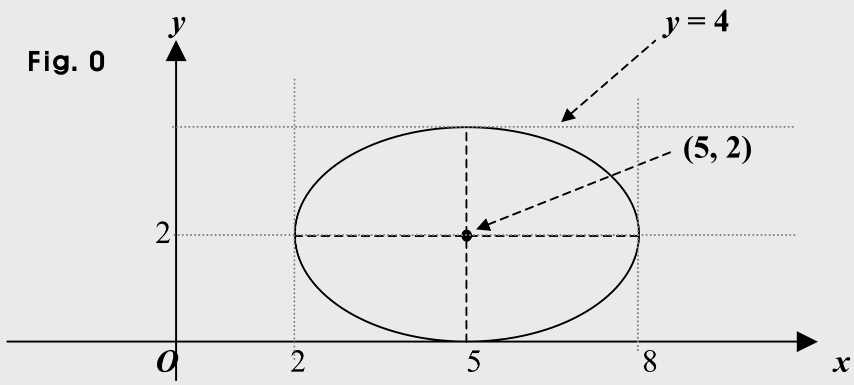
To begin with, putting an ellipse in an equation, we can get

$$\frac{(x-5)^2}{3^2}+\frac{(y-2)^2}{2^2}=1$$

What ellipse then, is it?

It is an ellipse horizontal, the center is at (5, 2), and its main axes are 6 and 4. So the major axis is parallel to the ***x***-axis, and is 6, and the ellipse is as below.

Fig. 0

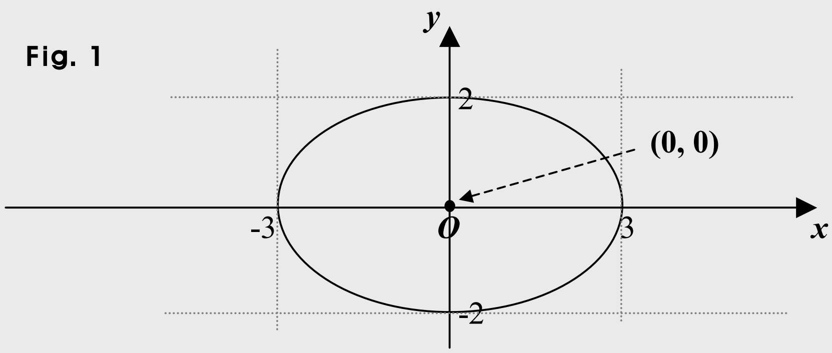


What ellipse then is the ellipse as follows?

$$\frac{x^2}{3^2}+\frac{y^2}{2^2}=1$$

The ellipse is horizontal, too, and its main axes are 6 and 4, also, but the center is different, and is at (0, 0), that is, at the origin.

Fig. 1



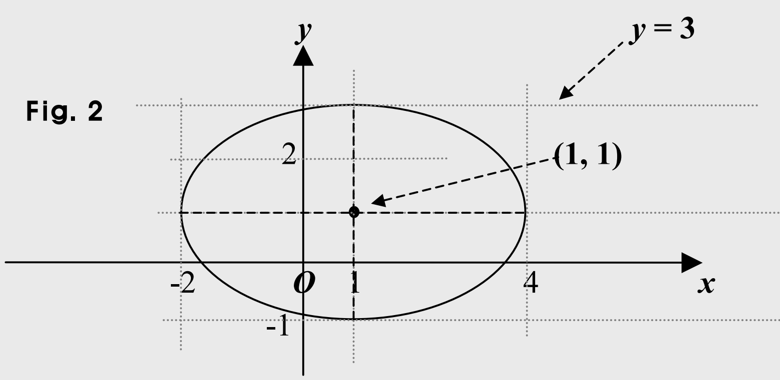
What if the center is at (1, 1), and the other information is the same?

Then, the ellipse will be as follows.

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

And we can put the ellipse above in a graph the way as follows.

Fig. 2



How then can we get such equations as above?

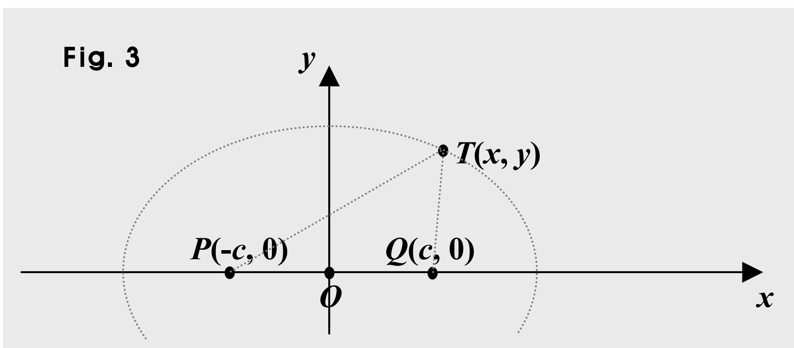
We know the fact that an ellipse is a set of points, from each of which, the sum of the two distances to two points called the foci is constant. So?

So using the fact above, we can get the equation of an ellipse.

Thus, to begin with, suppose $c > 0$, two points $P(-c, 0)$ and $Q(c, 0)$ are the two foci of an ellipse called E, and a point $T(x, y)$ is an arbitrary point in the ellipse E, that is, a random point representing all the points in E.

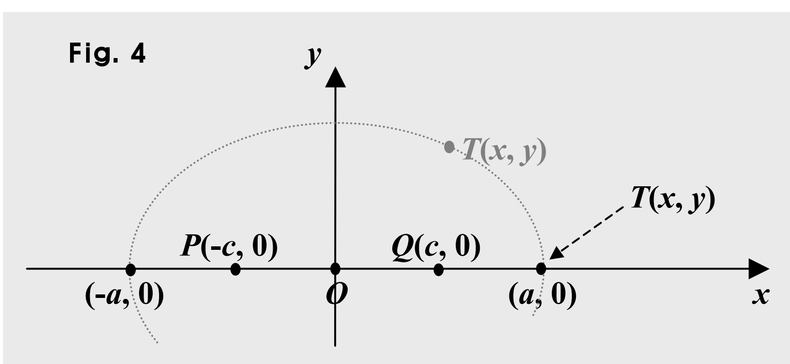
Then, the center is the origin, that is, $(0, 0)$, and c is the focal distance, because the center is the midpoint between the foci. So we can put the three points P, Q, and T in the x-y plane the way as follows:

Fig. 3



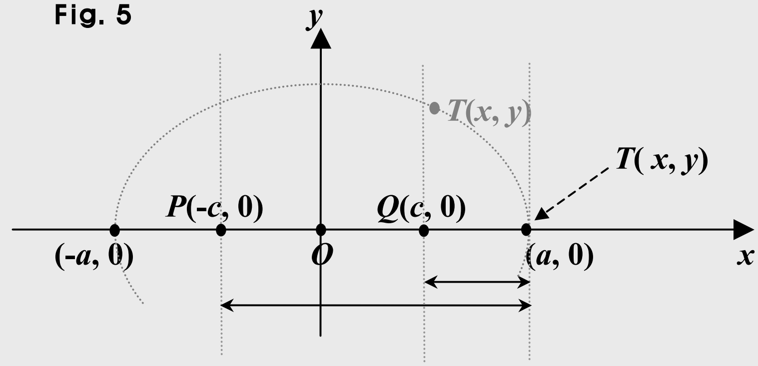
Suppose next, the point T is moving along the ellipse E, and is now at $(a, 0)$.
Then, we get this:

Fig. 4



Then, the sum of the two distances from $T(x, y)$ to the foci is $2a$. How?

Fig. 5



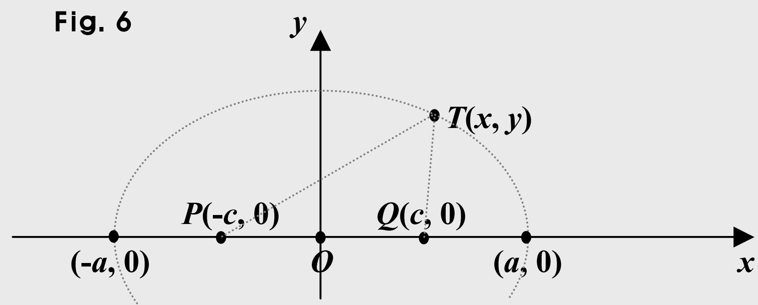
We can see that the length of TP is $a - (-c) = a + c$.
And the length of TQ is $a - c$.
So we get $TP + TQ = (a + c) + (a - c) = 2a$. So what?

We know the fact that from any point in an ellipse, the sum of the two distances to two points called the foci is constant. So?

So no matter where the point ***T*** may be in the ellipse ***E***, the sum of the two distances from the point ***T*** to the two foci is always the same, and is in this case, $2a$.

Using thus, the fact above, we can get the equation of ***E***. How then can we use it?

Fig. 6



We are going to get the sum of the two distances from $T(x, y)$ to the two foci, and then, set the sum equal to the sum we found above, which is $2a$, because the sum is constant, that is, the same no matter where the point ***T*** may be in the ellipse ***E***.

How then can we get the two distances from $T(x, y)$ to the two foci?

We can use the distance formula, often called Pythagorean Theorem.
What then do we get?

We get an equation expressed in terms of the coordinates of the arbitrary point $T(x, y)$ in the curve called the ellipse E. When then does the equation do?

The equation explains how x and y are related to each other. And we call such an equation a connective equation. So the connective equation connects x and y at $T(x, y)$ in the curve called the ellipse E. In other words, the equation explains every point in the ellipse E, so it defines the ellipse E. What equation then is it?

The equation is called the equation of the ellipse E.
So let's now get the equation. How then can we apply the distance formula?

Assuming d is the distance between two points, Δx is the difference in x-coordinates, and Δy is the difference in y-coordinates, we can put the formula the way as follows.

$$d^2 = (\Delta x)^2 + (\Delta y)^2$$

So using the distance formula, and assuming $TP = p$, and $TQ = q$, we can get

$p^2 = \{x - (-c)\}^2 + (y - 0)^2 = (x + c)^2 + y^2$, and $q^2 = (x - c)^2 + (y - 0)^2 = (x - c)^2 + y^2$.

So we get $TP = \sqrt{(x+c)^2 + y^2}$, and $TQ = \sqrt{(x-c)^2 + y^2}$.

Thus, we get $TP + TQ = \sqrt{(x+c)^2 + y^2} + \sqrt{(x-c)^2 + y^2} = 2a$. Then, we get

$$\sqrt{(x+c)^2 + y^2} = 2a - \sqrt{(x-c)^2 + y^2} \Rightarrow (x+c)^2 + y^2 = (2a - \sqrt{(x-c)^2 + y^2})^2$$

$$\Rightarrow (x+c)^2 + y^2 = 4a^2 - 4a\sqrt{(x-c)^2 + y^2} + (x-c)^2 + y^2$$

$$\Rightarrow 4a\sqrt{(x-c)^2 + y^2} = (x-c)^2 + y^2 - (x+c)^2 - y^2 + 4a^2$$

$$\Rightarrow 4a\sqrt{(x-c)^2+y^2} = x^2-2cx+c^2-x^2-2cx-c^2+4a^2 = 4a^2-4cx$$

$$\Rightarrow a\sqrt{(x-c)^2+y^2} = a^2-cx \Rightarrow a^2\{(x-c)^2+y^2\} = (a^2-cx)^2 = a^4-2a^2cx+c^2x^2$$

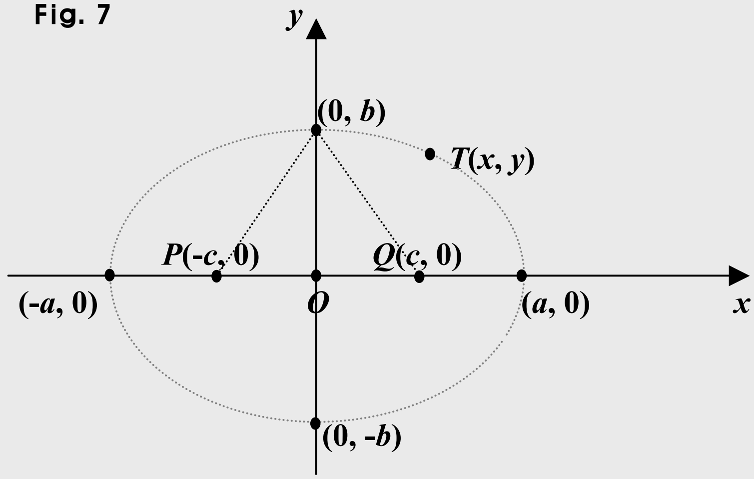
$$\Rightarrow a^2(x^2-2cx+c^2+y^2) = a^2x^2-2a^2cx+a^2c^2+a^2y^2 = a^4-2a^2cx+c^2x^2$$

$$\Rightarrow a^2x^2-c^2x^2+a^2c^2+a^2y^2 = a^4 \Rightarrow (a^2-c^2)x^2+a^2y^2 = a^4-a^2c^2 = a^2(a^2-c^2)$$

$$\Rightarrow x^2+\frac{a^2y^2}{a^2-c^2} = a^2 \Rightarrow \frac{x^2}{a^2}+\frac{y^2}{a^2-c^2} = 1.$$ What then about $a^2 - c^2$?

It is constant, so assuming b is constant, we can set $b^2 = a^2 - c^2$. And assuming $b > 0$, we can put two points $(0, b)$ and $(0, -b)$ the way as follows.

Fig. 7



How though?

Assuming now, that the point T is at $(0, b)$, we can see a triangle isosceles, and the triangle is ΔTPQ.

And we know that the sum of the two distances from T to the foci is $2a$.

That is, we have $TP + TQ = 2a$. So we get $TP = TQ = a$. So?

Assuming $\boldsymbol{O}$ is the origin, we can see that the triangle $\boldsymbol{TOP}$ is a right triangle, and its hypotenuse $\boldsymbol{TP}$ is $\boldsymbol{a}$, and the side $\boldsymbol{PO}$ is $\boldsymbol{c}$.

So applying the distance formula, we can get $\boldsymbol{a^2 = b^2 + c^2 \Rightarrow b^2 = a^2 - c^2}$.

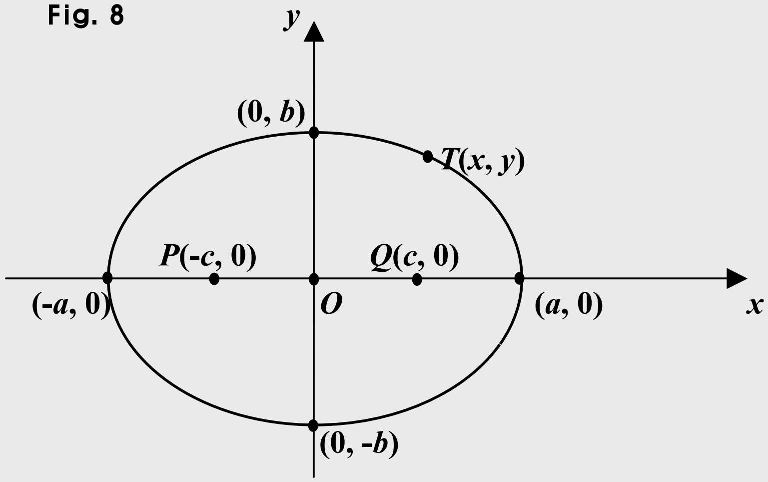
And we now have $\dfrac{x^2}{a^2} + \dfrac{y^2}{a^2 - c^2} = 1$.

So we get $\dfrac{x^2}{a^2} + \dfrac{y^2}{b^2} = 1$, and we know $\boldsymbol{a > b > 0}$, since $\boldsymbol{a^2 = b^2 + c^2 \Rightarrow a > b > 0}$.

Thus, the equation of the ellipse $\boldsymbol{E}$ is $\dfrac{x^2}{a^2} + \dfrac{y^2}{b^2} = 1$, where $\boldsymbol{a > b > 0}$.

And putting the ellipse $\boldsymbol{E}$ in the $\boldsymbol{x}$-$\boldsymbol{y}$ plane, we get this:

Fig. 8



So we can now say that the ellipse $\boldsymbol{E}$ is horizontal and centered at the origin, and that the major axis is $\boldsymbol{2a}$, and the minor axis is $\boldsymbol{2b}$.

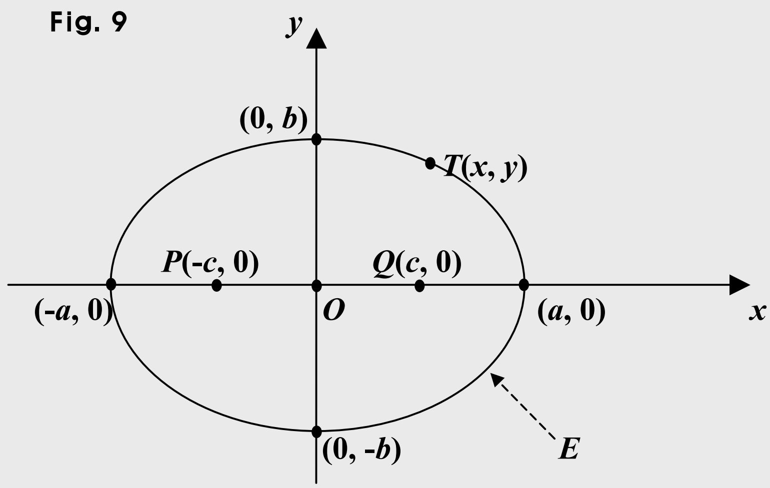
That is, the semi major axis is $\boldsymbol{a}$, and the semi minor is $\boldsymbol{b}$. In other words, the major radius is $\boldsymbol{a}$, and is parallel to the $\boldsymbol{x}$-axis, and the minor radius is $\boldsymbol{b}$.

And putting the ellipse $\boldsymbol{E}$ in its equation, we get $\dfrac{x^2}{a^2} + \dfrac{y^2}{b^2} = 1$, where $\boldsymbol{a > b > 0}$.

What then are the domain and the range?

The domain is $-a \le x \le a$, that is, $|x| \le a$, and the range is $-b \le y \le b$ or $|y| \le b$.

Fig. 9



Suppose now again, that the point ***T*** is moving along the ellipse ***E*** above.
Then, when will the point ***T*** be the farthest away from the center?

When the point ***T*** is in the ***x***-axis, ***T*** is the farthest away from the center.
And we know that the major radius is ***a***, and is parallel to the ***x***-axis, because the ellipse ***E*** is horizontal.
So the major radius is a part of the ***x***-axis, and ***a*** is the distance from ***T*** to the center when ***T*** is the farthest away from the center.

And the ellipse has two points where ***T*** gets located when ***T*** is the farthest away from the center. And the two are **(-*a*, 0)** and **(*a*, 0)**, which are called antipodal points.

Antipodal points are two points, and are in a line passing through the center located between the two points. So they are a pair of points facing each other over the center.

And in the case of an ellipse or a circle, antipodal points are the same distance away from the center. So in that case, the center is the midpoint between the antipodal points.

So the major axis is the line segment connecting the two antipodal points where ***T*** gets positioned when ***T*** is the farthest away from the center, and the two are **(-*a*, 0)** and **(*a*, 0)**.

Thus, the two antipodal points **(-*a*, 0)** and **(*a*, 0)** are the endpoints of the major axis, and the length of it is **2*a***.
What then about the case where the point ***T*** is the closest to the center?

When the point ***T*** is in the ***y***-axis, ***T*** is the closest to the center.

And we know the minor radius is ***b***, and is parallel to the ***y***-axis, that is, perpendicular to the ***x***-axis, because the ellipse ***E*** is horizontal.

So ***b*** is the distance from ***T*** to the center when ***T*** is the closest to the center.

And there are two points where ***T*** gets positioned when ***T*** is the closest to the center.
The two points are antipodal, too. And the two are **(0, *b*)** and **(0, -*b*)**.

So the minor axis is the line segment connecting the two antipodal points where ***T*** gets positioned when ***T*** is the closest to the center, and the two are **(0, *b*)** and **(0, -*b*)**.

Thus, the two antipodal points **(0, *b*)** and **(0, -*b*)** are the endpoints of the minor axis, and the length of it is **2*b***.

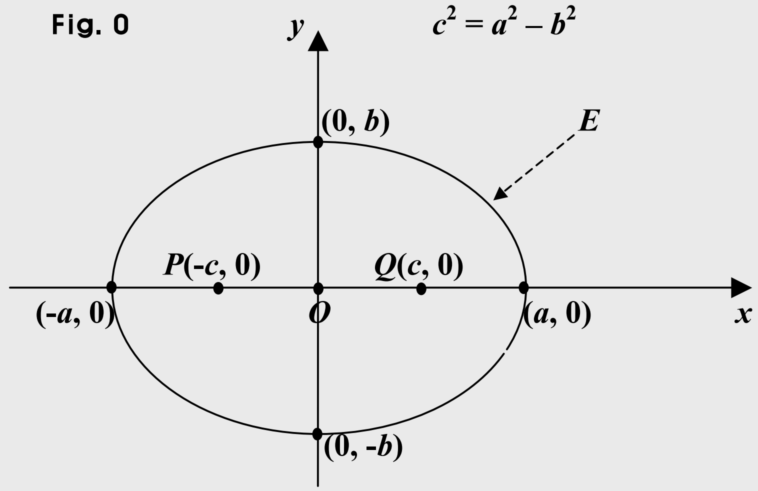
3.2. Equations for Ellipses 2

To begin with, putting an ellipse in an equation, we can get this:

$\frac{x^2}{a^2}+\frac{y^2}{b^2}=\mathbf{1}$, where $\boldsymbol{a > b > 0}$. What ellipse then is it?

It is an ellipse horizontal, where the center is at (0, 0), and its main axes are **2*a*** and **2*b***. So the major axis is parallel to the ***x***-axis, and is **2*a***, and the ellipse is as below.

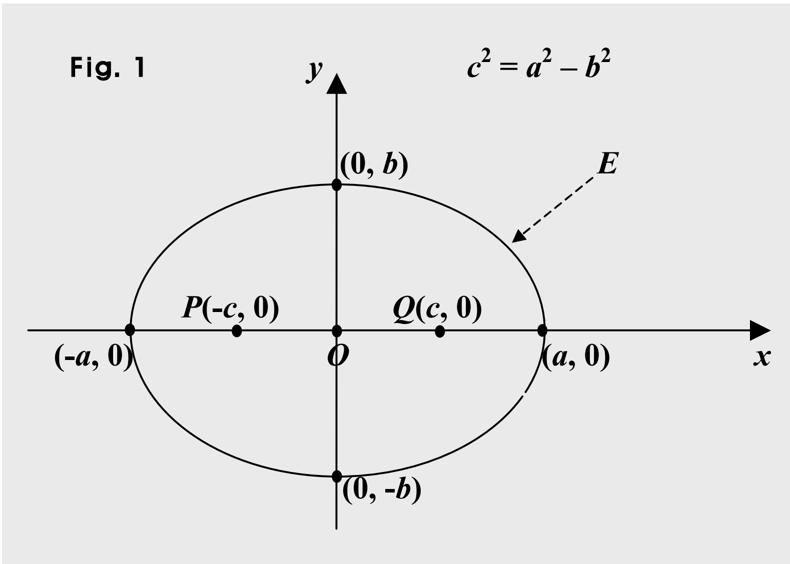
Fig. 0

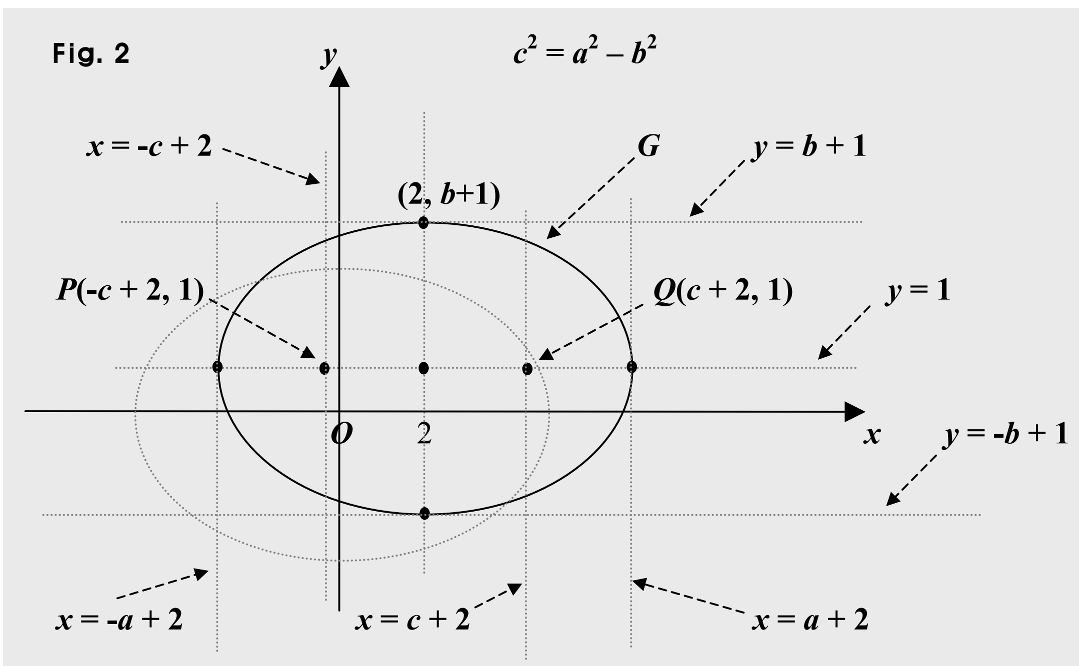


What if the ellipse ***E*** above is centered at a point other than the origin?

For instance, moving the center to (2, 1), we get a new ellipse, so we get a new equation. What then will the new equation be?

Assuming the new ellipse is ***G***, we can get ***G*** translating the ellipse ***E*** in the amount of 2 in the direction of the ***x***-axis, and in the amount of 1 in the direction of the ***y***-axis.





Then, the equation of the ellipse $\boldsymbol{G}$ is $\dfrac{(x-2)^2}{a^2}+\dfrac{(y-1)^2}{b^2}=1$, where $a > b > 0$.

(Note that we still get $c^2 = a^2 - b^2$, because a and b are constant, and so is c. And translations do not change shapes, so the focal distance c does not change either.)

And in general, we can put a perpendicular ellipse in an equation the way as follows.

$$\frac{(x-u)^2}{a^2}+\frac{(y-v)^2}{b^2}=1$$

Then, the ellipse above is centered at a point (u, v), and the main axes are $2a$ and $2b$.

If $a > b > 0$, the ellipse is horizontal, and the major axis is a, and parallel to the x-axis.

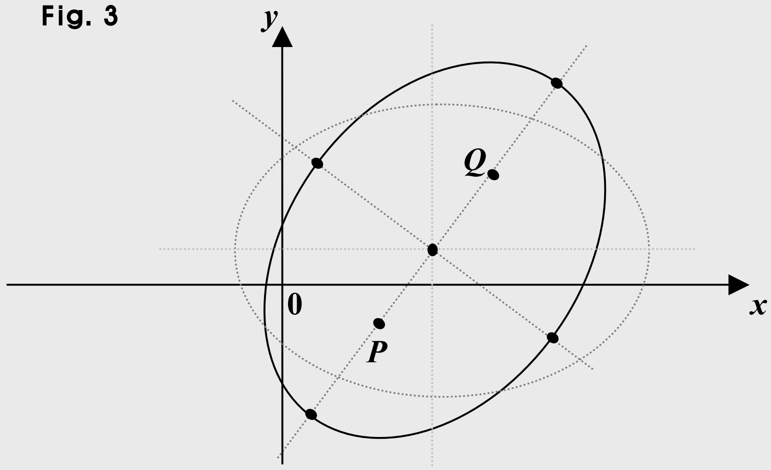
If $b > a > 0$, the ellipse is vertical, and the major axis is b, and parallel to the y-axis.

And the equation above is called the standard equation of an ellipse.

(Note however, the standard equation can indicate a perpendicular ellipse only. If an ellipse is perpendicular, one of the main axes is perpendicular to the x-axis.)

What if an ellipse is not perpendicular, thus, slanted as the ellipse below?

Fig. 3



Given the foci, you can derive the equation using the definition for ellipses, or can get the equation applying a transformation to a perpendicular ellipse.
The transformation is covered in the book, **GRAPH OPERATIONS**, and is covered in the section, Transformation by Turning. What then about the ellipse as follows?

The center is at (0, 0), its main axes are $\mathbf{2a}$ and $\mathbf{2b}$, and the major axis is this time, parallel to the $\boldsymbol{y}$-axis, and is $\mathbf{2b}$.

Then, the ellipse is vertical, and the equation is $\dfrac{x^2}{a^2}+\dfrac{y^2}{b^2}=1$, where $\boldsymbol{b > a > 0}$.

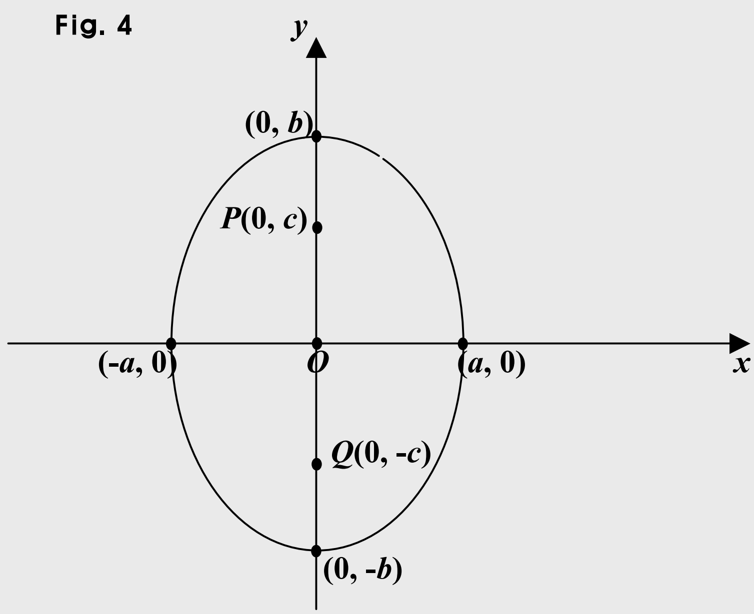
So the equation itself looks quite the same as the equation of the horizontal ellipse $\boldsymbol{E}$. We know however, the major > the minor. So the major radius is $\boldsymbol{b}$, because $\underline{\boldsymbol{b > a > 0}}$. What then are the domain and the range?

The domain is still $\boldsymbol{-a \le x \le a}$, that is, $\boldsymbol{|x| \le a}$, and also, the range is $\boldsymbol{-b \le y \le b}$ or $\boldsymbol{|y| \le b}$.

And we know the foci are in the major axis, which is this time, parallel to the $\boldsymbol{y}$-axis, and the center is the origin, so if $\boldsymbol{c}$ is the focal distance, the foci are $\mathbf{(0,\ c)}$ and $\mathbf{(0,\ -c)}$.

Assuming thus, the vertical ellipse above is $\boldsymbol{V}$, we can put $\boldsymbol{V}$ in a graph the way as follows.

Fig. 4



How then can we get the equation?

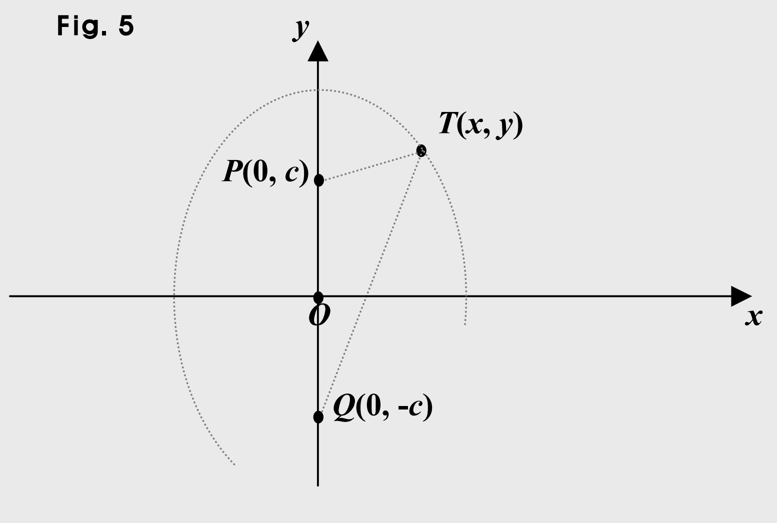
We know the definition for ellipses, which says the fact that an ellipse is a set of points, from each of which, the sum of the two distances to two points called the foci is constant.

Using thus, the fact again, we can get the equation of an ellipse vertical, too.

So to begin with, suppose $c > 0$, two points $P(0, c)$ and $Q(0, -c)$ are the two foci of an ellipse called V, and a point $T(x, y)$ is an arbitrary point in the ellipse V, i.e., a random point representing all the points in V.

Then, the center is the origin $O(0, 0)$, and we can say that c is the focal distance, because the center is the midpoint between the two foci. So we can put the three points P, Q, and T in the x-y plane the way as follows.

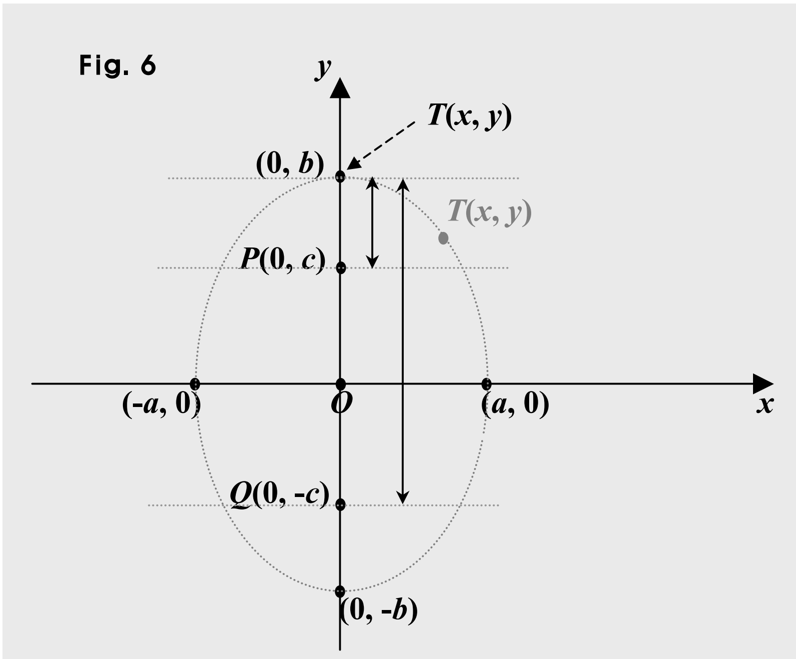
Fig. 5



Suppose next, the point T is moving now, along the ellipse V.

And we know the fact that the sum of the two distances from each point in V to the two foci is constant, that is, the same. So this time also, we are going to use the fact again.

So suppose this time, that the point T is now at the point $(0, b)$.
Then, the sum of the two distances is $2b$. How?



We can see that the length of ***TQ*** is ***b* – (-*c*) = *b* + *c***.

And the length of ***TP*** is ***b* – *c***. So we get ***TP* + *TQ* = (*b* – *c*) + (*b* + *c*) = 2*b***.

Therefore, no matter where the point ***T*** may be in the ellipse ***V***, the sum of the two distances from the point ***T*** to the two foci is always **2*b***, which is the major axis.

So we are going to get the sum of the two distances from ***T*(*x*, *y*)** to the two foci, and then, set the sum equal to **2*b***.
How then can we get the two distances from ***T*(*x*, *y*)** to the two foci?

We can use the distance formula, often called Pythagorean Theorem.
And using it, we can get an equation expressed in terms of ***x*** and ***y***, that is, we get the connective equation between ***x*** and ***y***, which are the coordinates of the arbitrary point ***T*(*x*, *y*)** in the curve called the ellipse ***V***. What then is the connective equation?

The equation explains every point in the ellipse ***V***, that is, it defines the ellipse ***V***.
So the equation is the equation of the ellipse ***V***.
Let's now get the equation. We can get it using the distance formula.
How then can we use the formula?

Assuming d is the distance between two points, Δx is the difference in x-coordinates, and Δy is the difference in y-coordinates, we can put the formula the way as follows.

$d^2 = (\Delta x)^2 + (\Delta y)^2$ And we have $P(0, c)$, $Q(0, -c)$, and $T(x, y)$.

So using the distance formula, and assuming $TP = p$, and $TQ = q$, we can get

$p^2 = (x - 0)^2 + (y - c)^2 = x^2 + (y - c)^2$, and $q^2 = (x - 0)^2 + \{y - (-c)\}^2 = x^2 + (y + c)^2$.

So we get $TP = \sqrt{x^2 + (y - c)^2}$, and $TQ = \sqrt{x^2 + (y + c)^2}$.

Thus, we get $TP + TQ = \sqrt{x^2 + (y - c)^2} + \sqrt{x^2 + (y + c)^2} = 2b$. Then, we get

$$\sqrt{x^2 + (y - c)^2} = 2b - \sqrt{x^2 + (y + c)^2} \Rightarrow x^2 + (y - c)^2 = (2b - \sqrt{x^2 + (y + c)^2})^2$$

$$\Rightarrow x^2 + (y - c)^2 = 4b^2 - 4b\sqrt{x^2 + (y + c)^2} + x^2 + (y + c)^2$$

$$\Rightarrow 4b\sqrt{x^2 + (y + c)^2} = (y + c)^2 - (y - c)^2 + 4b^2 = y^2 + 2cy + c^2 - y^2 + 2cy - c^2 + 4b^2$$

$$\Rightarrow 4b\sqrt{x^2 + (y + c)^2} = 4b^2 + 4cy \Rightarrow b\sqrt{x^2 + (y + c)^2} = b^2 + cy$$

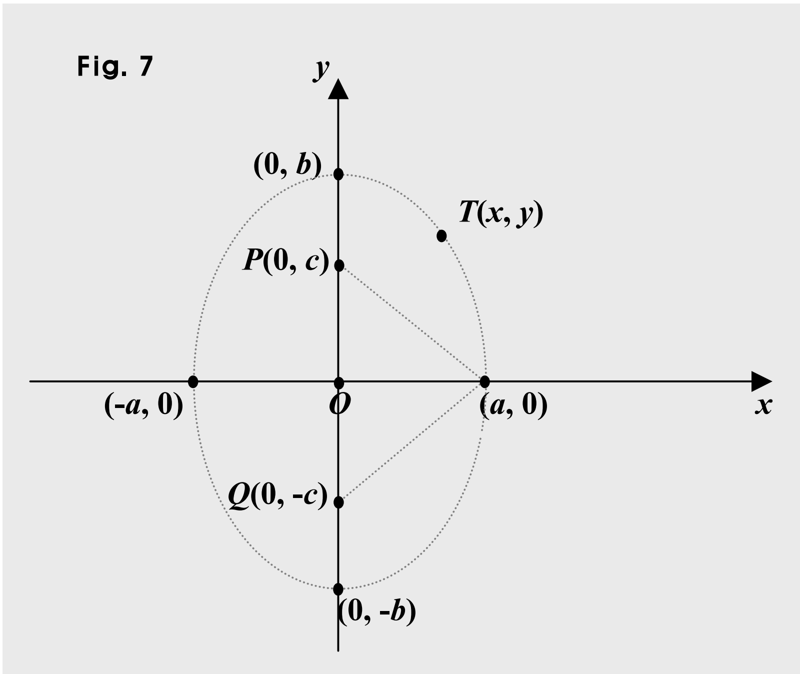
$$\Rightarrow b^2\{x^2 + (y + c)^2\} = (b^2 + cy)^2 = b^4 + 2b^2cy + c^2y^2$$

$$\Rightarrow b^2(x^2 + y^2 + 2cy + c^2) = b^2x^2 + b^2y^2 + 2b^2cy + b^2c^2 = a^4 + 2b^2cy + c^2y^2$$

$$\Rightarrow b^2x^2 + b^2y^2 + b^2c^2 - c^2y^2 = b^4 \Rightarrow b^2x^2 + (b^2 - c^2)y^2 = b^4 - b^2c^2 = b^2(b^2 - c^2)$$

$$\Rightarrow \frac{b^2x^2}{b^2 - c^2} + y^2 = b^2 \Rightarrow \frac{x^2}{b^2 - c^2} + \frac{y^2}{b^2} = 1.$$ What then about $b^2 - c^2$?

It is constant, so assuming a is constant, we can set $a^2 = b^2 - c^2$. And assuming $a > 0$, we can put two points $(a, 0)$ and $(-a, 0)$ the way as follows.



Assuming now, that the point T is at $(a, 0)$, we can see a triangle isosceles, and the triangle is ΔTPQ.
And we know that the sum of the two distances from T to the two foci is $2b$.

That is, we have $TP + TQ = 2b$. So we get $TP = TQ = b$.

Assuming thus, O is the origin, we can see the triangle TOP is a right triangle, and its hypotenuse TP is b, and the side PO is c.

So using the distance formula again, we can get $b^2 = a^2 + c^2 \Rightarrow a^2 = b^2 - c^2$.

And we have $\dfrac{x^2}{b^2 - c^2} + \dfrac{y^2}{b^2} = 1$. So we get $\dfrac{x^2}{a^2} + \dfrac{y^2}{b^2} = 1$,

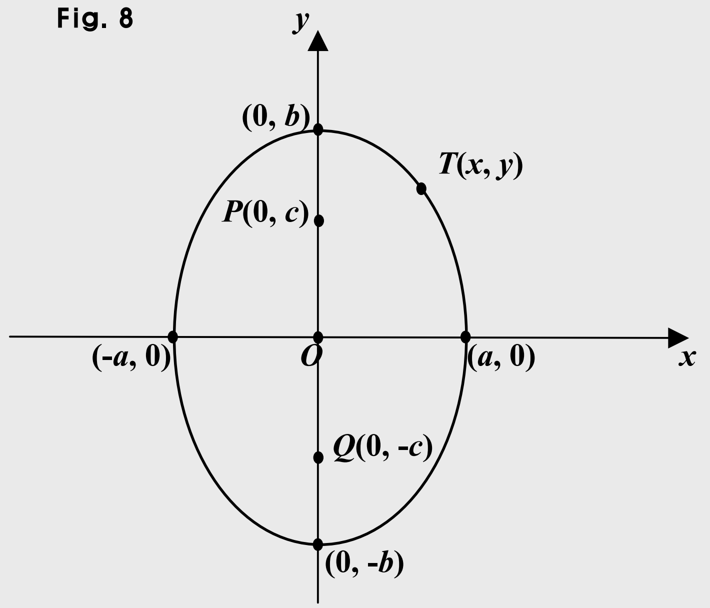
And we get $b > a > 0$, since $a^2 = b^2 - c^2 \Rightarrow b^2 = a^2 + c^2 \Rightarrow b > a > 0$.

Now, the equation above is a connective equation connecting x and y at $T(x, y)$, which is an arbitrary point in the curve called the ellipse V.

Thus, the equation of the ellipse V is $\dfrac{x^2}{a^2} + \dfrac{y^2}{b^2} = 1$, where $b > a > 0$.

And putting the ellipse V in the x-y plane, we get this:

Fig. 8



So we can now say that the ellipse V is vertical and centered at the origin, and that the major axis is $2b$, and the minor axis is $2a$.

That is, the semi major axis is b, and the semi minor is a. In other words, the major radius is b, and is parallel to the y-axis, and the minor radius is a.

And putting the ellipse V in its equation, we get $\frac{x^2}{a^2}+\frac{y^2}{b^2}=1$, where $b > a > 0$.

What then are the domain and the range?

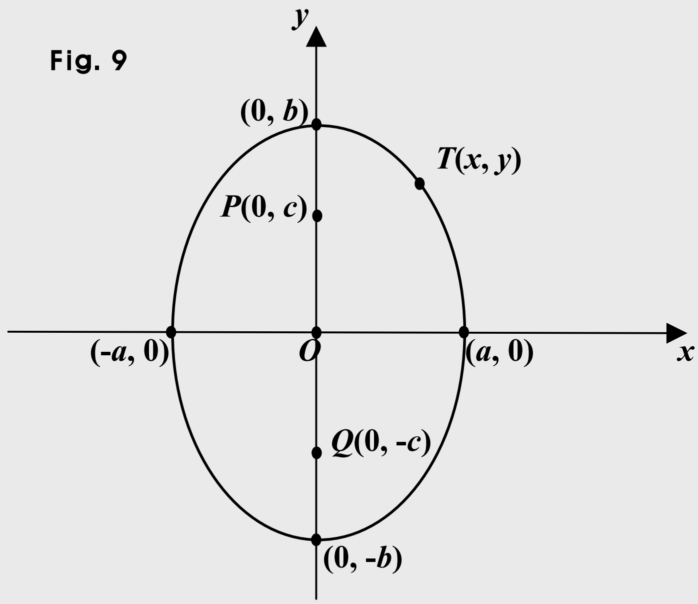
The domain is $-a \leq x \leq a$, that is, $|x| \leq a$, and the range is $-b \leq y \leq b$ or $|y| \leq b$.

Suppose now again, that the point T is moving along the ellipse V.
Then, when will the point T be the farthest away from the center?

When the point T is in the y-axis, T is the farthest away from the center. And we know the major radius is b, and is parallel to the y-axis, because the ellipse V is vertical.

So b is the distance from T to the center when T is the farthest away from the center.

Fig. 9



And the ellipse has two points where ***T*** gets located when ***T*** is the farthest away from the center. And the two are **(0, *b*)** and **(0, -*b*)**, which are called antipodal points.
So the major axis is the line segment connecting the two antipodal points where ***T*** gets positioned when ***T*** is the farthest away from the center, and the two are **(0, *b*)** and **(0, -*b*)**.

Thus, the two antipodal points **(0, *b*)** and **(0, -*b*)** are the endpoints of the major axis, and the length of it is **2*b***.
What then about the case where the point ***T*** is the closest to the center?

When the point ***T*** is in the ***x***-axis, ***T*** is the closest to the center.
And we know this time the minor radius is ***a***, and is parallel to the ***x***-axis, that is, perpendicular to the ***y***-axis, because the ellipse ***V*** is vertical.
So ***a*** is the distance from ***T*** to the center when ***T*** is the closest to the center.

And there are two points where ***T*** gets positioned when ***T*** is the closest to the center. The two points are antipodal, too. And the two are **(*a*, 0)** and **(-*a*, 0)**.
So the minor axis is the line segment connecting the two antipodal points where ***T*** gets positioned when ***T*** is the closest to the center, and the two are **(*a*, 0)** and **(-*a*, 0)**.

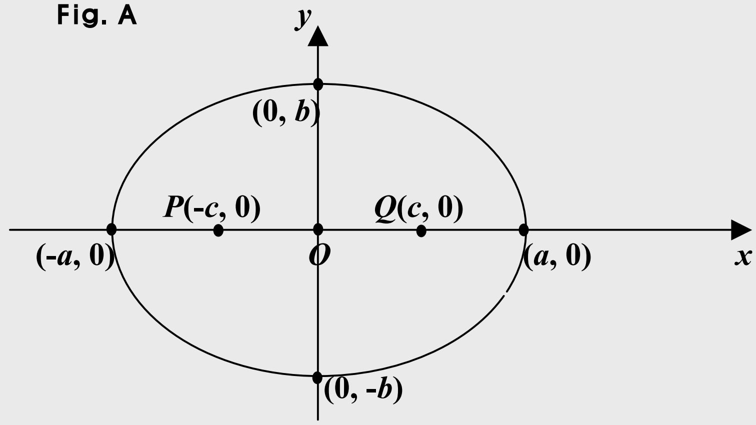
Thus, the two antipodal points **(*a*, 0)** and **(-*a*, 0)** are the endpoints of the minor axis, and the length of it is **2*a***.

So in sum, we have just covered two equations indicating ellipses in two kinds.

One is $\frac{x^2}{a^2}+\frac{y^2}{b^2}=1$, where $a > b > 0$, and $c^2 = a^2 - b^2$, where c is the focal distance.

The ellipse above is horizontal, and is centered at (0, 0), the major axis is $2a$ and is a part of (thus, parallel to) the x-axis, and the minor axis is $2b$. So the ellipse is as follows.

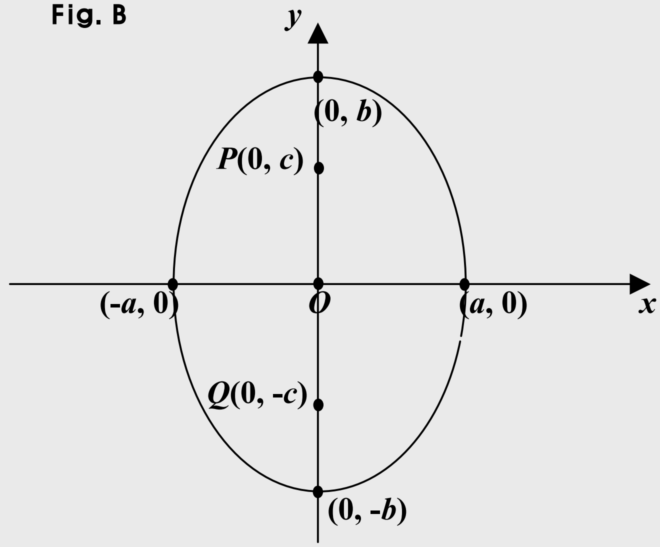
Fig. A



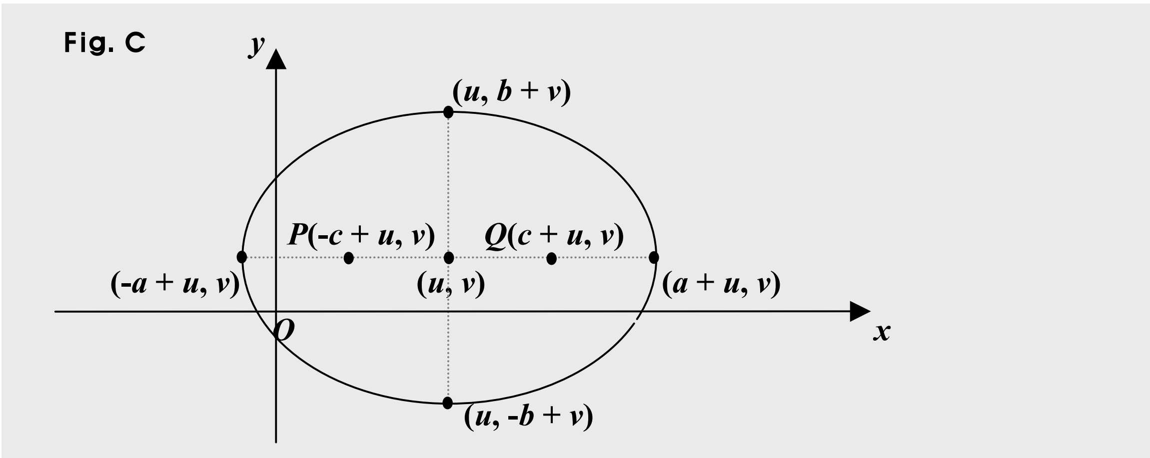
And the other is $\frac{x^2}{a^2}+\frac{y^2}{b^2}=1$, where $b > a > 0$, $c^2 = b^2 - a^2$, and c is the focal distance.

The ellipse above is vertical, and is centered at (0, 0), the major axis is $2b$ and is a part of (thus, parallel to) the y-axis, and the minor axis is $2a$. So the ellipse is as follows.

Fig. B



What if the center is not at the origin as in the case below?

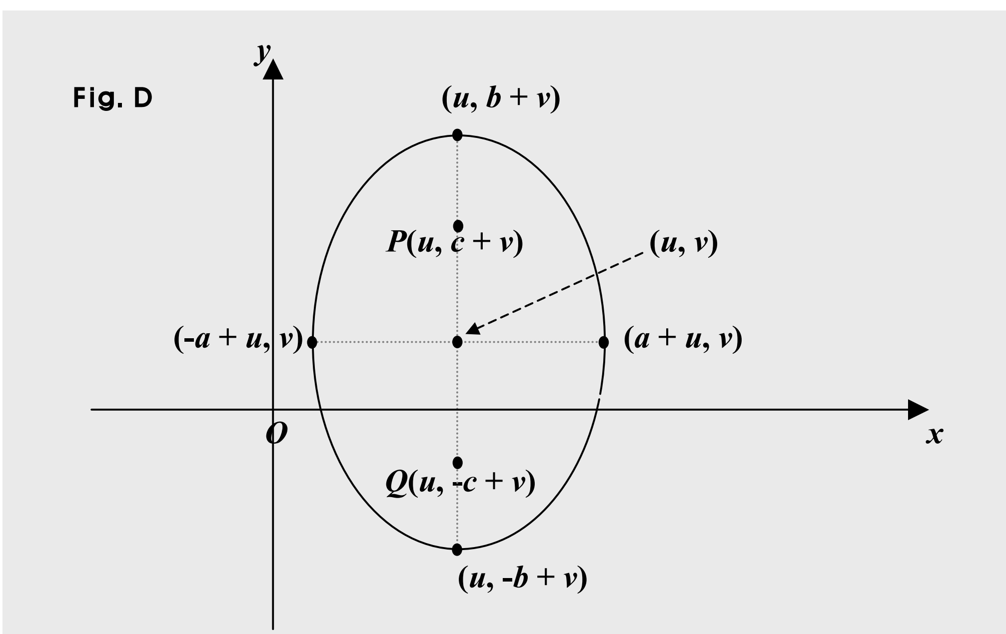


Then, the equation is $\frac{(x-u)^2}{a^2}+\frac{(y-v)^2}{b^2}=\mathbf{1}$, where $a > b > 0$, and $c^2 = a^2 - b^2$.

We can get the ellipse above translating an ellipse in amount of u along the x-axis, and in the amount of v along the y-axis. And the ellipse to be translated is as follows

$\frac{x^2}{a^2}+\frac{y^2}{b^2}=\mathbf{1}$, where $a > b > 0$, and $c^2 = a^2 - b^2$, where c is the focal distance.

What about the case as follows.



The equation is $\dfrac{(x-u)^2}{a^2}+\dfrac{(y-v)^2}{b^2}=1$, where $b > a > 0$, and $c^2 = b^2 - a^2$.

And note that in this case, we have this: $c^2 = b^2 - a^2$, and not this: $c^2 = a^2 - b^2$.

Also, assuming K is the ellipse above, we can get K translating an ellipse in amount of u along the x-axis, and in the amount of v along the y-axis. And the ellipse to be translated is as follows.

$\dfrac{x^2}{a^2}+\dfrac{y^2}{b^2}=1$, where $b > a > 0$, and $c^2 = b^2 - a^2$, where c is the focal distance.

Also, assuming L is the ellipse above, we can get L translating the ellipse K in amount of $-u$ along the x-axis, and in the amount of $-v$ along the y-axis.

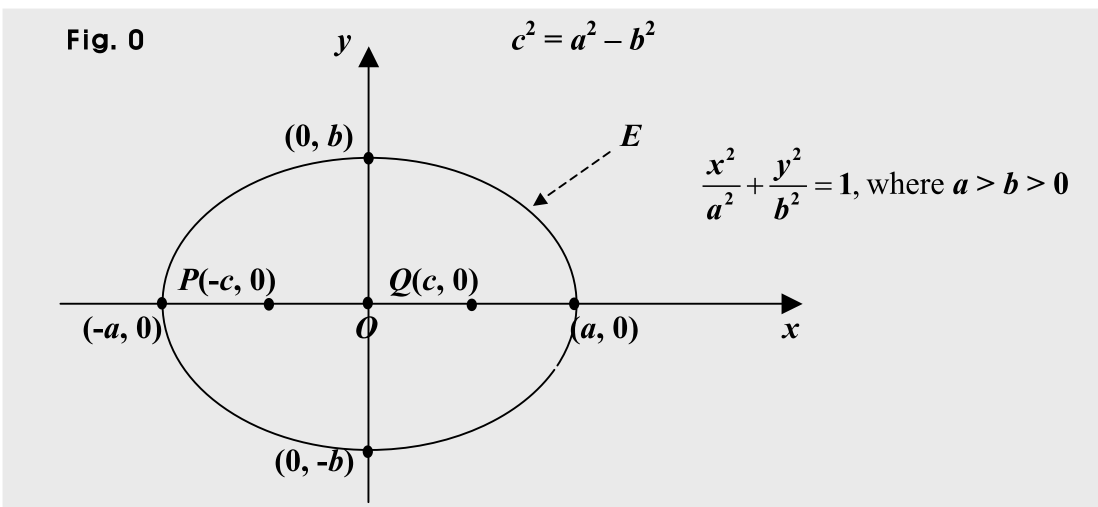
3.3. Elements of an Ellipse

Working with an ellipse, we often need to refer to some basic elements in it.
And the elements are as follows.

The center, foci, directrices, vertices, main axes, focal distance, and eccentricity.

So let's now take a close look at each of those elements.

To begin with, an ellipse has two vertices, which are at both ends of the major axis.



So in the ellipse E above, the two vertices are $(-a, 0)$ and $(a, 0)$. What about P and Q?

The two points $P(-c, 0)$ and $Q(c, 0)$ are the foci, and the center is $O(0, 0)$.

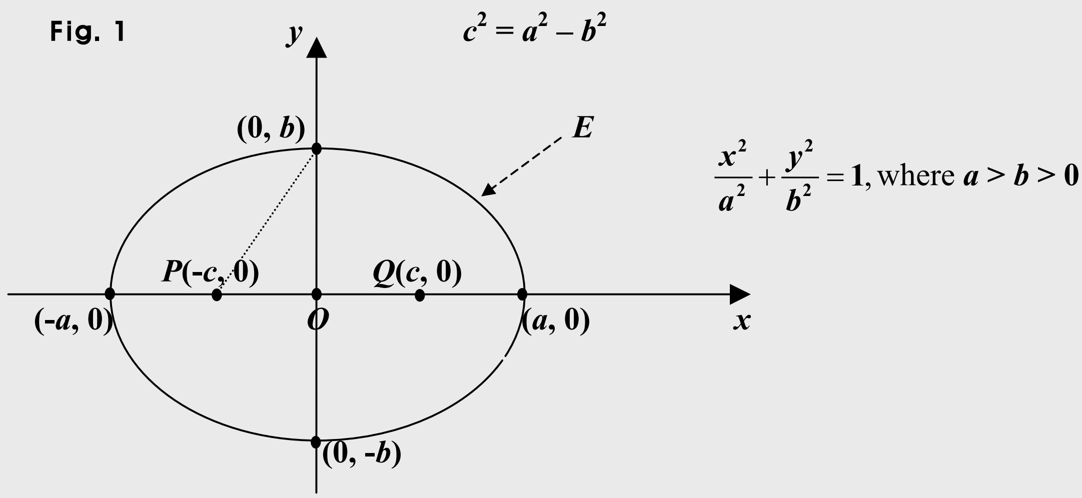
And we can notice that the major axis has all the five points, which are the two vertices, the two foci, and the center.
Also, we can notice that if an ellipse is horizontal, the y-coordinates are the same at all the five points, which share thus, the same y-coordinate.

Then, we can notice also that if an ellipse is vertical, the ***x***-coordinates are the same at all the five points: the vertices, foci, and center, which share thus, the same ***x***-coordinate.

Next, an ellipse is symmetric, and has two axes of symmetry, often called main axes.

An ellipse has two main axes, one is the major axis, longer than the other, called the minor axis. In the ellipse ***E*** in Fig. 1, the line segment connecting the vertices is the major axis, its length is **2*a***, and the line segment connecting two points **(0, *b*)** and **(0, *b*)** is the minor axis, the length of which is **2*b***. Usually though, we just call **2*a*** the major axis, and call **2*b*** the minor axis.

Fig. 1



So in the case of the horizontal ellipse ***E***, we have $a > b > 0$.   What then about ***a*** and ***b***?

We call ***a*** the semi major axis, called the major radius, too, and ***b*** is called the semi minor axis, called the minor radius, too.

And as we can see in the equation above, if an ellipse is horizontal, the major radius is with x^2, and is ***a***.   What then about an ellipse vertical?

If the equation above indicates an ellipse vertical, we get $b > a > 0$.
So the major radius is with y^2, and is ***b***.

Next, the focal distance is the distance from a focus to the center.

The two foci are symmetric about the center.
So the focal distance is half the distance between the foci.
Thus, in the ellipses $\boldsymbol{E}$ above, $\boldsymbol{c}$ is the focal distance.

Next, in the case of the ellipse $\boldsymbol{E}$, the two foci are $(-\boldsymbol{c}, \mathbf{0})$ and $(\boldsymbol{c}, \mathbf{0})$.

The sum of two distances from every point in an ellipse to its two foci is constant, that is, the same. And in the case of the ellipse $\boldsymbol{E}$ above, the sum is $\mathbf{2}\boldsymbol{a}$, the major axis.

What then about this $\boldsymbol{c}^2 = \boldsymbol{a}^2 - \boldsymbol{b}^2$?

It can be called the connective equation between the focal distance and the semi axes.
And in the case of the ellipse $\boldsymbol{E}$, we can get the equation from the three facts as follows.

1. One is that $\boldsymbol{c}$ is the focal distance, that is, the distance from one focus to the center.

2. Another is that the distance from one end of the minor axis to one focus is the major radius, that is, $\boldsymbol{a}$.

3. And the other fact is that a right triangle is made by the three points stated above, one is the one end of the minor axis, another is the one focus, and the other is the center.

And in the case of the ellipse $\boldsymbol{E}$, the three points can be $(\mathbf{0}, \boldsymbol{b})$, $(-\boldsymbol{c}, \mathbf{0})$, and $(\mathbf{0}, \mathbf{0})$.

So the three points above determine a right triangle, where the hypotenuse is $\boldsymbol{a}$, which is the major radius. Using thus, the distance formula, we can get $\boldsymbol{a}^2 = \boldsymbol{b}^2 + \boldsymbol{c}^2$.

So we get $\boldsymbol{c}^2 = \boldsymbol{a}^2 - \boldsymbol{b}^2$. Therefore, the focal distance is $\boldsymbol{c} = \sqrt{\boldsymbol{a}^2 - \boldsymbol{b}^2}$.

And in the case of the ellipse $\boldsymbol{E}$ horizontal, the two foci are $(-\boldsymbol{c}, \mathbf{0})$ and $(\boldsymbol{c}, \mathbf{0})$.

So we can put them this way, too: $(-\sqrt{\boldsymbol{a}^2 - \boldsymbol{b}^2}, \mathbf{0})$ and $(\sqrt{\boldsymbol{a}^2 - \boldsymbol{b}^2}, \mathbf{0})$.
What if the ellipse is vertical?

Then, we have $\boldsymbol{b} > \boldsymbol{a} > \mathbf{0}$. So we get $\boldsymbol{c}^2 = \boldsymbol{b}^2 - \boldsymbol{a}^2$.

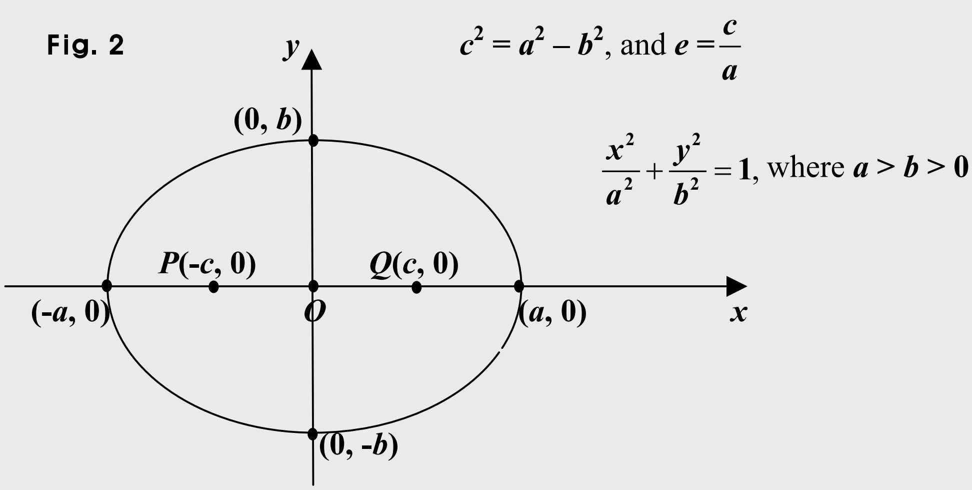
Therefore, in the case of an ellipse vertical, the focal distance is $\boldsymbol{c} = \sqrt{\boldsymbol{b}^2 - \boldsymbol{a}^2}$.

And if an ellipse is vertical and centered at the origin, the two foci are **(0, *c*)** and **(0, -*c*)**.

So we can put the foci this way, too: $(0, \sqrt{b^2 - a^2})$ and $(0, -\sqrt{b^2 - a^2})$.

Next, an ellipse has a special number called an ***eccentricity***, which is a ratio of a focal distance to a major radius, that is, ***the focal distance over the major radius***.

Fig. 2



So of an ellipse, the eccentricity specifies the degree, to which the ellipse is out of round. Thus, it can tell us how flat or round an ellipse is.
Usually, an eccentricity is denoted by e, and we have $0 < e < 1$ for an ellipse.

And if the eccentricity e were 0, it would be a circle.

If e is close to 1, the ellipse is very flat, and looks like a bar.

So the bigger the eccentricity, the flatter the ellipse. What number though, is e?

The eccentricity e is a ratio, which is the focal distance over the major radius.

So in the case of an ellipse horizontal, we get $e = \dfrac{c}{a}$, since $a > b > 0$.

And in the case of an ellipse vertical, we get $e = \dfrac{c}{b}$, since $b > a > 0$.

So we can put the focal distance c the ways below, too:

If $a > b > 0$, since $e = \dfrac{c}{a}$, we get $c = ae$, and if $b > a > 0$, since $e = \dfrac{c}{b}$, we get $c = be$.

How though, would it be a circle if it were the case where $\boldsymbol{e} = \mathbf{0}$?

If $\boldsymbol{a} > \boldsymbol{b} > \mathbf{0}$, we have $\boldsymbol{e} = \dfrac{\boldsymbol{c}}{\boldsymbol{a}}$, and if $\boldsymbol{b} > \boldsymbol{a} > \mathbf{0}$, we have $\boldsymbol{e} = \dfrac{\boldsymbol{c}}{\boldsymbol{b}}$.

So either way, we get $\boldsymbol{c} = \mathbf{0}$ if $\boldsymbol{e} = \mathbf{0}$.

And next, if $\boldsymbol{a} > \boldsymbol{b} > \mathbf{0}$, we have $\boldsymbol{c} = \sqrt{\boldsymbol{a}^2 - \boldsymbol{b}^2}$, and if $\boldsymbol{b} > \boldsymbol{a} > \mathbf{0}$, we have $\boldsymbol{c} = \sqrt{\boldsymbol{b}^2 - \boldsymbol{a}^2}$.

So either way, if $\boldsymbol{c} = \mathbf{0}$, we get $\boldsymbol{a} = \boldsymbol{b}$, that is, both the major and minor radii are the same.

Therefore, if $\boldsymbol{e} = \mathbf{0}$, it would be not an ellipse but a circle with a radius of $\boldsymbol{a}$ (or $\boldsymbol{b}$).
What if this time, the eccentricity $\boldsymbol{e} = \mathbf{1}$?

If first, $\boldsymbol{a} > \boldsymbol{b} > \mathbf{0}$, we have $\boldsymbol{e} = \dfrac{\boldsymbol{c}}{\boldsymbol{a}}$. So if $\boldsymbol{e} = \mathbf{1}$, we would get $\boldsymbol{c} = \boldsymbol{a}$.

And if $\boldsymbol{a} > \boldsymbol{b} > \mathbf{0}$, we have $\boldsymbol{c} = \sqrt{\boldsymbol{a}^2 - \boldsymbol{b}^2}$. So if $\boldsymbol{c} = \boldsymbol{a}$, we would get $\boldsymbol{b} = \mathbf{0}$.

And if next, $\boldsymbol{b} > \boldsymbol{a} > \mathbf{0}$, we have $\boldsymbol{e} = \dfrac{\boldsymbol{c}}{\boldsymbol{b}}$. So if $\boldsymbol{e} = \mathbf{1}$, we would get $\boldsymbol{c} = \boldsymbol{b}$.

And if $\boldsymbol{b} > \boldsymbol{a} > \mathbf{0}$, we have $\boldsymbol{c} = \sqrt{\boldsymbol{b}^2 - \boldsymbol{a}^2}$. So if $\boldsymbol{c} = \boldsymbol{b}$, we would get $\boldsymbol{a} = \mathbf{0}$.

So either way, if the eccentricity $\boldsymbol{e} = \mathbf{1}$, the minor axis would be 0.
And if the minor axis were 0, it would be not an ellipse but a line segment, which is in fact, the major axis, and the foci would be at both ends of the line segment.

So it would be a line segment with the length of the major axis if the eccentricity $\boldsymbol{e} = \mathbf{1}$.

How though, if $\boldsymbol{e} = \mathbf{1}$, would the foci be at both ends of the major axis?

Let's take another look at the eccentricity, but this time, a bit differently.

If an ellipse is horizontal, that is, if $\boldsymbol{a} > \boldsymbol{b} > \mathbf{0}$, we have $\boldsymbol{e} = \dfrac{\boldsymbol{c}}{\boldsymbol{a}}$, and $\boldsymbol{c} = \sqrt{\boldsymbol{a}^2 - \boldsymbol{b}^2}$.

Then, we can put the eccentricity $\boldsymbol{e}$ this way, too: $\boldsymbol{e} = \dfrac{\sqrt{\boldsymbol{a}^2 - \boldsymbol{b}^2}}{\boldsymbol{a}} = \sqrt{\dfrac{\boldsymbol{a}^2 - \boldsymbol{b}^2}{\boldsymbol{a}^2}} = \sqrt{\mathbf{1} - \dfrac{\boldsymbol{b}^2}{\boldsymbol{a}^2}}$.

Then, as $\boldsymbol{b}$ changes and approaches $\boldsymbol{a}$, the eccentricity $\boldsymbol{e}$ changes and approaches 0. And since $\boldsymbol{c} = \sqrt{\boldsymbol{a}^2 - \boldsymbol{b}^2}$, the focal distance $\boldsymbol{c}$ approaches 0. Thus, the foci approach the center.

So as $\boldsymbol{b}$ changes the way above, the ellipse changes and approaches a circle of radius $\boldsymbol{a}$.

Next, as $\boldsymbol{b}$ changes and approaches 0, the eccentricity $\boldsymbol{e}$ changes and approaches 1. And since $\boldsymbol{c} = \sqrt{\boldsymbol{a}^2 - \boldsymbol{b}^2}$, the focal distance $\boldsymbol{c}$ approaches $\boldsymbol{a}$. Thus, the foci approach both ends of the main axis.

So as $\boldsymbol{b}$ changes the way above, the ellipse changes and approaches a line segment of length $\mathbf{2}\boldsymbol{a}$, which is the major axis, and the foci approach both ends of the line segment.

And the same is true for an ellipse vertical, too, that is, the case where $\boldsymbol{b} > \boldsymbol{a} > \mathbf{0}$.
If an ellipse is vertical, that is, if $\boldsymbol{b} > \boldsymbol{a} > \mathbf{0}$, we have $\boldsymbol{e} = \dfrac{\boldsymbol{c}}{\boldsymbol{b}}$, and $\boldsymbol{c} = \sqrt{\boldsymbol{b}^2 - \boldsymbol{a}^2}$.
Then, we can put the eccentricity $\boldsymbol{e}$ this way, too: $\boldsymbol{e} = \dfrac{\sqrt{\boldsymbol{b}^2 - \boldsymbol{a}^2}}{\boldsymbol{b}} = \sqrt{\dfrac{\boldsymbol{b}^2 - \boldsymbol{a}^2}{\boldsymbol{b}^2}} = \sqrt{\mathbf{1} - \dfrac{\boldsymbol{a}^2}{\boldsymbol{b}^2}}$.

Then, as $\boldsymbol{a}$ changes and approaches $\boldsymbol{b}$, the eccentricity $\boldsymbol{e}$ changes and approaches 0. And since $\boldsymbol{c} = \sqrt{\boldsymbol{b}^2 - \boldsymbol{a}^2}$, the focal distance $\boldsymbol{c}$ approaches 0. Thus, the foci approach the center. So as $\boldsymbol{a}$ changes the way above, the ellipse changes and approaches a circle of radius $\boldsymbol{b}$.

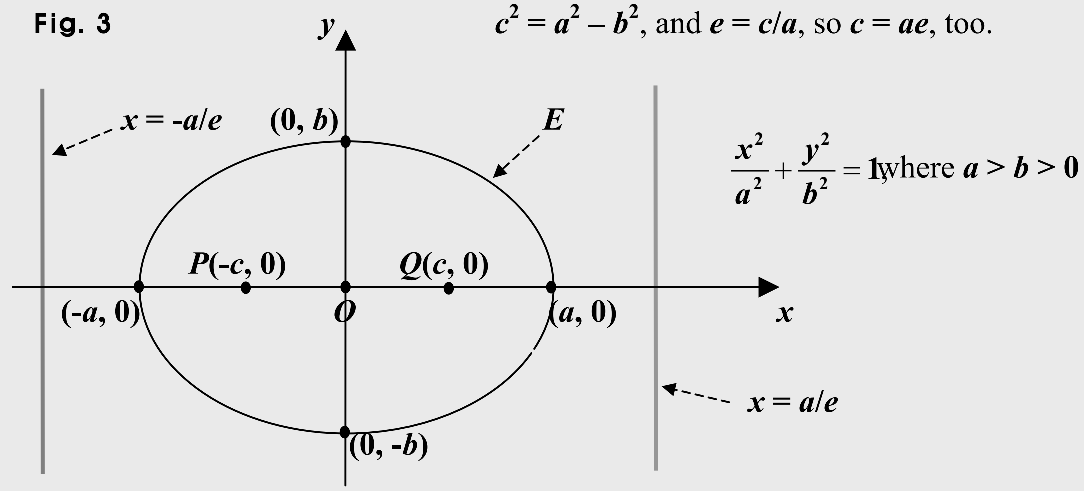
Next, as $\boldsymbol{a}$ changes and approaches 0, the eccentricity $\boldsymbol{e}$ changes and approaches 1. And since $\boldsymbol{c} = \sqrt{\boldsymbol{b}^2 - \boldsymbol{a}^2}$, the focal distance $\boldsymbol{c}$ approaches $\boldsymbol{b}$. Thus, the foci approach both ends of the main axis.

So as $\boldsymbol{a}$ changes the way above, the ellipse changes and approaches a line segment of length $\mathbf{2}\boldsymbol{b}$, which is the major axis, and the foci approach both ends of the line segment.

And next, as in the cases of parabolas, ellipses have ***directrices***, too.

Unlike a parabola though, an ellipse has not one but two directrices.

Fig. 3



So as we can see in the figure above, the two vertical lines in gray are the two directrices of the ellipse E, one is on the left of the ellipse, and the other is on the right of the ellipse.

And if an ellipse is horizontal, a directrix is a vertical line, that is, a line perpendicular to the x-axis. And the focus on the left is said to correspond to the directrix on the left, and the focus on the right is said to correspond to the directrix on the right.

So in the figure above, the vertical line $x = -a/e$ is the directrix corresponding to the focus P, and the other line $x = a/e$ is the directrix corresponding to the focus Q.

How though, are the two directrices of E the two lines $x = -\frac{a}{e}$, and $x = \frac{a}{e}$?

To begin with, in a parabola, the distance from a point to the focus is the same as the distance from the point to the directrix.

In an ellipse however, the distance from a point to a focus is less than the distance from the point to the directrix corresponding to the focus. It's because $0 < e < 1$ for an ellipse.

So next, we can use the fact above the way as follows.
Suppose first, d is the distance from a point in a parabola to the focus, and D is the distance from the point to the directrix.

Then, we get $d = D$. In other words, we get $\frac{d}{D} = 1$.

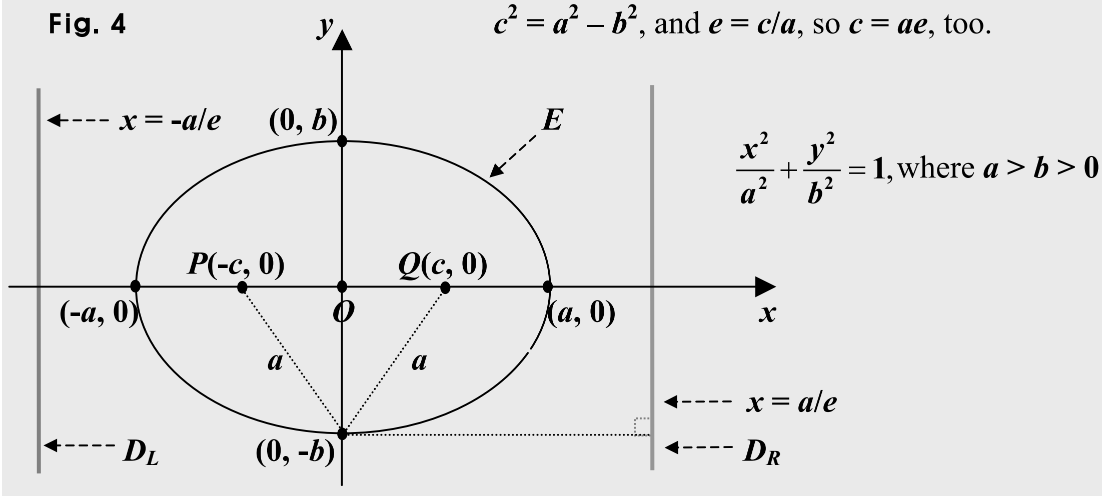
Suppose this time also, ***d*** is the distance from a point in an ellipse to one focus, and ***D*** is the distance from the point to the directrix corresponding to the focus.

Then, we get $d = eD$ where e is the eccentricity. In other words, we get $\frac{d}{D} = e$.

And the same is true for the other directrix, too. What then are the two directrices?

In the case of the ellipse ***E*** below, the two directrices are $x = -\frac{a}{e}$, and $x = \frac{a}{e}$. Why?

Fig. 4



Let's now check to see if the directrices are $x = -\frac{a}{e}$, and $x = \frac{a}{e}$.

Suppose first, $x = m$ is the equation of the directrix D_R shown in the figure above.

Then, in the figure above, we can see that ***m*** is the distance from **(0, -*b*)** (which is a point in ***E***) to the directrix D_R, and ***a*** is the distance from **(0, -*b*)** to the focus ***Q***.

Next, we have $d = eD$ where ***d*** is the distance from a point to a focus, and ***D*** is the distance from the point to the directrix corresponding to the focus. So $d = a$, and $D = m$.

Thus, we get $d = eD \Rightarrow a = em \Rightarrow m = \frac{a}{e}$. So the directrix D_R is the line $x = \frac{a}{e}$.

Let's use this time, another point **(*a*, 0)** to check to see if it is the case.

Then, first, taking the distance from **(*a*, 0)** to the focus ***Q***, we get $a - ae$, since $c = ae$.

Next, taking the distance from **(*a*, 0)** to the directrix $x = m$, that is, D_R, we get $m - a$.

And next, we have $\boldsymbol{d = eD}$ where $\boldsymbol{d}$ is the distance from a point to a focus, and $\boldsymbol{D}$ is the distance from the point to the directrix corresponding to the focus.

So we get $\boldsymbol{d = a - ae}$, and $\boldsymbol{D = m - a}$. Thus, we get

$$\boldsymbol{d = eD \Rightarrow a - ae = e(m - a) = em - ae \Rightarrow a - ae = em - ae \Rightarrow a = em \Rightarrow m = \frac{a}{e}}.$$

So the directrix $\boldsymbol{D_R}$ is the line $\boldsymbol{x = \frac{a}{e}}$.

And the same is true for the other directrix $\boldsymbol{D_L}$, too, because the ellipse is symmetric about the $\boldsymbol{y}$-axis. We know however, $\boldsymbol{D_L}$ is on the left of the origin.

So the directrix $\boldsymbol{D_L}$ is the line $\boldsymbol{x = -\frac{a}{e}}$.

And let's now find for instance, the directrices of this ellipse: $\boldsymbol{\frac{x^2}{4^2} + \frac{y^2}{2^2} = 1}$.

We have $\boldsymbol{x = \pm\frac{a}{e}}$, $\boldsymbol{e = \frac{c}{a}}$, and $\boldsymbol{c = \sqrt{a^2 - b^2}}$. So first, we can get $\boldsymbol{x = \pm\frac{a^2}{c}}$.

And we have $\boldsymbol{a = 4}$, and $\boldsymbol{b = 2}$.

So finding first, the focal distance, we get $\boldsymbol{c = \sqrt{a^2 - b^2} = \sqrt{4^2 - 2^2} = \sqrt{12} = 2\sqrt{3}}$.

Thus, next, we can get $\boldsymbol{\frac{a^2}{c} = \frac{16}{2\sqrt{3}} = \frac{8}{\sqrt{3}} = \frac{8\sqrt{3}}{3}}$.

So the directrices are $\boldsymbol{x = \pm\frac{8\sqrt{3}}{3} \approx \pm 4.6188}$.

What then about the directrices of this ellipse: $\boldsymbol{\frac{(x-u)^2}{a^2} + \frac{(y-v)^2}{b^2} = 1}$, where $\boldsymbol{a > b > 0}$?

We know that the ellipse $\boldsymbol{E}$ is $\boldsymbol{\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1}$, where $\boldsymbol{a > b > 0}$.

So assuming the ellipse in question is $\boldsymbol{G}$, we can get $\boldsymbol{G}$ translating the ellipse $\boldsymbol{E}$ in the amount of $\boldsymbol{u}$ in the direction of the $\boldsymbol{x}$-axis, and in the amount of $\boldsymbol{v}$ along the $\boldsymbol{y}$-axis.

And we know that the center of the ellipse $\boldsymbol{G}$ is $\boldsymbol{(u, v)}$.
So in short, moving the center of the ellipse $\boldsymbol{E}$ to the point $\boldsymbol{(u, v)}$, we can get $\boldsymbol{G}$.

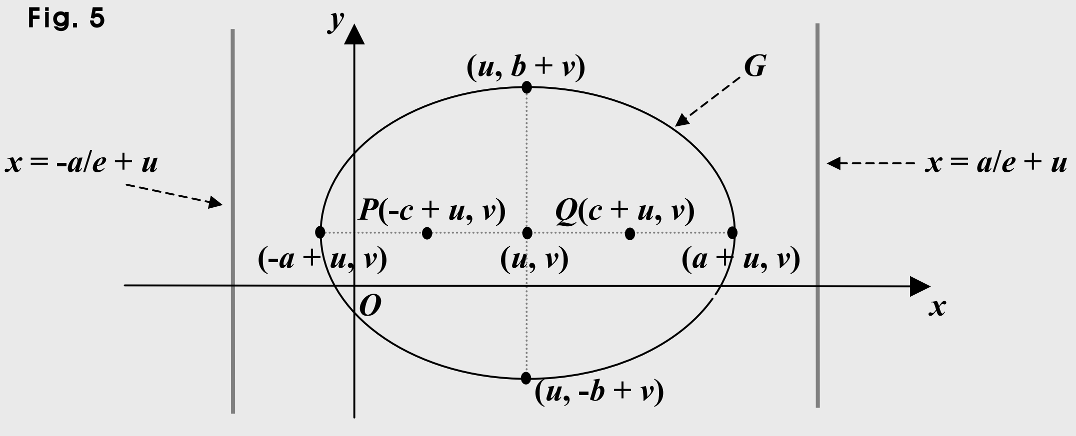
Then, the directrices of $\boldsymbol{E}$ will be translated exactly the way the ellipse $\boldsymbol{E}$ gets translated.

And we know that the directrices of $\boldsymbol{E}$ are $\boldsymbol{x = \pm\frac{a}{e}}$.

So the directrices of $\boldsymbol{G}$ will be as follows. $\boldsymbol{x = \pm\frac{a}{e} + u}$.

That is, the two directrices are $x = -\frac{a}{e} + u$, and $x = \frac{a}{e} + u$.

Fig. 5



Let's now, for instance, find the directrices of this ellipse: $\frac{(x-2)^2}{4^2} + \frac{(y-1)^2}{2^2} = \mathbf{1}$.

First, we've already found the directrices of $\frac{x^2}{2^2} + \frac{y^2}{4^2} = \mathbf{1}$, and they are $\boldsymbol{x} = \pm\frac{8\sqrt{3}}{3}$.

So next, translating the directrices above by 2 along the $\boldsymbol{x}$-axis, and by 1 along the $\boldsymbol{y}$-axis, we get the directrices of the ellipse we want to find the directrices of.
We know however, translating a vertical line along the $\boldsymbol{y}$-axis, we see no change in the line. So assuming (2, 1) is a new center, we just add the $\boldsymbol{x}$-coordinate of the new center to the directrices of the ellipse centered at the origin.

Then, we get $\boldsymbol{x} = \pm\frac{8\sqrt{3}}{3} + \mathbf{2}$, which are the directrices we want. And more specifically,

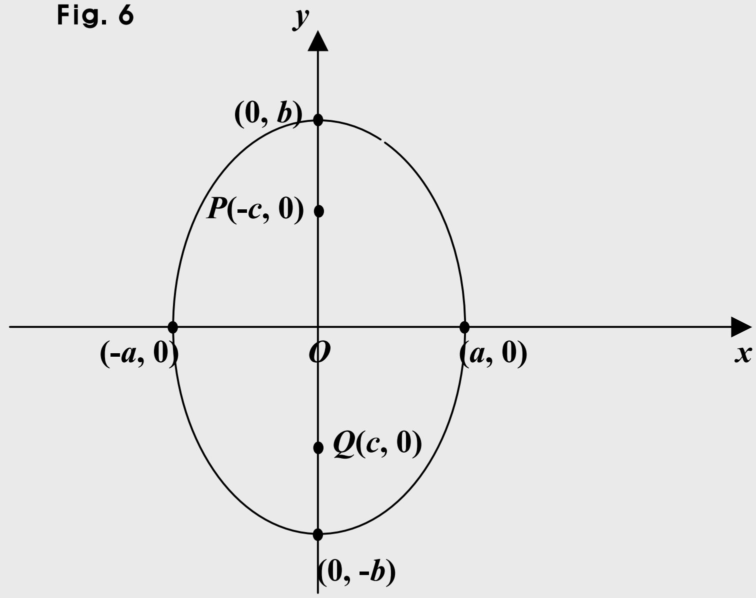
one is $\boldsymbol{x} = -\frac{8\sqrt{3}}{3} + \mathbf{2} = \frac{6-8\sqrt{3}}{3} \approx \mathbf{-2.62}$, which corresponds to the left focus.

And the other is $\boldsymbol{x} = \frac{8\sqrt{3}}{3} + \mathbf{2} = \frac{6+8\sqrt{3}}{3} \approx \mathbf{6.62}$, which corresponds to the right focus.

What if the ellipse is vertical, that is, if it is $\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$, where $b > a > 0$?

Putting first, the ellipse above in a graph, we can put it the way as follows.

Fig. 6



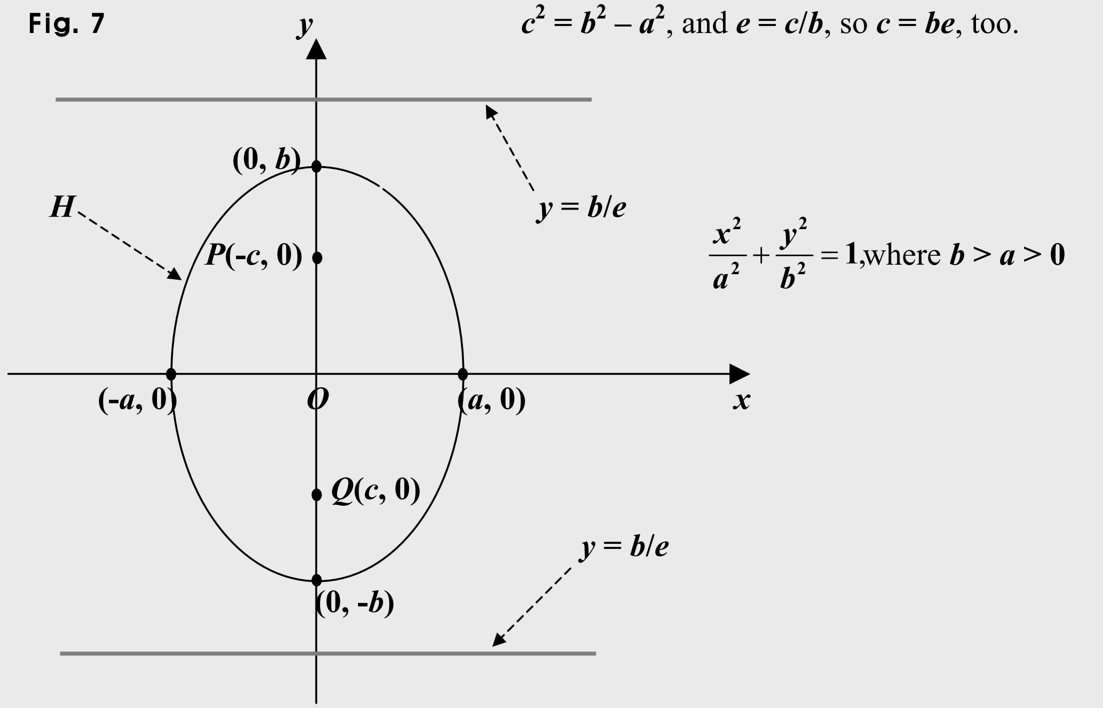
To begin with, we know the fact that if in an ellipse, d is the distance from a point to a focus, and D is the distance from the point to the directrix corresponding to the focus, we get $d = eD$, where e is the eccentricity. In other words, we get $\frac{d}{D} = e$.
Next, the directrices are perpendicular to the major axis.

And we know if an ellipse is horizontal, the major axis is parallel to the x-axis.
So the directrices of an ellipse horizontal are perpendicular to the x-axis.
What then about the directrices of an ellipse vertical?

If the ellipse is vertical, the major axis is parallel to the y-axis.
So the directrices are horizontal, that is, perpendicular to the y-axis.
And next, the directrices of E are $x = \pm\frac{a}{e}$, where a is the major radius, and e is the eccentricity. What then are the directrices of the ellipse vertical?

If the vertical ellipse is V, the directrices of V are $y = \pm\frac{b}{e}$, where b is the major radius.

Fig. 7



So for instance, what are the directrices of this ellipse: $\frac{x^2}{2^2} + \frac{y^2}{4^2} = 1$?

We know finding the directrices of $\frac{x^2}{4^2} + \frac{y^2}{2^2} = 1$, we get $x = \pm\frac{8\sqrt{3}}{3}$.

So the directrices of $\frac{x^2}{2^2} + \frac{y^2}{4^2} = 1$ are $y = \pm\frac{8\sqrt{3}}{3}$. How?

We have $y = \pm\frac{b}{e}$, $e = \frac{c}{b}$, and $c = \sqrt{b^2 - a^2}$. So first, we can get $y = \pm\frac{b^2}{c}$.

And we have $a = 2$, and $b = 4$.

So finding first, the focal distance, we get $c = \sqrt{a^2 - b^2} = \sqrt{4^2 - 2^2} = \sqrt{12} = 2\sqrt{3}$.

Thus, the directrices are $\boldsymbol{y=\pm\frac{b^2}{c}=\pm\frac{16}{2\sqrt{3}}=\pm\frac{8}{\sqrt{3}}=\pm\frac{8\sqrt{3}}{3}}$. That is, $\boldsymbol{y=\pm\frac{8\sqrt{3}}{3}}$.

What then about the directrices of this ellipse: $\boldsymbol{\frac{(x-u)^2}{a^2}+\frac{(y-v)^2}{b^2}=1}$, where $\boldsymbol{b>a>0}$?

Assuming the ellipse above is $\boldsymbol{J}$, we can get $\boldsymbol{J}$ translating the ellipse $\boldsymbol{V}$ in the amount of $\boldsymbol{u}$ in the direction of the $\boldsymbol{x}$-axis, and in the amount of $\boldsymbol{v}$ in the direction of the $\boldsymbol{y}$-axis.

And we know that the center of the ellipse $\boldsymbol{J}$ is $\boldsymbol{(u, v)}$.
So in short, moving the center of the ellipse $\boldsymbol{V}$ to the point $\boldsymbol{(u, v)}$, we can get $\boldsymbol{J}$.

Then, the directrices of $\boldsymbol{V}$ will be translated exactly the way the ellipse $\boldsymbol{V}$ gets translated.

And we know that the directrices of $\boldsymbol{V}$ are $\boldsymbol{y=\pm\frac{b}{e}}$.

So the directrices of $\boldsymbol{J}$ will be as follows. $\boldsymbol{y=\pm\frac{b}{e}+v}$.

That is, the two directrices are $\boldsymbol{y=-\frac{b}{e}+v}$, and $\boldsymbol{y=\frac{b}{e}+v}$.

Let's now, for instance, find the directrices of this ellipse: $\boldsymbol{\frac{(x-3)^2}{2^2}+\frac{(y+5)^2}{4^2}=1}$.

First, we've already found the directrices of $\boldsymbol{\frac{x^2}{2^2}+\frac{y^2}{4^2}=1}$, and they are $\boldsymbol{y=\pm\frac{8\sqrt{3}}{3}}$.

So next, translating those directrices by 3 along the $\boldsymbol{x}$-axis, and by -5 along the $\boldsymbol{y}$-axis, we get the directrices of the ellipse we want to find the directrices of.

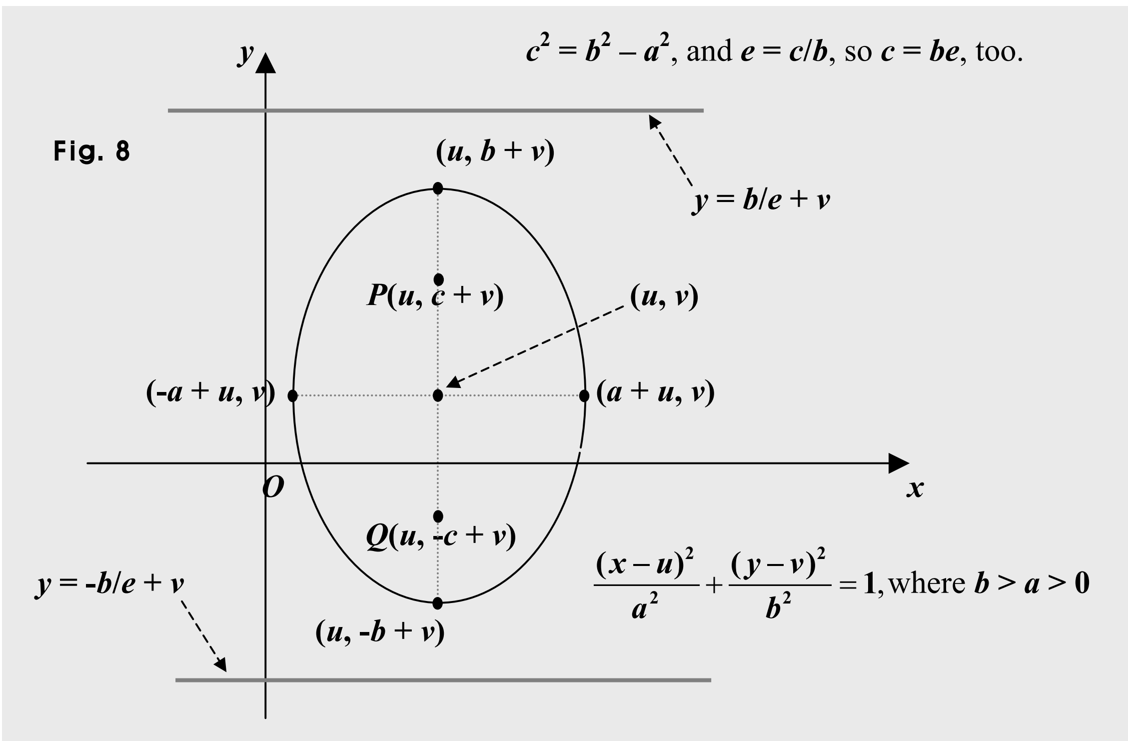
We know however, the directrices above are horizontal, and translating horizontal lines along the $\boldsymbol{x}$-axis, we see no change.

So assuming (3, -5) is a new center, we just add the $\boldsymbol{y}$-coordinate of the new center to the directrices above. Note however, we can do so, because the ellipse getting translated is centered at the origin.

Then, we get $y = \pm\frac{8\sqrt{3}}{3} - 5$, which are the directrices we want. And more specifically,

one is $y = -\frac{8\sqrt{3}}{3} - 5$, which corresponds to the lower focus.

And the other is $y = \frac{8\sqrt{3}}{3} - 5$, which corresponds to the upper focus.



3.4. Summary on Ellipses

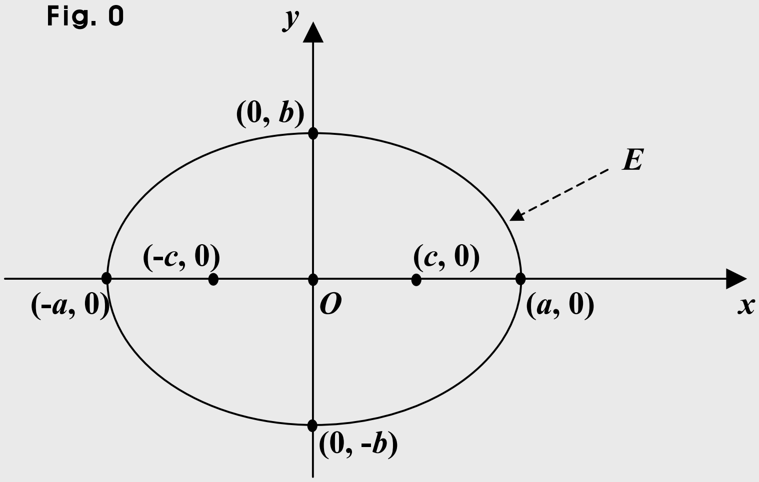
Now, summing up, we can put together the ideas on ellipses the way as follows.
To begin with, putting an ellipse in a standard equation, we can get this:

$$\frac{x^2}{a^2}+\frac{y^2}{b^2}=1, \text{ where } a > b > 0$$

What ellipse then is it?

If ***E*** is the ellipse, ***E*** is centered at **(0, 0)**, the origin, and since $a > b$, it is horizontal, so the major axis is parallel to the ***x***-axis, and is $2a$, and the minor axis is parallel to the ***y***-axis, and is $2b$. And assuming c is the focal distance, we get $c^2 = a^2 - b^2$.

Fig. 0



Next, since the major radius is a, assuming e is the eccentricity, we get $e = \dfrac{c}{a}$, where c is the focal distance, of course. And we have $0 < e < 1$, since $0 < c < a$.

Next, since the ellipse ***E*** is horizontal, the center and foci share the same ***y***-coordinate. And the center is the midpoint between the foci. So since the center is **(0, 0)**, and the focal distance is c, the two foci are $(-c, 0)$ and $(c, 0)$.

Also, since $\boldsymbol{E}$ is horizontal, the center and vertices share the same $\boldsymbol{y}$-coordinate, too. And the center is the midpoint between the vertices. The distance from the center to a vertex is the major radius. So since the major radius is $\boldsymbol{a}$, and the center is **(0, 0)**, the two vertices are $(-\boldsymbol{a}, \mathbf{0})$ and $(\boldsymbol{a}, \mathbf{0})$.

And next, since $\boldsymbol{E}$ is horizontal, the two directrices are two lines parallel to the $\boldsymbol{y}$-axis. The distance from the center to each directrix is $\boldsymbol{a}/\boldsymbol{e}$, where $\boldsymbol{a}$ is the major radius, and $\boldsymbol{e}$ is the eccentricity. So since the center is **(0, 0)**, the two directrices are $\boldsymbol{x} = \pm\dfrac{\boldsymbol{a}}{\boldsymbol{e}}$.

And we know $\boldsymbol{e} = \dfrac{\boldsymbol{c}}{\boldsymbol{a}}$. So we can put the two directrices this way, too: $\boldsymbol{x} = \pm\dfrac{\boldsymbol{a}^2}{\boldsymbol{c}}$.

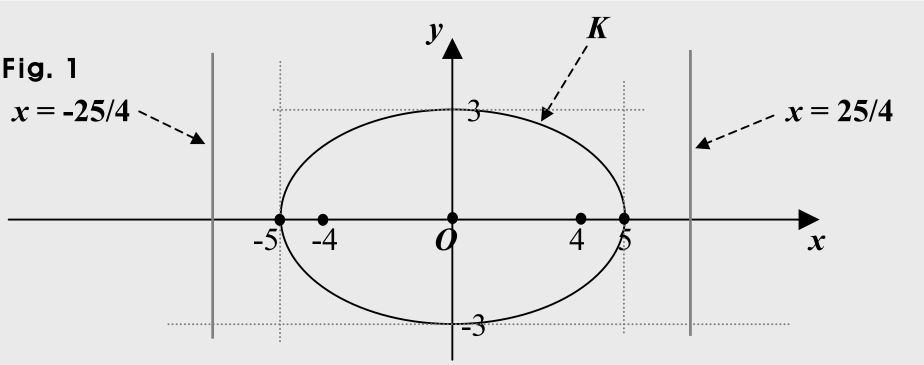
So for instance, putting a horizontal ellipse in an equation, we can get $\dfrac{\boldsymbol{x}^2}{\mathbf{5}^2} + \dfrac{\boldsymbol{y}^2}{\mathbf{3}^2} = \mathbf{1}$.
What ellipse then is it?

If $\boldsymbol{K}$ is the ellipse, $\boldsymbol{K}$ is centered at (0, 0), the origin, and is horizontal, so the major axis is parallel to the $\boldsymbol{x}$-axis, and is 10, and the minor axis is 6.

And the focal distance is 4, because if $\boldsymbol{c}$ is the focal distance, we get $\boldsymbol{c}^2 = \boldsymbol{a}^2 - \boldsymbol{b}^2$, where $\boldsymbol{a}$ is **5** and is the major radius, and $\boldsymbol{b} = \mathbf{3}$ and is the minor radius.
So next, assuming $\boldsymbol{e}$ is the eccentricity, we get $\boldsymbol{e} = \dfrac{\boldsymbol{c}}{\boldsymbol{a}} = \dfrac{4}{5}$, where $\boldsymbol{a}$ is the major radius,

and $\boldsymbol{c}$ is the focal distance. And the directrices are $\boldsymbol{x} = \pm\dfrac{\boldsymbol{a}}{\boldsymbol{e}} = \pm\dfrac{\boldsymbol{a}^2}{\boldsymbol{c}} = \dfrac{5^2}{4} = \dfrac{25}{4}$.

Fig. 1



Next, since the ellipse $\boldsymbol{K}$ is horizontal, the center is (0, 0), and the focal distance is 4, the two foci are (-4, 0) and (4, 0). (Note that horizontal means the same $\boldsymbol{y}$-coordinate.)

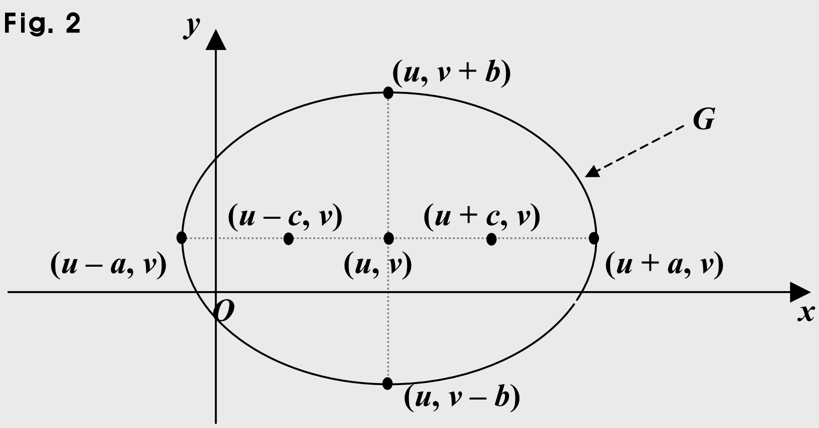
Also, since ***K*** is horizontal, the major radius, that is, the semi major axis is 5 and the center is (0, 0), the two vertices are (-5, 0) and (5, 0).

What then about the ellipse as follows $\frac{(x-u)^2}{a^2}+\frac{(y-v)^2}{b^2}=1$, where $a > b > 0$?

If G is the ellipse above, the ellipse G is centered at (u, v), and since $a > b$, G is horizontal, so the major axis is parallel to the x-axis, and is $2a$, and the minor axis is parallel to the y-axis, and is $2b$.
And assuming c is the focal distance, we get $c^2 = a^2 - b^2$, where a is the major radius, and b is the minor radius.

Next, since the major radius is a, assuming e is the eccentricity, we get $e = \frac{c}{a}$, where c is the focal distance, of course. Also, we have $0 < e < 1$, too, since $0 < c < a$.

Fig. 2



Next, since G is horizontal, the two directrices are two lines parallel to the y-axis. The distance from the center to each directrix is $\frac{a}{e}$, where a is the major radius, and e is the eccentricity. So since the center is (u, v), the two directrices are $x = \pm\frac{a}{e} + u$.

And we know $e = \frac{c}{a}$. So we can put the two directrices this way, too: $x = \pm\frac{a^2}{c} + u$.
Next, since the ellipse G is horizontal, the center is at (u, v), and the focal distance is c, the two foci are $(u - c, v)$ and $(u + c, v)$.

Also, since G is horizontal, the center is (u, v), and the major radius is a, the two vertices are $(u - a, v)$ and $(u + a, v)$.

And notice that translating the ellipse E in the amount of u along the x-axis, and in the amount of v along the y-axis, we get the ellipse G.

Also, of course, translating the ellipse G in the amount of $-u$ along the x-axis, and in the amount of $-v$ along the y-axis, we get the ellipse E.

So for instance, putting a horizontal ellipse in an equation, we can get this:

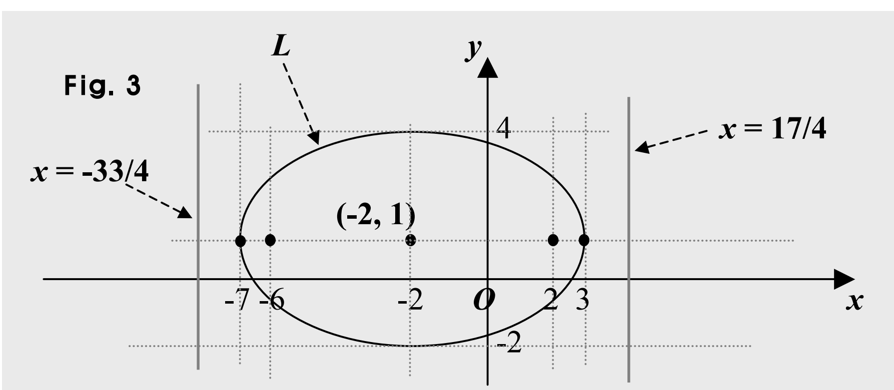
$$\frac{(x+2)^2}{5^2} + \frac{(y-1)^2}{3^2} = 1. \qquad \text{What ellipse then is it?}$$

If L is the ellipse above, the ellipse L is centered at (-2, 1), and is horizontal, so the major axis is parallel to the x-axis, and is 10, and the minor axis is 6.

The focal distance is 4, because if c is the focal distance, we get $c^2 = a^2 - b^2$, where $a = 5$, and a is the major radius, and $b = 3$, and b is the minor radius.

So next, since the major radius is 5, and the focal distance is 4, assuming e is the eccentricity, we get $e = \dfrac{c}{a} = \dfrac{4}{5}$, where c is the focal distance, and a is the major radius.

Fig. 3



And since the center is (-2, 1), the directrices are $x = \pm\dfrac{a}{e} - 2 = \pm\dfrac{a^2}{c} - 2 = \pm\dfrac{25}{4} - 2.$

Next, since the ellipse L is horizontal, the focal distance is 4 and the center is (-2, 1), the two foci are (-6, 1) and (2, 1).

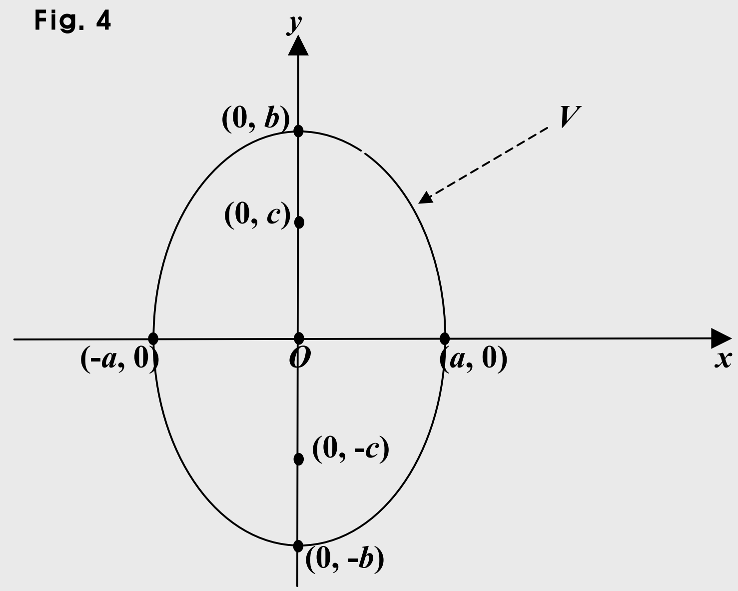
Also, since L is horizontal, the center is (-2, 1), and the major radius, that is, the semi major axis is 5, the two vertices are (-7, 1) and (3, 1).

Next, putting a vertical ellipse in a standard equation, we can get this:

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = \mathbf{1}, \text{ where } b > a > \mathbf{0}. \qquad \text{What ellipse then is it?}$$

If V is the ellipse above, V is centered at **(0, 0)**, the origin, and since $b > a$, it is vertical, so the major axis is parallel to the y-axis, and is $2b$, and the minor axis is parallel to the x-axis, and is $2a$. And assuming c is the focal distance, we get $c^2 = b^2 - a^2$.

Fig. 4



Next, since the major radius is b, assuming e is the eccentricity, we get $e = \dfrac{c}{b}$, where c is the focal distance, of course. And also, we have $0 < e < 1$, too, since $0 < c < b$.
Next, since the ellipse V is vertical, the center and foci share the same x-coordinate.

And the center is the midpoint between the foci. So since the center is **(0, 0)**, and the focal distance is c, the two foci are $(0, c)$ and $(0, -c)$.

Also, since V is vertical, the center and the vertices share the same x-coordinate, too.
And the center is the midpoint between the vertices. So since the semi major axis, that is, the major radius is b, and the center is **(0, 0)**, the two vertices are $(0, b)$ and $(0, -b)$.

Next, since $\boldsymbol{V}$ is horizontal, the two directrices are two lines parallel to the $\boldsymbol{x}$-axis.

The distance from the center to each directrix is $\dfrac{\boldsymbol{b}}{\boldsymbol{e}}$, where $\boldsymbol{b}$ is the major radius, and $\boldsymbol{e}$ is the eccentricity. So since the center is **(0, 0)**, the two directrices are $\boldsymbol{x} = \pm\dfrac{\boldsymbol{b}}{\boldsymbol{e}}$.

And we know $\boldsymbol{e} = \dfrac{\boldsymbol{c}}{\boldsymbol{b}}$. So we can put the two directrices this way, too: $\boldsymbol{x} = \pm\dfrac{\boldsymbol{b}^2}{\boldsymbol{c}}$.

So for instance, putting a vertical ellipse in an equation, we can get $\dfrac{\boldsymbol{x}^2}{\mathbf{3}^2} + \dfrac{\boldsymbol{y}^2}{\mathbf{5}^2} = \mathbf{1}$.

What ellipse then is it?

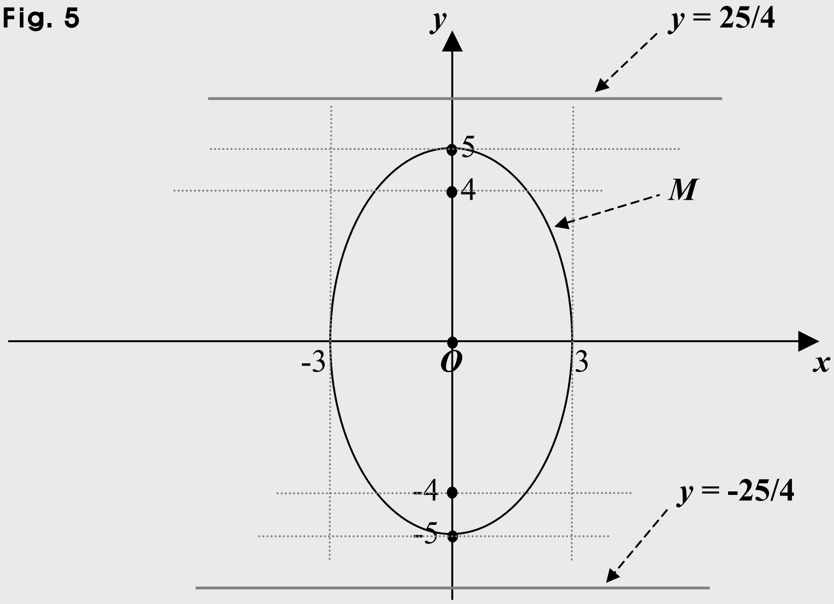
If $\boldsymbol{M}$ is the ellipse above, $\boldsymbol{M}$ is centered at (0, 0), the origin, and is vertical, so the major axis is parallel to the $\boldsymbol{y}$-axis, and is 10, and the minor axis is 6.

The focal distance is 4, because if $\boldsymbol{c}$ is the focal distance, we get $\boldsymbol{c}^2 = \boldsymbol{b}^2 - \boldsymbol{a}^2$, where $\boldsymbol{b} = \mathbf{5}$, and $\boldsymbol{b}$ is the major radius, and $\boldsymbol{a} = \mathbf{3}$, and $\boldsymbol{a}$ is the minor radius.

So next, assuming $\boldsymbol{e}$ is the eccentricity, we get $\boldsymbol{e} = \dfrac{\boldsymbol{c}}{\boldsymbol{b}} = \dfrac{4}{5}$, where $\boldsymbol{b}$ is the major radius, and $\boldsymbol{c}$ is the focal distance.

And since $\boldsymbol{M}$ is vertical, the major radius is $\boldsymbol{b}$, and the center is **(0, 0)**, the two directrices are $\boldsymbol{y} = \pm\dfrac{\boldsymbol{b}}{\boldsymbol{e}} = \pm\dfrac{\boldsymbol{b}^2}{\boldsymbol{c}} = \pm\dfrac{5^2}{4} = \dfrac{25}{4}$.

Fig. 5



Next, since the ellipse $\boldsymbol{M}$ is vertical, the center is (0, 0), and the focal distance is 4, the two foci are (0, 4) and (0, -4). (Note that vertical means the same $\boldsymbol{x}$-coordinate.)

Also, since $\boldsymbol{M}$ is vertical, the semi major axis, that is, the major radius is 5 and the center is (0, 0), the two vertices are (0, 5) and (0, -5).

What then, about the ellipse as follows? $\dfrac{(x-u)^2}{a^2}+\dfrac{(y-v)^2}{b^2}=1$, where $b > a > 0$

If $\boldsymbol{J}$ is the ellipse above, $\boldsymbol{J}$ is centered at (u, v), and is vertical, so the major axis is parallel to the $\boldsymbol{y}$-axis, and is $2b$, and the minor axis is parallel to the $\boldsymbol{x}$-axis, and is $2a$.

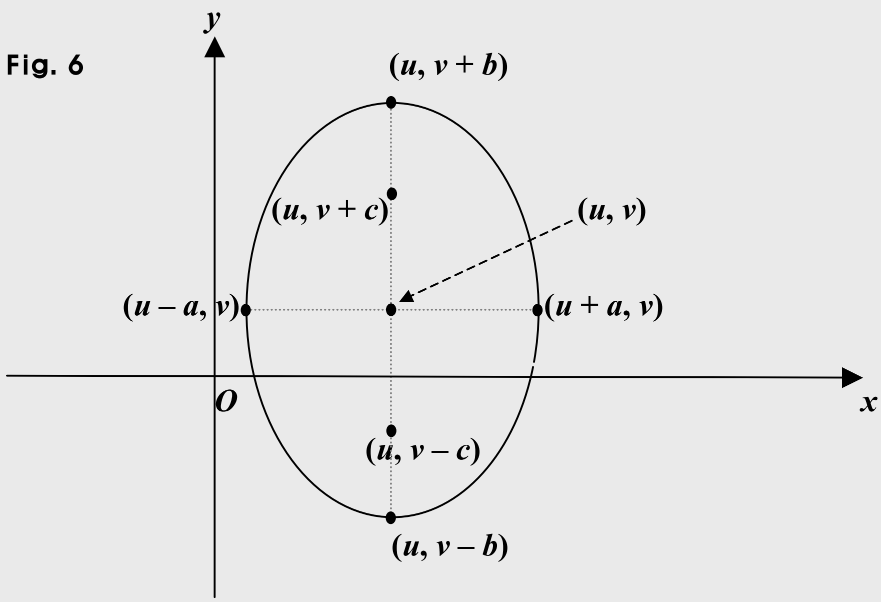
And assuming c is the focal distance, we get $c^2 = b^2 - a^2$.

Next, since the major radius is b, assuming e is the eccentricity, we get $e = c/b$, where c is the focal distance, of course. Also, we have $0 < e < 1$, too, since $0 < c < b$.

Next, since $\boldsymbol{J}$ is vertical, the two directrices are two lines parallel to the $\boldsymbol{x}$-axis. The distance from the center to each directrix is $\dfrac{b}{e}$, where b is the major radius, and e is the eccentricity. So since the center is (u, v), the two directrices are $y = \pm\dfrac{b}{e} + v$.

And we know $\boldsymbol{e} = \dfrac{\boldsymbol{c}}{\boldsymbol{b}}$. So we can put the two directrices this way, too: $\boldsymbol{y} = \pm\dfrac{\boldsymbol{b}^2}{\boldsymbol{c}} + \boldsymbol{v}$.

Fig. 6



Next, since $\boldsymbol{J}$ is vertical, the center and foci share the same $\boldsymbol{x}$-coordinate. So since the center is $(\boldsymbol{u}, \boldsymbol{v})$, and the focal distance is $\boldsymbol{c}$, the two foci are $(\boldsymbol{u}, \boldsymbol{v} + \boldsymbol{c})$ and $(\boldsymbol{u}, \boldsymbol{v} - \boldsymbol{c})$.

Also, since $\boldsymbol{J}$ is vertical, the center and the vertices share the same $\boldsymbol{x}$-coordinate. So since the major radius, that is, the semi major axis is $\boldsymbol{b}$, and the center is $(\boldsymbol{u}, \boldsymbol{v})$, the two vertices are $(\boldsymbol{u}, \boldsymbol{v} + \boldsymbol{b})$ and $(\boldsymbol{u}, \boldsymbol{v} - \boldsymbol{b})$.
So for instance, putting a vertical ellipse in an equation, we can get this:

$$\frac{(\boldsymbol{x}+\mathbf{2})^2}{\mathbf{3}^2} + \frac{(\boldsymbol{y}-\mathbf{1})^2}{\mathbf{5}^2} = \mathbf{1}. \qquad \text{What ellipse then is it?}$$

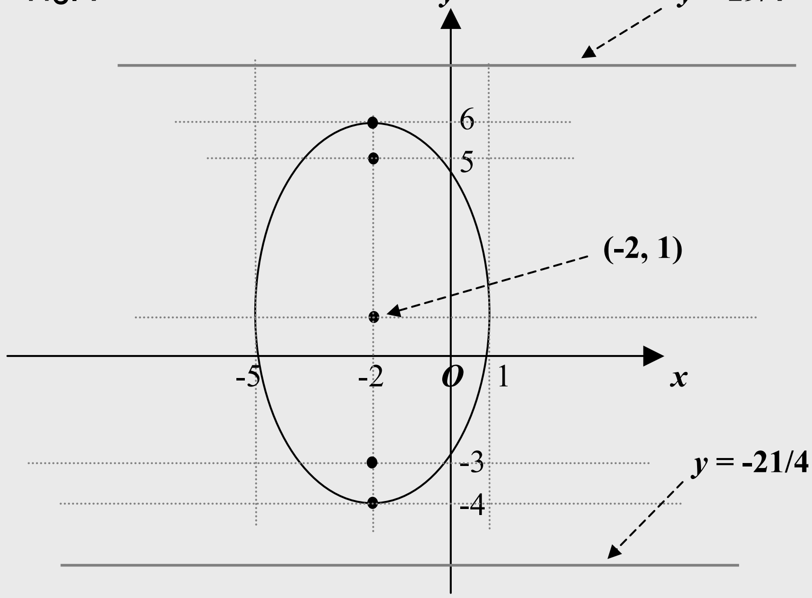
If $\boldsymbol{N}$ is the ellipse above, $\boldsymbol{N}$ is centered at (-2, 1), and is vertical, so the major axis is parallel to the $\boldsymbol{y}$-axis, and is 10, and the minor axis is 6.

And if $\boldsymbol{c}$ is the focal distance, we get $\boldsymbol{c}^2 = \boldsymbol{b}^2 - \boldsymbol{a}^2$, where $\boldsymbol{b} = \mathbf{5}$, $\boldsymbol{a} = \mathbf{3}$, $\boldsymbol{b}$ is the major radius, and $\boldsymbol{a}$ is the minor radius. So the focal distance is 4,

So next, assuming $\boldsymbol{e}$ is the eccentricity, we get $\boldsymbol{e} = \dfrac{\boldsymbol{c}}{\boldsymbol{b}} = \dfrac{4}{5}$, where $\boldsymbol{b}$ is the major radius, and $\boldsymbol{c}$ is the focal distance.

And since N is vertical, the major radius is 5, and the center is **(-2, 1)**, the two directrices are $y = \pm\frac{b}{e} + 1 = \pm\frac{b^2}{c} + 1 = \pm\frac{5^2}{4} + 1 = \pm\frac{25}{4} + 1$.

Fig. 7



Next, since the ellipse N is vertical, the center and foci share the same x-coordinate. So since the center is (-2, 1), and the focal distance is 4, the two foci are (-2, 5) and (-2, -3).

Also, since N is vertical, the center and the vertices share the same x-coordinate, too. So since the major radius, that is, the semi major axis is 5, and the center is (-2, 1), the two vertices are (-2, 6) and (-2, -4).

And we call all the equations above standard equations for ellipses, and the ellipses above can be said to be perpendicular, because each main axis is perpendicular to a coordinate axis.
What about an equation then that can indicate an ellipse perpendicular or not?

Putting an ellipse in such an equation, we can put it the way as follows.

$$ax^2 + by^2 + cxy + ux + vy + w = 0, \text{ where } 4ab > c^2.$$

Note that a and b can be the same, and that $4ab > c^2$ can mean this, too: $ab > 0$, which is saying that a and b have the same sign. And of course, a, b, c, u, v, and w are constant.

Note however, for some values of the constants, the equation does not indicate an ellipse, even if $4ab > c^2$.
For instance, $x^2 + y^2 + xy + x + y - 1 = 0$ and $x^2 + y^2 + xy - 1 = 0$ are ellipse equations, but $x^2 + y^2 + xy + x + y + 2 = 0$ and $x^2 + y^2 + xy + 1 = 0$ are not.

That's because no real numbers for x and y can satisfy the last two equations above.
What then, about an equation that is in the general form and indicates an ellipse perpendicular only?

Putting an ellipse in such an equation, we can put it the way as follows.

$$ax^2 + by^2 + ux + vy + w = 0, \text{ where } a \neq b, \text{ and } ab > 0.$$

So the equation above indicates an ellipse with a main axis perpendicular to the x-axis.
Notice that it does not have the xy-term, that is, cxy in the equation shown earlier.
Note however, for some values of the constants, the equation does not indicate an ellipse, even if $a \neq b$, and $ab > 0$.

For instance, $40x + 36y - 4x^2 - 9y^2 - 100 = 0$ and $4x^2 + 9y^2 - 1 = 0$ are ellipse equations, but $40x + 36y - 4x^2 - 9y^2 - 136 = 0$ and $4x^2 + 9y^2 + 1 = 0$ are not. That's because,

$4x^2 + 9y^2 + 1 = 0 \Rightarrow 4x^2 + 9y^2 = -1 \Rightarrow \dfrac{x^2}{3^2} + \dfrac{y^2}{2^2} = -1$, which is not an ellipse.

And in fact, no real number for x and y can satisfy the equation above.
So the equation has no curve if x and y are real.

Next, moving on to the next equation, we get

$$40x + 36y - 4x^2 - 9y^2 - 136 = 0 \Rightarrow 4x^2 + 9y^2 - 40x - 36y + 136 = 0$$

$$\Rightarrow 4x^2 - 40x + 9y^2 - 36y + 136 = 4(x^2 - 10x) + 9(y^2 - 4y) + 136$$

$$= 4(x^2 - 10x + 25 - 25) + 9(y^2 - 4y + 4 - 4) + 136$$

$$= 4(x^2 - 10x + 25) - 100 + 9(y^2 - 4y + 4) - 36 + 136$$

$= 4(x - 5)^2 + 9(y - 2)^2 = 0 \Rightarrow \dfrac{(x-5)^2}{3^2} + \dfrac{(y-2)^2}{2^2} = 0$, which is not an ellipse.

So putting more specifically a perpendicular ellipse in a general equation, we can put it the way as follows.

$$ax^2 + by^2 + ux + vy + w = 0, \text{ where } a \neq b,\ a > 0,\ b > 0, \text{ and } bu^2 + av^2 > 4abw.$$

That's simply because, putting the equation into the sum of perfect squares, we get

$$ax^2 + by^2 + ux + vy + w = ax^2 + ux + by^2 + vy + w$$

$$= a(x^2 + \frac{u}{a}x) + b(y^2 + \frac{v}{b}y) + w = a(x^2 + \frac{u}{a}x + \frac{u^2}{4a^2} - \frac{u^2}{4a^2}) + b(y^2 + \frac{v}{b}y + \frac{v^2}{4b^2} - \frac{v^2}{4b^2}) + w$$

$$= a(x + \frac{u}{2a})^2 - \frac{u^2}{4a} + b(y + \frac{v}{2b})^2 - \frac{v^2}{4b} + w = 0$$

$$\Rightarrow a(x + \frac{u}{2a})^2 + b(y + \frac{v}{2b})^2 = \frac{u^2}{4a} + \frac{v^2}{4b} - w > 0.$$

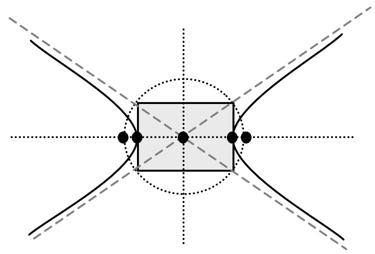
Meanwhile, $\frac{u^2}{4a} + \frac{v^2}{4b} - w = \frac{bu^2 + av^2 - 4abw}{4ab}$. So we get this:

$$\frac{u^2}{4a} + \frac{v^2}{4b} - w > 0 \Rightarrow \frac{bu^2 + av^2 - 4abw}{4ab} > 0 \Rightarrow bu^2 + av^2 - 4abw > 0, \text{ since } ab > 0.$$

Thus, we get $bu^2 + av^2 > 4abw$.

Note however, in this case, we need to assume $a > 0$, and $b > 0$, as well as $a \neq b$.

4.0. What is a hyperbola?



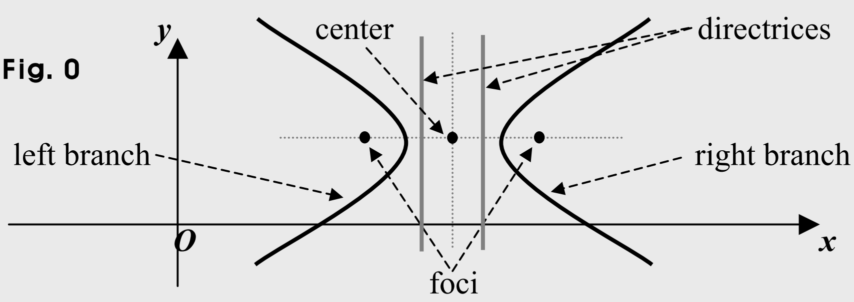
Like an ellipse, a hyperbola is a conic section, often just called a conic, too.
So it is a conic, and is a basic curve. Unlike other conics though, it is not just a single curve, but is made of two. Each is called a branch (or arm). So a hyperbola is made of two branches. And a branch looks like a parabola, yet is quite different from a parabola.

Though looking quite different, a hyperbola and an ellipse have quite a few in common.

A hyperbola has a center and two axes of symmetry, called the main axes, too. One is a major axis, and the other is a minor axis, called an imaginary axis, too. And the two axes are perpendicular to each other, and meet at the center. The major axis is however, often called the transverse axis, and the minor axis is often called the conjugate axis.

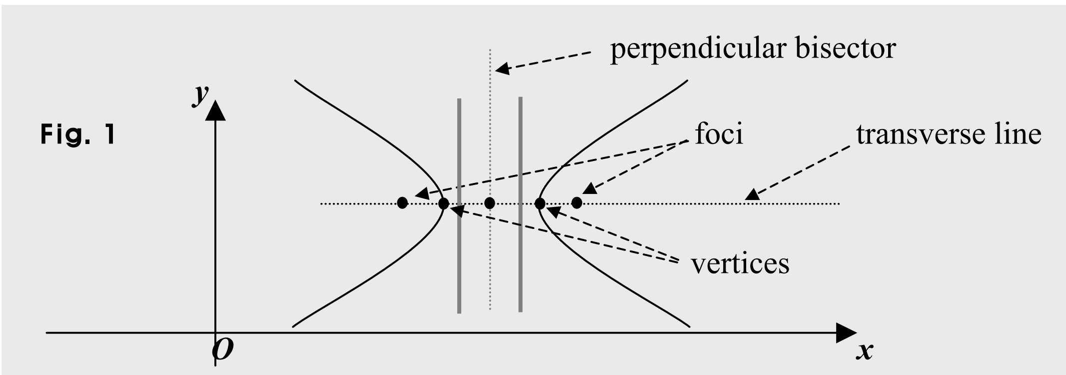
Like an ellipse, a hyperbola has two foci, too, each of which has a corresponding directrix. Thus, a hyperbola has two directrices, too, and each branch has a focus and a directrix corresponding to the focus.

Fig. 0



So we can say that a hyperbola is made of two braches that look like parabolas.

In the figure below, the two dotted lines are perpendicular to each other, and meet at the center. And the dotted line passing through the foci is called the transverse line. And the other line dotted is called the perpendicular bisector of the transverse line.



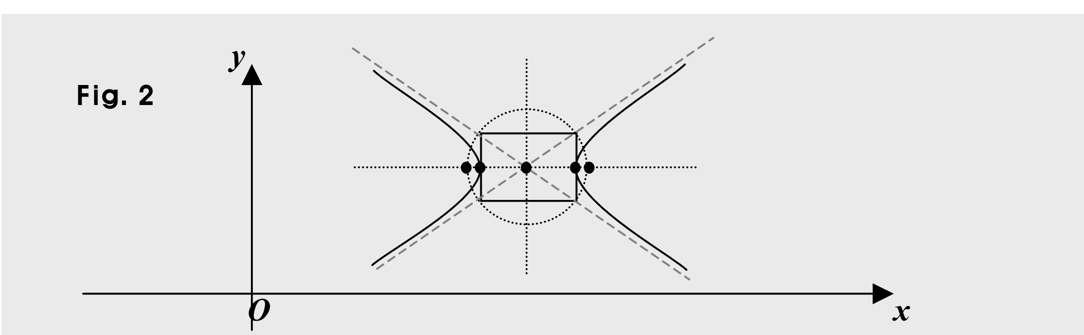
Also, like an ellipse, a hyperbola has two vertices. Each branch has one vertex. And the two vertices are the endpoints of the transverse axis, and thus, are in the transverse line.

So the length of the transverse (major) axis is the distance between the two vertices, and the transverse line has the vertices, together with the foci and the center.

Next, a hyperbola has a focal distance, too, which is, of course, the distance from either focus to the center, which is thus, the midpoint between the foci.

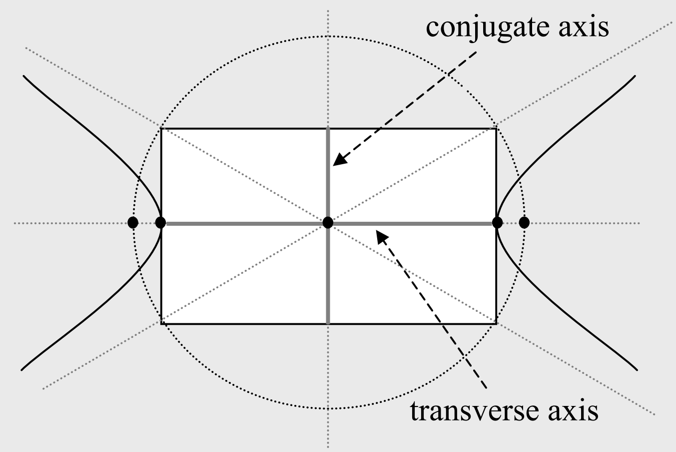
Where then are the two main axes, that is, the transverse axis and the conjugate axis?

A hyperbola is said to have an invisible rectangle.
And it can be a square, too. Anyway, taking the length of the diagonal of the rectangle, we get twice the focal distance. So half the diagonal is the focal distance, too.
And making the rectangle visible, we can put it the way as follows.



We can see in the figure, the rectangle is tangent to both branches at the vertices of the hyperbola, and its diagonal is the diameter of the circle that passes through all the vertices of the rectangle. The circle is said to be circumscribed about the rectangle. And such a circle is called a circumcircle. So the diagonal is the diameter of the circumcircle.

Fig. 3



And the diameter is the distance between the two foci.
So half the diameter, that is, half the diagonal is the focal distance.
In other words, the radius is the focal distance.

Then, the transverse axis is the line segment connecting the vertices of the hyperbola, that is, the length of the rectangle. And the conjugate axis is the width of the rectangle.

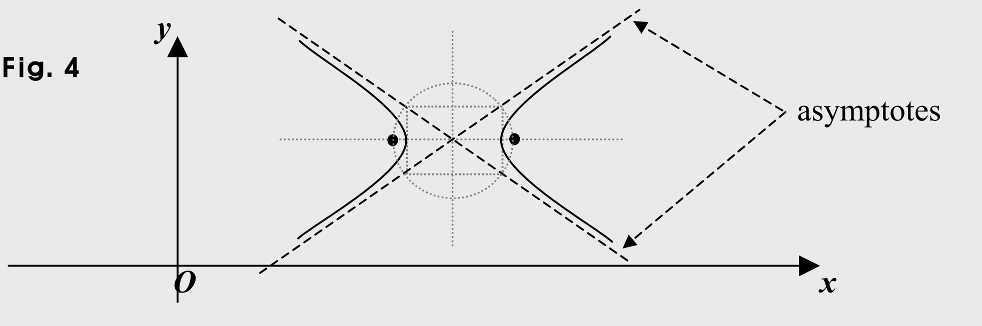
Also, half the length of the rectangle is called the semi major axis, and half the width is called the semi minor axis.

In short, half the transverse is the semi major, and half the conjugate is the semi minor.

So the rectangle has most of the information (specifications) of the hyperbola that has the rectangle.
Thus, specifying the dimensions of the rectangle, together with the center, we can describe specifically (that is, define) the hyperbola.

So understanding and using the rectangle is important.
Next, unlike other conics, a hyperbola has two special lines, called asymptotes.

Fig. 4



The two dashed lines above are the asymptotes, and are crossing each other at the center. So each diagonal of the rectangle is a part of each asymptote. What is an asymptote?

Suppose a point is moving along a branch of a hyperbola.
Then, there is a line that the point approaches as the point gets farther away from the vertex. No matter how far the point may be however, from the vertex, the point will never meet the line. And we call the line an asymptote, which a dashed line above.

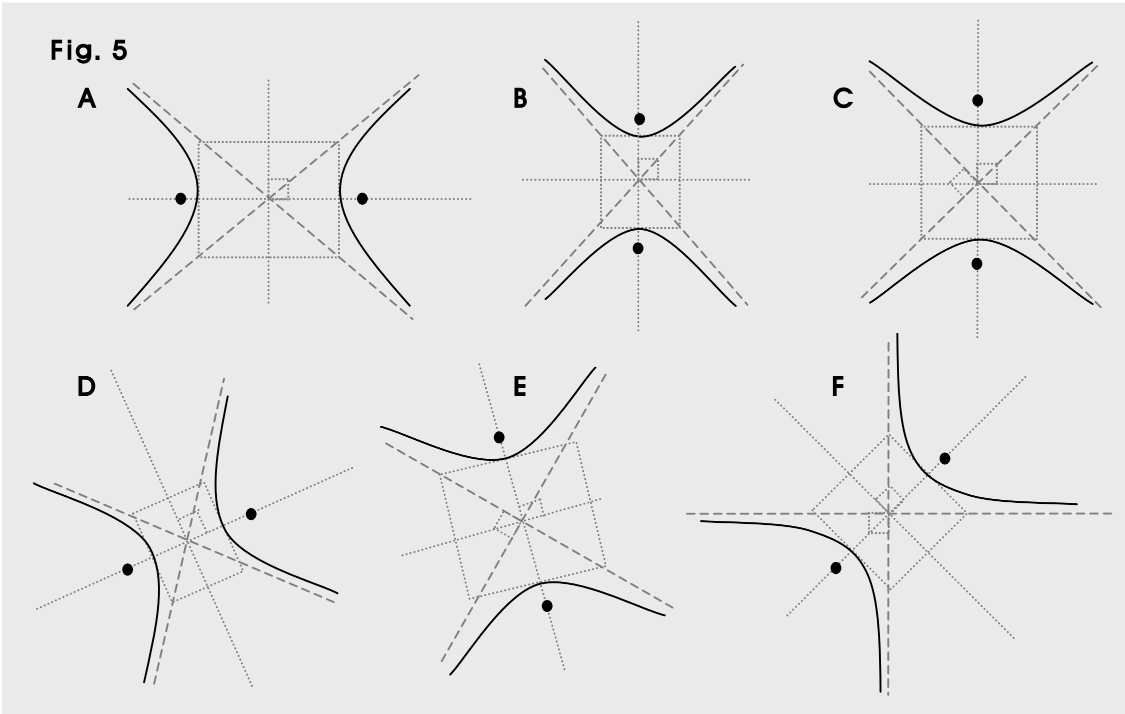
So as the point gets farther and farther away from the vertex, the distance between the point and the asymptote gets smaller and smaller, but won't be 0. In other words, the distance can be as good as 0, but cannot be 0. So the point doesn't meet the asymptote.

And a hyperbola has two asymptotes, and the two meet at the center, but not necessarily at a right angle, 90^{o}. Thus, the asymptotes do not have to be perpendicular to each other. So they can meet at an angle other than 90^{o}.

And as in the case of ellipses, **in this book**, if the asymptotes are perpendicular to each other, the hyperbola is said to be rectangular. So it is called a rectangular hyperbola. Note however, in other books, it may not be called that way.

If a hyperbola is rectangular, the invisible rectangle is a square, because the asymptotes are perpendicular to each other.

Also, in this book, if the transverse axis is horizontal, that is, parallel to the $\boldsymbol{x}$-axis, it is called a horizontal hyperbola, and if the axis is vertical, that is, parallel to the $\boldsymbol{y}$-axis, it is called a vertical hyperbola. In other books though, it may not be called that way.



In the figure above, **A** is a horizontal hyperbola, **B** is a vertical hyperbola, **C** is a vertical rectangular hyperbola, and **E** and **F** are rectangular hyperbolas. And **D** is just called a nonstandard hyperbola. If a hyperbola is standard, it is either horizontal or vertical. So **F** is a nonstandard hyperbola, too. And **A**, **B**, and **C** are standard hyperbolas.

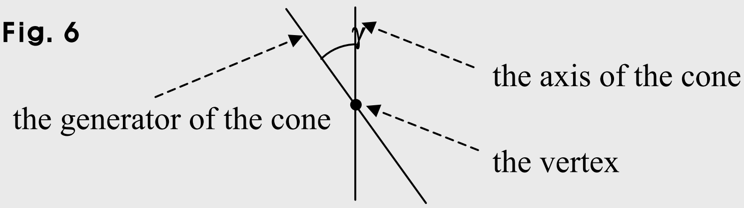
In high school math or basic courses in university math, we usually work with standard ones as **A**, **B**, and **C** above, or if it is nonstandard, it is probably a rectangular hyperbola where asymptotes are parallel or perpendicular to the coordinate axes as **F** above. So in high school math, we don't normally use nonstandard hyperbolas as **D** and **E** above. How then can we get a hyperbola?

As shown in the figure below, if cutting a right cone with a plane, we can get a cross section called a hyperbola. So it's called a conic section, just called a conic, for short. A right cone is a cone, where the line connecting the vertex of the cone and the center of the base is perpendicular to the base, and thus, makes a right angle (90^o) with the base.

And the line stated above is called the axis of the cone. And in the study of conics, such a right cone is said to be made or generated the way as follows.

First, take a line as the axis of the cone to be made.
Next, make another line meet the axis of the cone at a point at an angle γ, which is an acute angle, that is, $\mathbf{0 < \gamma < 90^o}$.

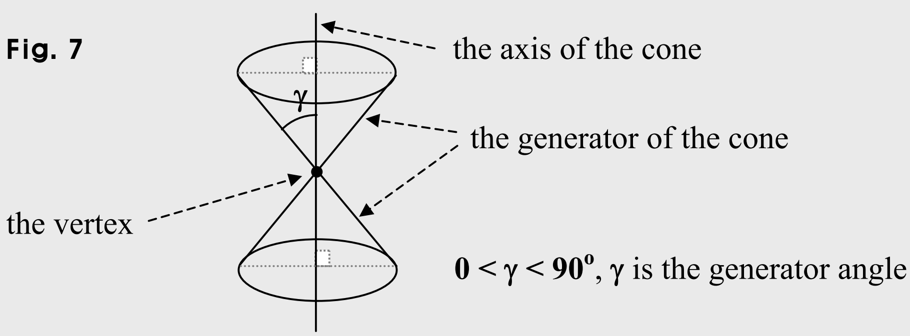
Fig. 6



Then, we call the other line the *generator of the cone*, and the point where the other line, that is, the generator of the cone meets the axis of the cone is the vertex of the cone.

So the vertex of the cone is the point where the generator meets the axis of the cone.
And in the figure above, the angle γ is called the *generator angle*.
Then, rotating the generator around the axis fixing the generator at the vertex, we get a double-cone as shown below.

Fig. 7

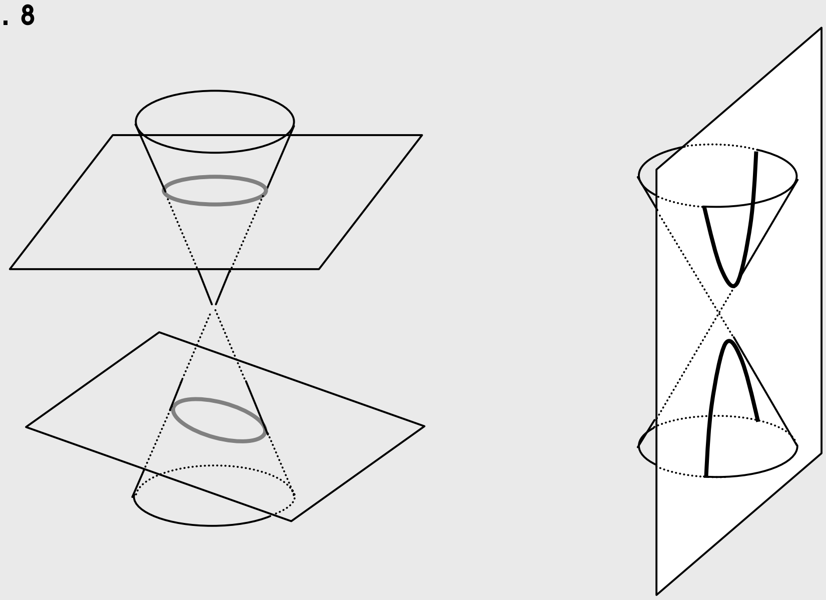


And note that the generator and the axis are lines, and lines have infinite lengths, so the lengths of the cones are infinite, too. So we cannot show them all. Showing thus, such a cone or a curve in math, we just show some part of it.

How then can we get a hyperbola cutting the cone with a plane?

The cross section in black below is a pair of curves, and is a hyperbola. So if α is the angle between the plane and the axis of the cone, and $\mathbf{0} \leq \alpha < \gamma$, and if the plane does not include the vertex, the cross section is a pair of curves, and is a conic called a hyperbola.

Fig. 8

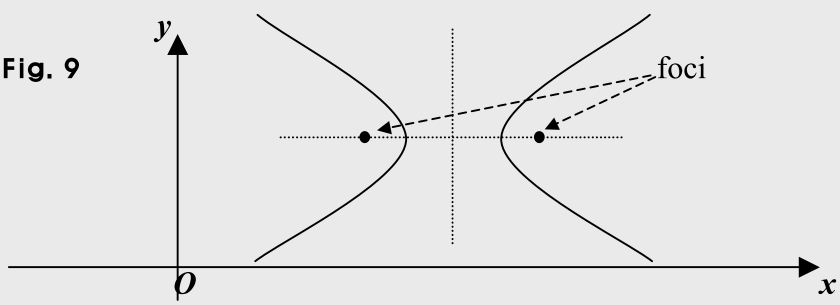


And a sketchy definition of a hyperbola can be as follows.

Suppose we designate two particular points in the ***x-y*** plane as shown in the figure below.

Then, a hyperbola is a set of points, from each of which, the difference between the two distances to the two particular points is constant, that is, the same. And the designated two points are called the foci of the hyperbola.

Fig. 9



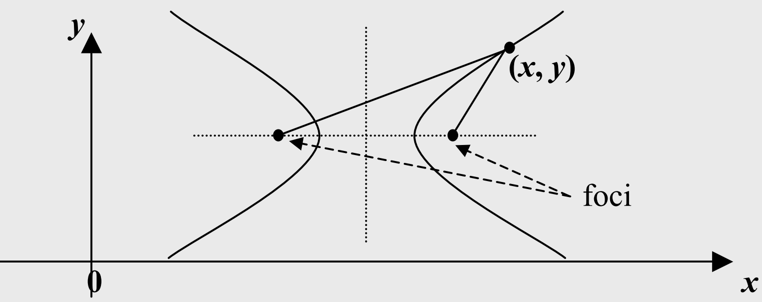
Also, we can describe (define) a hyperbola the way as follows, too:

Suppose a point is moving along a curve in the ***x-y*** plane, and the difference between the two distances from the moving point to two particular points is constant.

Then, the two particular points are called the foci, and the curve is a hyperbola.

If therefore, a point **(x, y)** is in a hyperbola, no matter where the point **(x, y)** may be, the difference between the two distances from the point **(x, y)** to the two foci is constant.

Fig. A



So using the fact above, we can actually make a hyperbola the way as follows.

First, set the two foci and another point by fastening three nails or thumb tacks to a panel as shown below. Then, make a loop using a string and attach to the loop a spring coil or a rubber band. Then, put the loop on the panel so that it wraps around the nails as below.

Fig. B

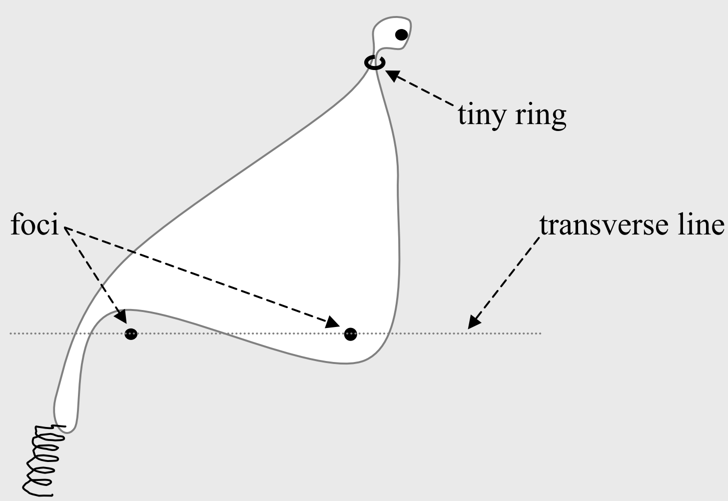
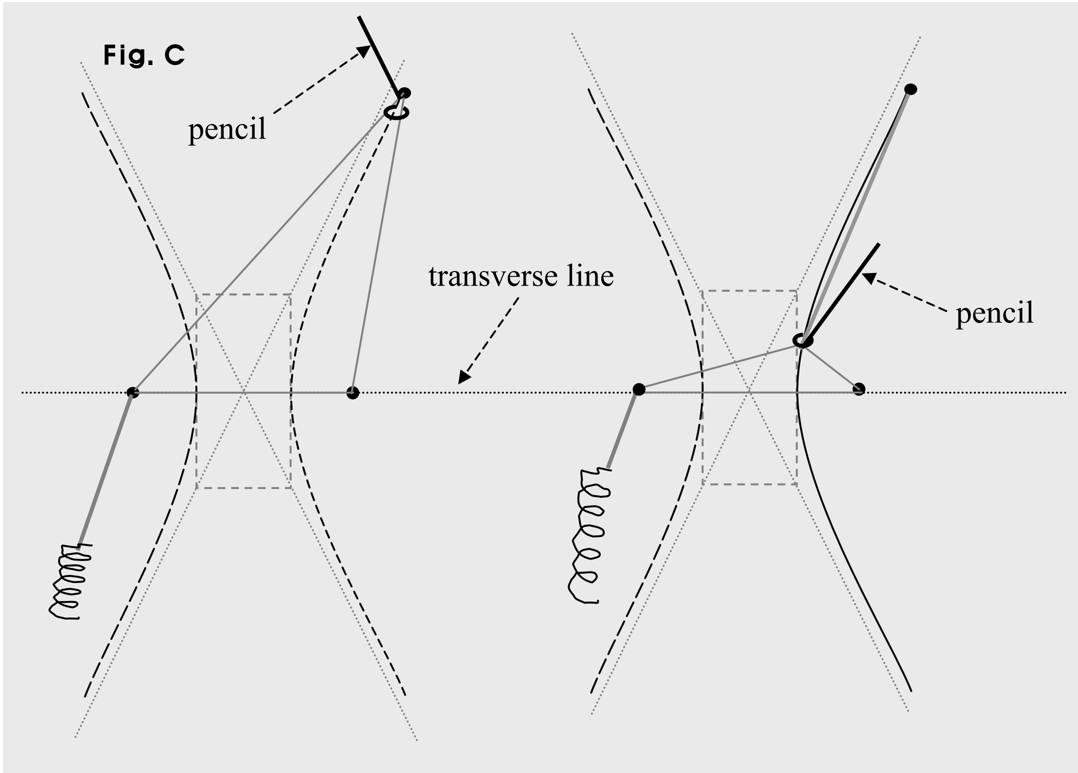
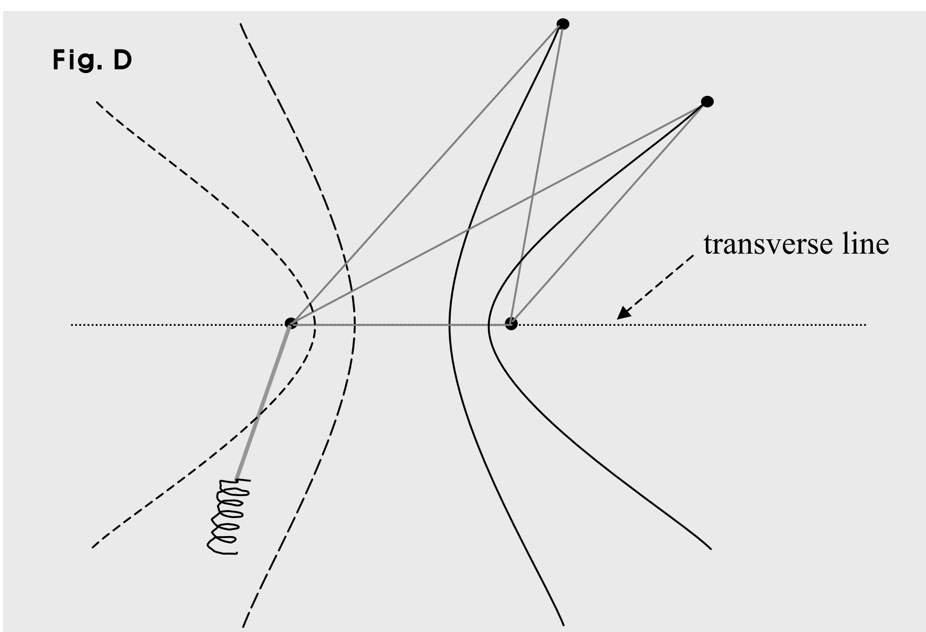


Fig. C



And as shown below, changing the position of the beginning point, we can change the location of the vertex and the bending in the branch. The more it bends, the closer the vertex is to the focus.

Fig. D



So we can see in the figure above that if a point is in the hyperbola, the difference between the two distances from the point to the foci is constant, that is, the same, because the length of the loop is constant.

How then, can we explain or describe a particular hyperbola?

Normally, we put a hyperbola in the ***x-y*** plain, and the hyperbola is standard, so the transverse axis is parallel to a coordinate axis. And we can describe such a hyperbola specifying the information as follows.

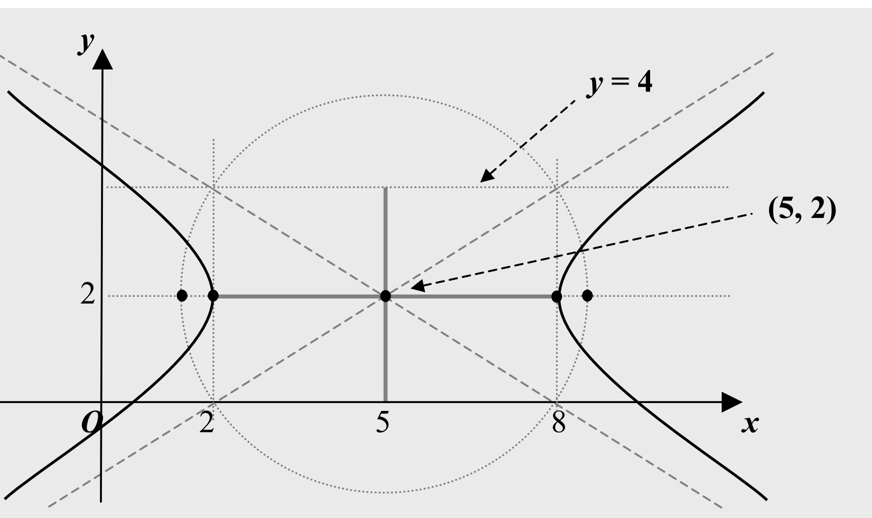
The center of the hyperbola and the lengths of the transverse and conjugate axes, along with the orientation, which indicates if the hyperbola is horizontal or vertical

If the transverse axis is horizontal, thus, parallel to the ***x***-axis, the hyperbola is horizontal, and if the axis is vertical, thus, perpendicular to the ***x***-axis, the hyperbola is vertical.

(Note that though, it is the case in this book, and it may not be the case in other books. That is to say that in other books, hyperbolas may not be classified the way above, and thus, may not be said to be horizontal or vertical.)

Anyway, for instance, if the center is (5, 2), and the transverse axe is 6, and is parallel to the ***x***-axis, and the conjugate axis is 4, the hyperbola is horizontal, and is as follows.

Fig. E

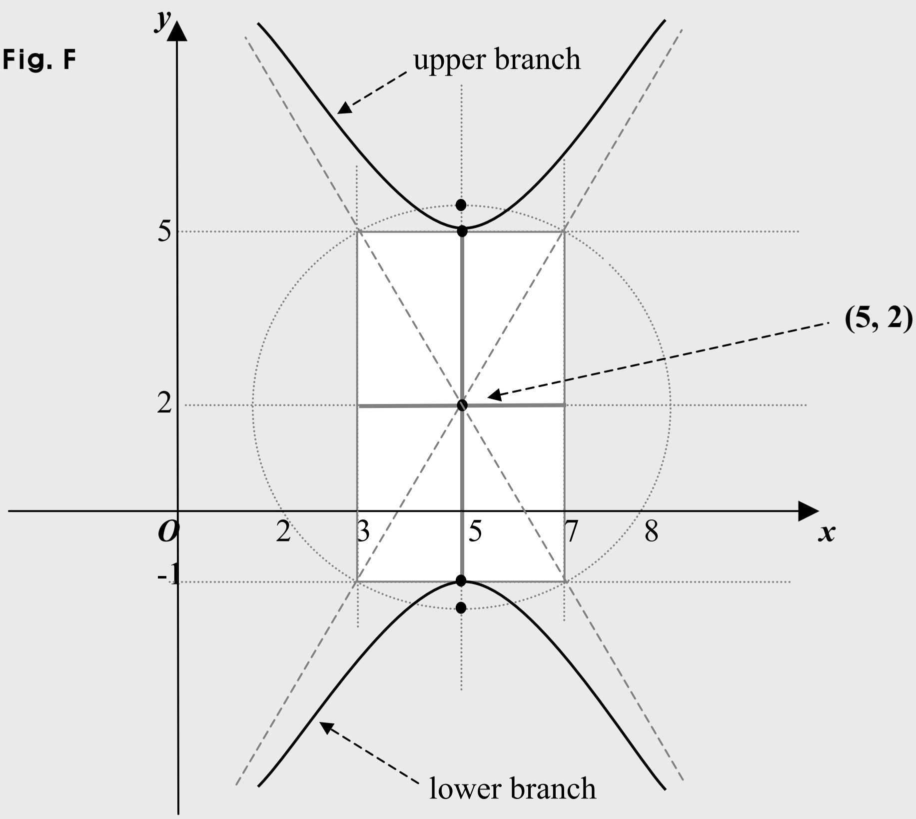


And showing a hyperbola, we often show the asymptotes, too.

And in fact, if specifying the dimensions of the rectangle, that is, the length and the width, along with the center and orientation, we can recognize a particular hyperbola.

So if saying that the rectangle is 6 by 4, centered at (5, 2), and horizontal, we mean the hyperbola shown above.
And if saying that the rectangle is 4 by 6, centered at (5, 2), and vertical, we mean the hyperbola shown below.

Fig. F



In math though, making a hyperbola, we define it. So defining a hyperbola particular, we make the particular hyperbola. And defining a hyperbola, we use an equation indicating the hyperbola. So we can define the particular hyperbola producing the equation of it.

Assuming thus, ***H*** is the horizontal hyperbola above, we can define ***H*** the way as follows.

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

Notice that if a hyperbola is horizontal, the sign of x^2 is positive, and that the denominator of the coefficient of x^2 is the square of half the transverse axis.

And assuming V is the vertical hyperbola above, we can define V the way as follows.

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

Notice that if a hyperbola is vertical, the sign of y^2 is positive, and that the denominator of the coefficient of y^2 is the square of half the transverse axis.

And each equation above is in the standard form, and thus, is a standard equation of a hyperbola.
Putting a standard equation in the general form, we just expand (or simplify) it.
So expanding the equation of the hyperbola H, we get

$4x^2 - 9y^2 - 40x + 36y + 28 = 0$, which is a general equation of a hyperbola.

And in general, putting a horizontal hyperbola in the standard form, we can put it the way as follows.

$$\frac{(x-u)^2}{a^2}-\frac{(y-v)^2}{b^2}=1\text{, where } a \text{ and } b > 0\text{. Then, the transverse axis is } 2a.$$

And in the case of a vertical hyperbola, we can put it the way as follows.

$$\frac{(y-v)^2}{b^2}-\frac{(x-u)^2}{a^2}=1\text{, where } a \text{ and } b > 0\text{. Then, the transverse axis is } 2b.$$

So if we define a particular hyperbola placed in the x-y plane, we want to specify the center and the orientation, together with the transverse and conjugate axes (or the semis).
What then is the center?

If the equation is either of the two above, the center of the hyperbola is a point (u, v).
Note that though, the standard equations above can indicate hyperbolas standard only.

Is there any equation then, that can indicate a hyperbola standard or not?

Putting a hyperbola in such an equation, we can put it the way as follows.

$$ax^2 + by^2 + cxy + ux + vy + w = 0, \text{ where } ab < 0, \text{ and } 4ab < c^2$$

Note that $ab < 0$ is saying that a and b have opposite signs. And of course, in the equation above, a, b, c, u, v, and w are constant. Note however, for some values of the constants, the equation does not indicate a hyperbola, even if $ab < 0$, and $4ab < c^2$.

What then about an equation that is in the general form and can indicate a hyperbola standard only?

Putting a hyperbola in such an equation, we can put it the way as follows.

$$ax^2 + by^2 + ux + vy + w = 0, \text{ where } ab < 0$$

So the equation above indicates a hyperbola with the transverse axis parallel to a coordinate axis. And notice that it does not have the xy-term, that is, cxy in the equation shown earlier. Note however, for some values of the constants, the equation does not indicate a hyperbola, even if $ab < 0$.

And we can give a name to an equation, too. For instance, assuming H is the equation above, we can put H the way as follows.

$$H(x, y) = ax^2 + by^2 + ux + vy + w = 0, \text{ where } ab < 0.$$

What then about the standard equation?

Assuming S is the equation $\frac{(x-u)^2}{a^2} - \frac{(y-v)^2}{b^2} = 1$, we can put S the way as follows.

$$S(x, y) = \frac{(x-u)^2}{a^2} - \frac{(y-v)^2}{b^2} - 1 = 0, \text{ where } a \text{ and } b > 0.$$

And assuming U is the equation $\frac{(y-v)^2}{b^2} - \frac{(x-u)^2}{a^2} = 1$, we can put U the way below

$$U(x, y) = \frac{(y-v)^2}{b^2} - \frac{(x-u)^2}{a^2} - 1 = 0, \text{ where } a \text{ and } b > 0.$$

What if $a = b$ though, in the equation above?

Then, the equation still indicates a hyperbola. What hyperbola then, is it?

It is the hyperbola where the both main axes are a (or b), and the center is at (u, v).

Setting $b = a$ in S, we get $\dfrac{(x-u)^2}{a^2} - \dfrac{(y-v)^2}{b^2} - 1 = 0 \Rightarrow \dfrac{(x-u)^2}{a^2} - \dfrac{(y-v)^2}{a^2} = 1$

So the equation indicates a hyperbola where the transverse axis is $2a$, which is also, the conjugate axis, and the center is (u, v). And of course, the hyperbola is horizontal.

And notice that it is not the case the standard equation of a hyperbola above is defined for x real and y real. What then is the domain and the range?

The domain is $x \leq -a + u$ or $x \geq a + u$, and the range is a set of all real numbers.
So more specifically, we can put the standard equation of a hyperbola the way as follows.

$\dfrac{(x-u)^2}{a^2} - \dfrac{(y-v)^2}{b^2} = 1$, where a and $b > 0$, and $x \leq -a + u$ or $x \geq a + u$

Also, like an ellipse, a hyperbola has an eccentricity.
Unlike an ellipse however, the eccentricity is bigger than 1. What is an eccentricity?

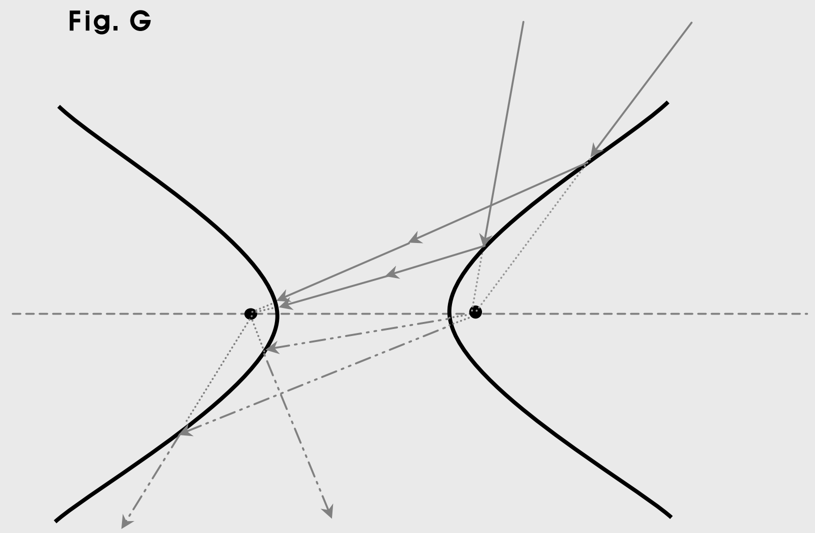
An eccentricity is a ratio of a focal distance to a semi major axis, so in the case of a hyperbola, the eccentricity is the focal distance over half the transverse axis.

So of a hyperbola, the eccentricity specifies the degree, to which the branches bend.
Thus, it can tell us how flat the branches are in a hyperbola or how sharp the vertices are.
An eccentricity is denoted by e, and we have $e > 1$ for a hyperbola.

If the eccentricity e is close to 1, the foci are close to the vertices, and the branches bend a lot, so the vertices are very sharp.
Therefore, the bigger the eccentricity, the flatter the branches.
So if e is very large, the two branches look like two parallel lines.
And a hyperbola has a physical property as follows.

For instance, a light beam directed to one focus of a hyperbolic mirror is reflected toward the other focus. Such a property can be used for the construction of a reflecting telescope, and the idea of a hyperbola is used in a navigation system.

Fig. G



How then can we get the foci and the eccentricity?

Suppose a hyperbola is $\dfrac{(x-u)^2}{a^2}-\dfrac{(y-v)^2}{b^2}=1$.

Then, the foci are $(u-c, v)$ and $(u+c, v)$, where $c^2 = b^2 + a^2$, and c is the focal distance, that is, half the diagonal of the rectangle (the radius of the circumcircle).

Then, the eccentricity $e = c/a$. And the hyperbola is horizontal.

Suppose however, a hyperbola is $\dfrac{(y-v)^2}{b^2}-\dfrac{(x-u)^2}{a^2}=1$.

Then, the foci are $(u, v+c)$ and $(u, v-c)$, where $c^2 = b^2 + a^2$, and c is the focal distance.

Then, the eccentricity $e = \dfrac{c}{b}$. And the hyperbola is vertical.

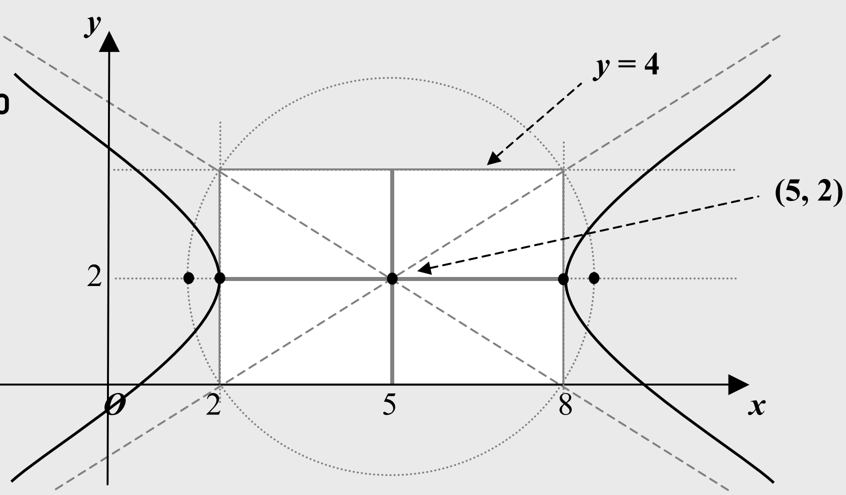
4.1. Equations for Hyperbolas 1

To begin with, putting a hyperbola in an equation, we can get $\frac{(x-5)^2}{3^2}-\frac{(y-2)^2}{2^2}=1$.

What hyperbola then is it?

It is a hyperbola horizontal, the center is at (5, 2), the major axis is 6, and the minor is 4. So the transverse axis is parallel to the x-axis, and is 6, and the hyperbola is as follows.

Fig. 0

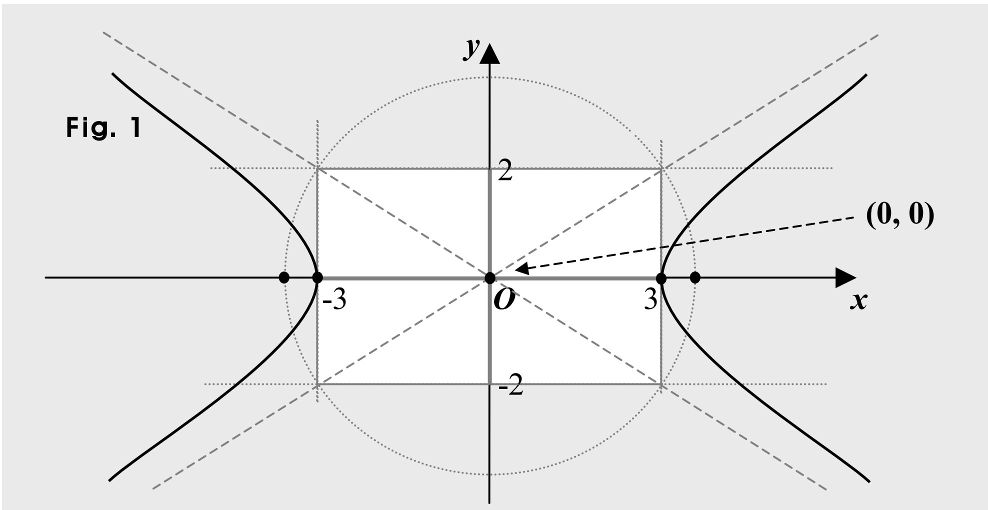


The transverse axis has vertices as its endpoints. And the minor axis is called the conjugate axis, so the conjugate axis is 4, and is the line segment having the center as the midpoint and parallel to the y-axis. What hyperbola then is the hyperbola as follows?

$$\frac{x^2}{3^2}-\frac{y^2}{2^2}=1$$

The hyperbola is horizontal, too. And the transverse axes is 6, and the conjugate is 4, also, but the center is different, and is at (0, 0), that is, at the origin.

Fig. 1

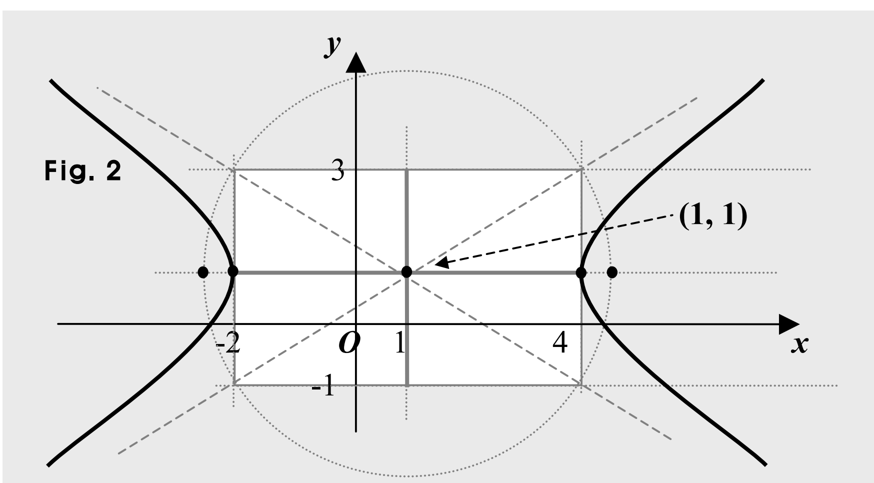


What if the center is at (1, 1), and the other information is the same?

Then, the hyperbola will be as follows $\frac{(x-1)^2}{3^2} - \frac{(y-1)^2}{2^2} = 1$.

And we can put the hyperbola above in a graph the way below.

Fig. 2



How then can we get such equations as above?

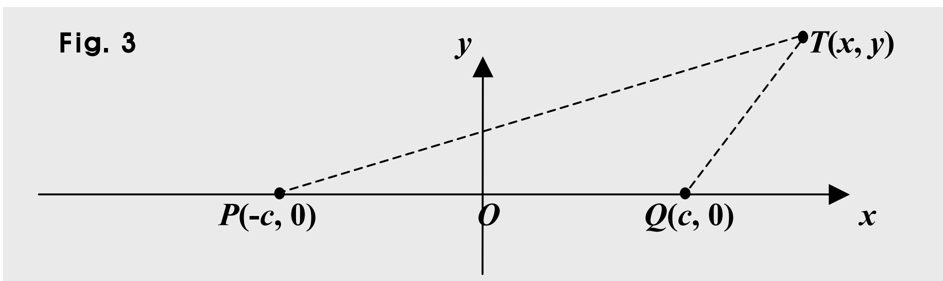
We know the fact that a hyperbola is a set of points, from each of which, the difference between the two distances to two points called the foci is constant. So?

So using the fact above, we can get the equation of a hyperbola.

Thus first, suppose $c > 0$, two points $P(-c, 0)$ and $Q(c, 0)$ are the two foci of a hyperbola called H, and a point $T(x, y)$ is an arbitrary point in the hyperbola H, that is, a random point representing all the points in H.

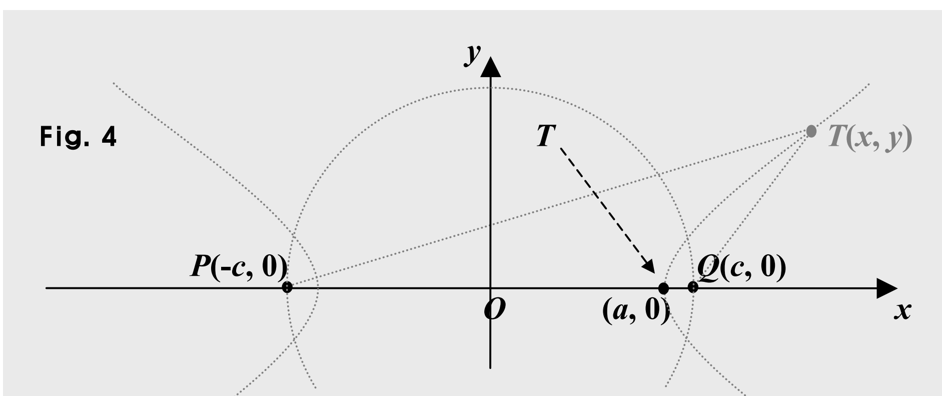
Then, the center is the origin, **(0, 0)**, and c is the focal distance, because the center is the midpoint between the foci. So we can put the three points P, Q, and T in the x-y plane the way as follows.

Fig. 3



Suppose next, the point T is moving along the hyperbola H, and is now at $(a, 0)$.
Then, we get this:

Fig. 4



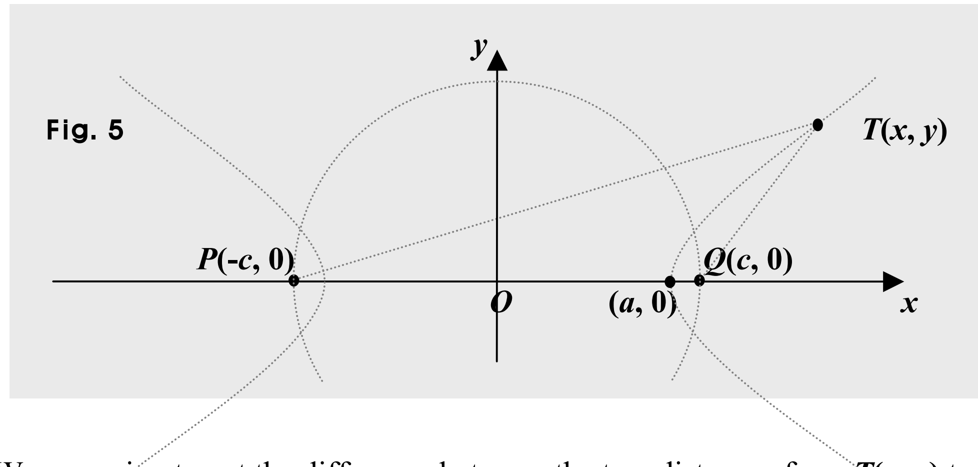
Then, the difference between the two distances from $T(x, y)$ to the two foci is as follows.

$\overline{TP} = a - (-c) = a + c$, and $\overline{TQ} = c - a$. So $\overline{TP} - \overline{TQ} = a + c - (c - a) = 2a$.
And we know the fact that a hyperbola is a set of points, from each of which, the difference between the two distances to two points called the foci is constant. So?

So no matter where the point T may be in the hyperbola H, the difference between the two distances from the point T to the two foci is always the same, and is in this case, $2a$.

Using thus, the fact above, we can get the equation of ***H***. How then, can we use the fact?

Fig. 5



We are going to get the difference between the two distances from ***T*(*x*, *y*)** to the two foci, and then, set the difference equal to **2*a***, because the difference is constant, that is, the same no matter where the point ***T*** may be in the hyperbola ***H***.

How then can we get the two distances from ***T*(*x*, *y*)** to the two foci?

We can use the distance formula, often called Pythagorean Theorem.
What then, do we get?

We get an equation expressed in terms of the coordinates of the point ***T***, that is, we get an equation in terms of ***x*** and ***y***. And we call such an equation a connective equation.

So in this case, the connective equation is the equation that connects the two variables used as the coordinates of the arbitrary point ***T*(*x*, *y*)** in the curve called the hyperbola ***H***.
How then do we call the equation?

The equation explains every point in the hyperbola ***H***, so it indicates the hyperbola ***H***.
The equation is thus, called the equation of the hyperbola ***H***.
So let's now get the equation.   How then can we apply the distance formula?

Assuming ***d*** is the distance between two points, **Δ*x*** is the difference in ***x***-coordinates, and **Δ*y*** is the difference in ***y***-coordinates, we can put the formula the way as follows.

$$d^2 = (\Delta x)^2 + (\Delta y)^2$$

So using the distance formula, and assuming $\boldsymbol{TP = p}$, and $\boldsymbol{TQ = q}$, we can get

$p^2 = \{x - (-c)\}^2 + (y - 0)^2 = (x + c)^2 + y^2$, and $q^2 = (x - c)^2 + (y - 0)^2 = (x - c)^2 + y^2$.

So we get $TP = \sqrt{(x+c)^2 + y^2}$, and $TQ = \sqrt{(x-c)^2 + y^2}$.

Thus, we get $TP - TQ = \sqrt{(x+c)^2 + y^2} - \sqrt{(x-c)^2 + y^2} = 2a$. Then, we get

$$\sqrt{(x+c)^2 + y^2} = 2a + \sqrt{(x-c)^2 + y^2} \Rightarrow (x+c)^2 + y^2 = (2a + \sqrt{(x-c)^2 + y^2})^2$$

$$\Rightarrow (x+c)^2 + y^2 = 4a^2 + 4a\sqrt{(x-c)^2 + y^2} + (x-c)^2 + y^2$$

$$\Rightarrow 4a\sqrt{(x-c)^2 + y^2} = (x+c)^2 + y^2 - (x-c)^2 - y^2 - 4a^2$$

$$\Rightarrow 4a\sqrt{(x-c)^2 + y^2} = x^2 + 2cx + c^2 - x^2 + 2cx - c^2 - 4a^2 = 4cx - 4a^2$$

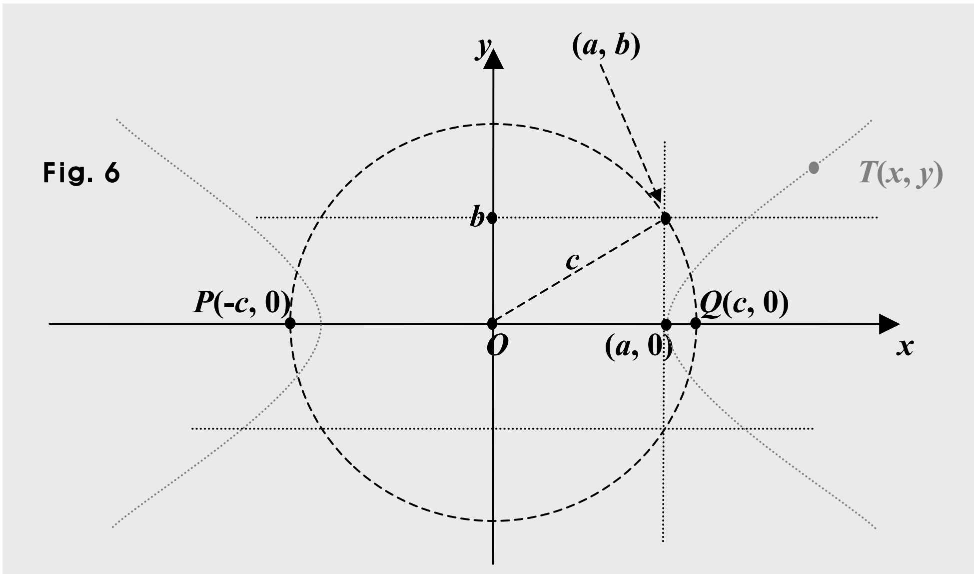
$$\Rightarrow a\sqrt{(x-c)^2 + y^2} = cx - a^2 \Rightarrow a^2\{(x-c)^2 + y^2\} = (cx - a^2)^2 = c^2x^2 - 2a^2cx + a^4$$

$$\Rightarrow a^2(x^2 - 2cx + c^2 + y^2) = a^2x^2 - 2a^2cx + a^2c^2 + a^2y^2 = c^2x^2 - 2a^2cx + a^4$$

$$\Rightarrow a^2x^2 - c^2x^2 + a^2c^2 + a^2y^2 = a^4 \Rightarrow (a^2 - c^2)x^2 + a^2y^2 = a^4 - a^2c^2 = a^2(a^2 - c^2)$$

$$\Rightarrow x^2 + \frac{a^2y^2}{a^2 - c^2} = a^2 \Rightarrow \frac{x^2}{a^2} + \frac{y^2}{a^2 - c^2} = 1.$$ What then about $a^2 - c^2$?

We can add to the graph a circle of radius c centered at the origin the way as follows.



Then, the circle's center is the center of the hyperbola. So?

So the radius is the focal distance of the hyperbola. What then?

Assuming $b > 0$, and putting a point (a, b) in the circle the way above, we can get a right triangle where the vertices are $(a, 0)$, (a, b), and the origin. So the three sides in the right triangle are a, b, and c, where c is the hypotenuse. So?

Applying thus, the distance formula, we can get this: $c^2 = b^2 + a^2 \Rightarrow b^2 = c^2 - a^2$.

And we now have $\dfrac{x^2}{a^2} + \dfrac{y^2}{a^2 - c^2} = 1$. So we get $\dfrac{x^2}{a^2} - \dfrac{y^2}{b^2} = 1$.

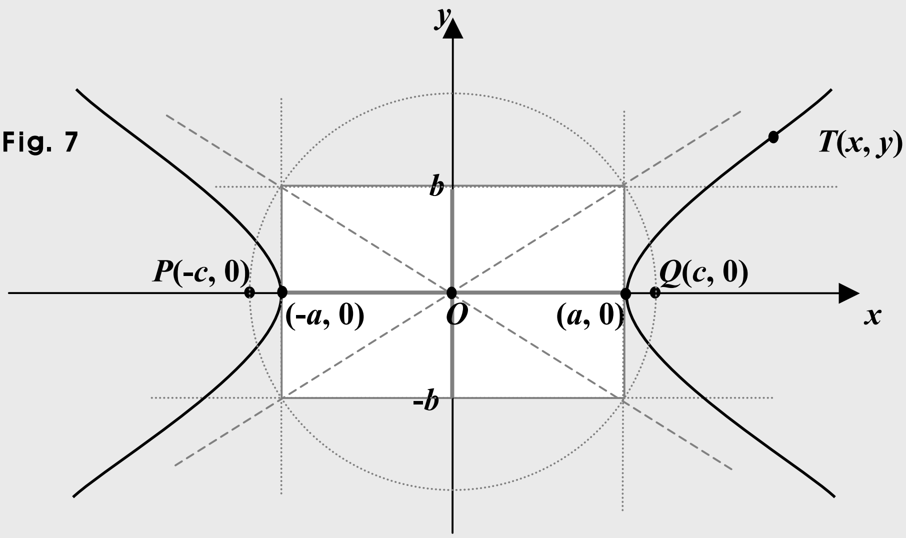
And the equation above is an equation expressed in terms of x and y. And we call such an equation a connective equation connecting x and y.

So in this case, the connective equation is the equation that connects the two variables used as the coordinates of the arbitrary point $T(x, y)$ in the curve called the hyperbola H.

Thus, the equation of the hyperbola H is $\dfrac{x^2}{a^2} - \dfrac{y^2}{b^2} = 1$.

And putting the hyperbola H in the x-y plane, we can get this:

Fig. 7



So we can now say that the hyperbola $\boldsymbol{H}$ is horizontal and centered at the origin, the major axis is $\mathbf{2}\boldsymbol{a}$, and is parallel to the $\boldsymbol{x}$-axis, and the minor axis is $\mathbf{2}\boldsymbol{b}$.

That is, the transverse axis is $\mathbf{2}\boldsymbol{a}$, and is horizontal, and the conjugate axis $\mathbf{2}\boldsymbol{b}$.

Then, putting the hyperbola $\boldsymbol{H}$ in the equation, we get $\dfrac{x^2}{a^2} - \dfrac{y^2}{b^2} = 1$.

What then are the domain and the range?

The domain is $x \le -a$ or $x \ge a$, that is, $|x| \ge a$. And the range is $y \in R$ where R is a set of all real numbers, that is, the range is a set of all real numbers.

What then about the domain and the range of this hyperbola: $\dfrac{x^2}{2^2} - \dfrac{y^2}{3^2} = 1$?

The domain is $x \le -2$ or $x \ge 2$, that is, $|x| \ge 2$.

And the range is $y \in R$ where R is a set of all real numbers, that is, the range is a set of all real numbers.

This time though, the transverse axis is 4, and thus, is smaller than the conjugate axis, which is 6.

What then, about the domain and the range of this hyperbola: $\dfrac{(x-1)^2}{2^2} - \dfrac{(y-2)^2}{3^2} = 1$?

Translating the hyperbola $\frac{x^2}{2^2} - \frac{y^2}{3^2} = 1$ by 1 along the x-axis, and by 2 along the y-axis, we get the hyperbola $\frac{(x-1)^2}{2^2} - \frac{(y-2)^2}{3^2} = 1$.

And the domain gets translated the way the hyperbola gets translated, too.
In this case however, the range does not actually get translated, because it is a set of all real numbers. So the range remains the same. And getting the new domain, we get

$x \le -2 + 1 = -1$ or $x \ge 2 + 1 = 3$, so the new domain is $x \le -1$ or $x \ge 3$.
What then, about the domain and the range of this hyperbola: $\frac{(x+1)^2}{2^2} - \frac{(y+2)^2}{3^2} = 1$?

Translating this time, the hyperbola $\frac{x^2}{2^2} - \frac{y^2}{3^2} = 1$ by -1 along the x-axis, and by -2 along the y-axis, we get this hyperbola: $\frac{(x+1)^2}{2^2} - \frac{(y+2)^2}{3^2} = 1$.

And the domain gets translated the way the hyperbola gets translated, too.
In this case, too, though, the range does not actually get translated, because it is a set of all real numbers. So the range remains the same. And getting the new domain, we get

$x \le -2 - 1 = -3$ or $x \ge 2 - 1 = 1$, so the new domain is $x \le -3$ or $x \ge 1$.
What then about this hyperbola: $\frac{(x-u)^2}{a^2} - \frac{(y-v)^2}{b^2} = 1$?

Translating the hyperbola $\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1$ by u along the x-axis, and by v along the y-axis, we get this hyperbola: $\frac{(x-u)^2}{a^2} - \frac{(y-v)^2}{b^2} = 1$.

And the domain gets translated the way the hyperbola gets translated, too.
We know that the domain of $\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1$ is $x \le -a$ or $x \ge a$.

So getting the domain of $\frac{(x-u)^2}{a^2} - \frac{(y-v)^2}{b^2} = 1$, we get $x \le -a + u$, or $x \ge a + u$.

In this case, too, though, the range does not actually get translated, because it is a set of all real numbers. So the range remains the same.

And note that unlike an ellipse, the major axis of a hyperbola, that is, the transverse axis does not have to be bigger than the minor axis, which is often called the conjugate axis.

So the transverse axis does not have to be larger than the conjugate axis.
In other words, the conjugate axis can be larger than the transverse axis, too.

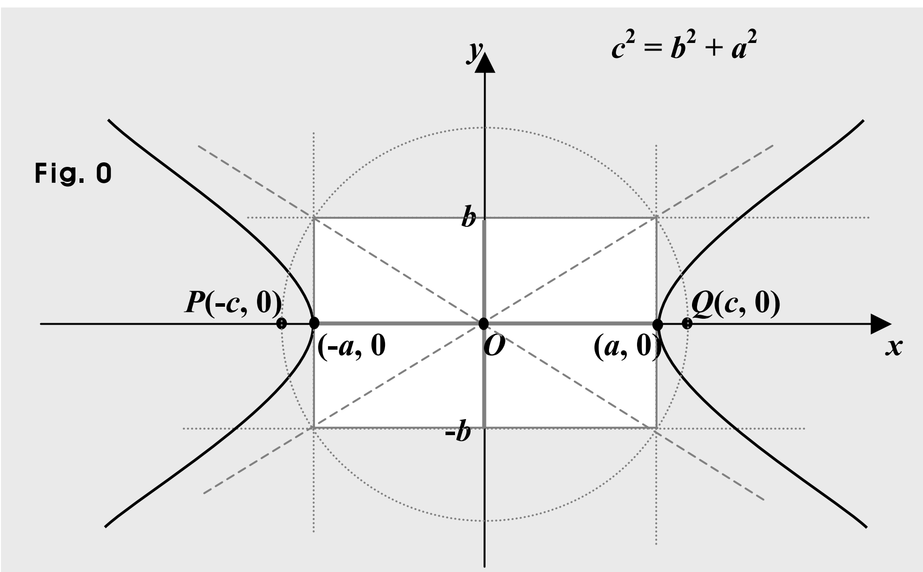
4.2. Equations for Hyperbolas 2

To begin with, putting a hyperbola in an equation, we can get $\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1$.

What hyperbola then is it?

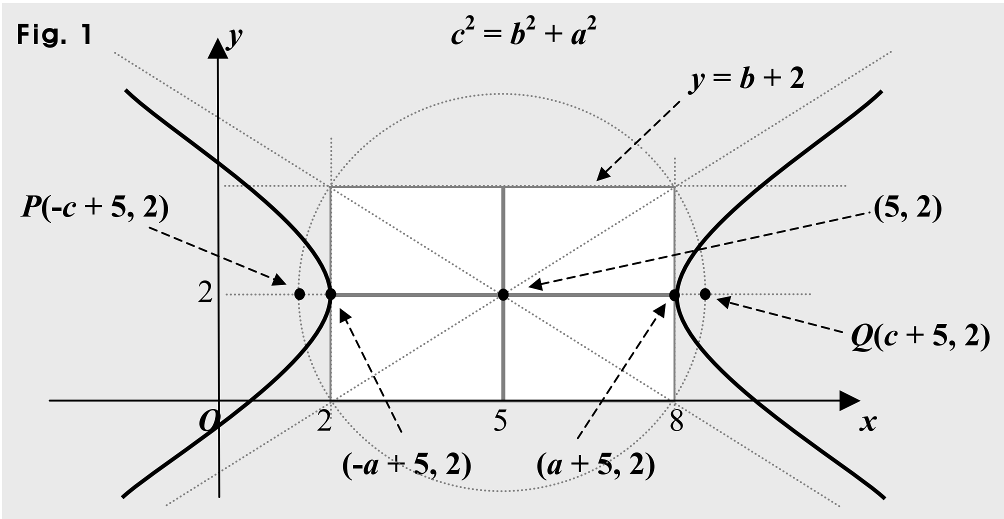
It is a hyperbola horizontal, where the center is at (0, 0), and its transverse axis is $2a$ and is parallel to the x-axis, and the conjugate axis is $2b$. And the hyperbola is as follows.

Fig. 0



What if the hyperbola above is centered at a point other than the origin?
For instance, moving the center to (5, 1), we get a new hyperbola, so we get a new equation. What then will the new equation be?

Assuming the hyperbola above is H, and the new hyperbola is G, we can get G translating H in the amount of 5 along the x-axis, and in the amount of 2 along the y-axis.



Then, the equation of the hyperbola G is $\dfrac{(x-5)^2}{a^2} - \dfrac{(y-2)^2}{b^2} = 1$.

(Note that we still get $c^2 = b^2 + a^2$, because a and b are constant, and so is c. And translations do not change shapes, so the focal distance c does not change either.)

And in general, we can put a standard hyperbola in an equation the way as follows.

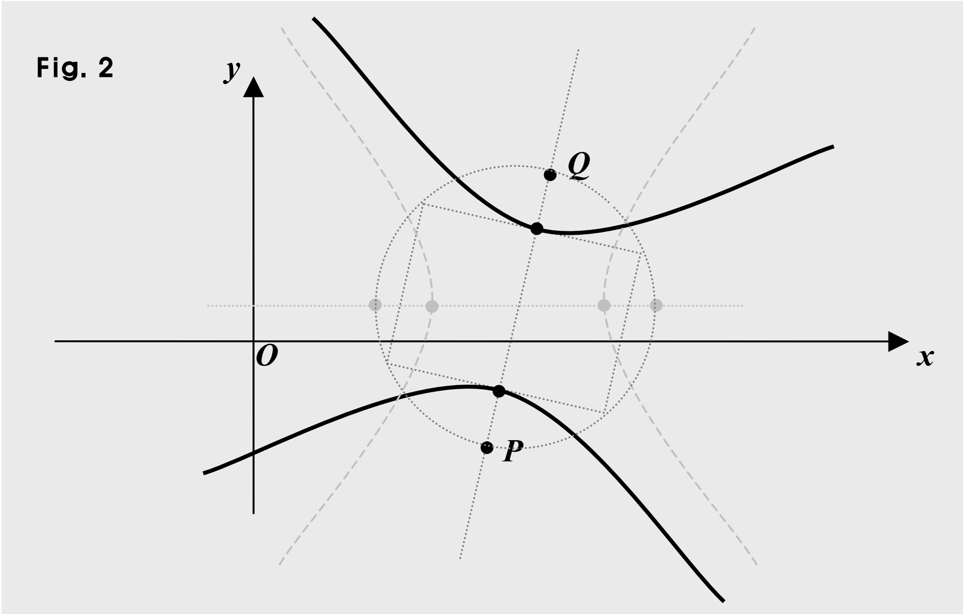
$$\frac{(x-u)^2}{a^2} - \frac{(y-v)^2}{b^2} = 1$$

Then, the center is at a point (u, v), the transverse axis is $2a$, and is horizontal, that is, parallel to the x-axis, and the conjugate axis is $2b$.

And the equation above is called the standard equation of a hyperbola.
(Note however, that the standard equation above can indicate a hyperbola standard only. If a hyperbola is standard, the transverse axis is parallel to a coordinate axis.)

What if a hyperbola is not standard, thus, tilted or slanted as the hyperbola as follows?

Fig. 2



Given the foci, you can derive the equation using the definition for hyperbolas, or can get the equation applying a transformation to a hyperbola standard. The transformation is covered in the section, Transformation by Turning, in the book, **GRAPH OPERATIONS**.

What then, about the ellipse described below?

The center is at (0, 0), the transverse axis is this time, $\mathbf{2}\boldsymbol{b}$, and is parallel to the $\boldsymbol{y}$-axis, and the conjugate axis is $\mathbf{2}\boldsymbol{a}$.

Then, the hyperbola is vertical, and the equation is $\dfrac{y^2}{b^2} - \dfrac{x^2}{a^2} = 1$.

So this time, the signs of the quadratic terms got swapped, and the equation looks quite opposite of the equation of a hyperbola horizontal.
What then, are the domain and the range?

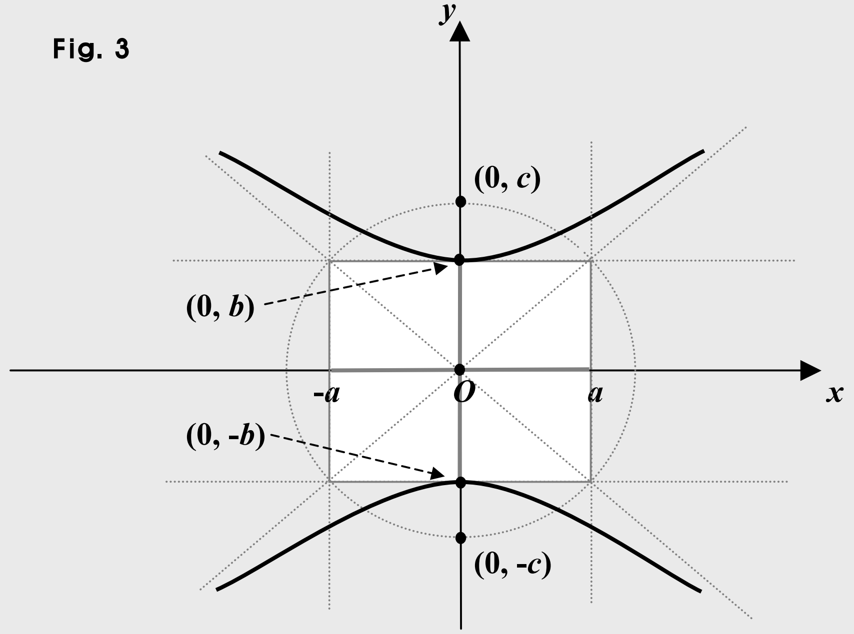
A domain is a set of all the $\boldsymbol{x}$-values, and a range is a set of all the $\boldsymbol{y}$-values.

So the domain is a set of all real numbers, and the range is $y \geq b$ or $y \leq -b$, i.e., $|y| \geq b$.
And we know the foci are in the transverse line, a part of which is the transverse axis, which is this time, parallel to the $\boldsymbol{y}$-axis, and is the line segment from one vertex **(0, *b*)** to the other vertex **(0, -*b*)**.

And the center is the origin, so if $\boldsymbol{c}$ is the focal distance, the foci are **(0, *c*)** and **(0, -*c*)**.

Assuming thus, the vertical hyperbola above is ***V***, we can put ***V*** in a graph the way as follows.

Fig. 3



How then can we get the equation of ***V***?

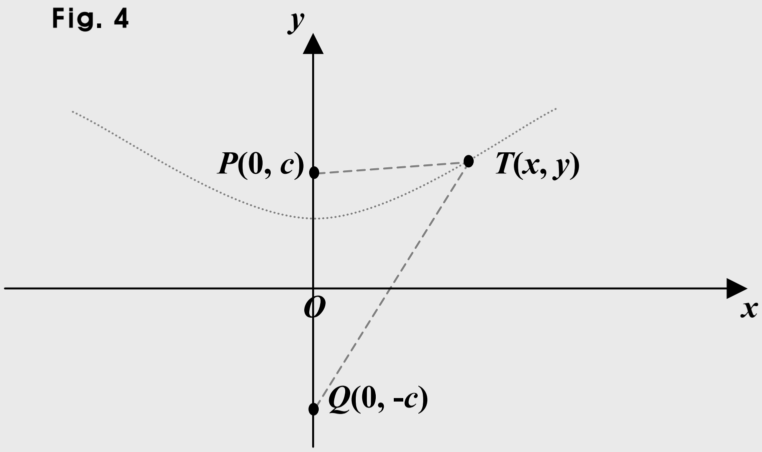
We know the definition for hyperbolas is saying that a hyperbola is a set of points, from each of which, the difference between the two distances to the two foci is constant.

Using thus, the fact above again, we can get the equation of a hyperbola vertical, too.

So to begin with, suppose $c > 0$, two points $P(0, c)$ and $Q(0, -c)$ are the two foci of a hyperbola called ***V***, and a point $T(x, y)$ is an arbitrary point in the hyperbola ***V***, that is, a random point representing all the points in ***V***.

Then, the center is the origin, that is, **(0, 0)**, and we can say that ***c*** is the focal distance, because the center is the midpoint between the two foci. So we can put the three points ***P***, ***Q***, and ***T*** in the ***x-y*** plane the way as follows.

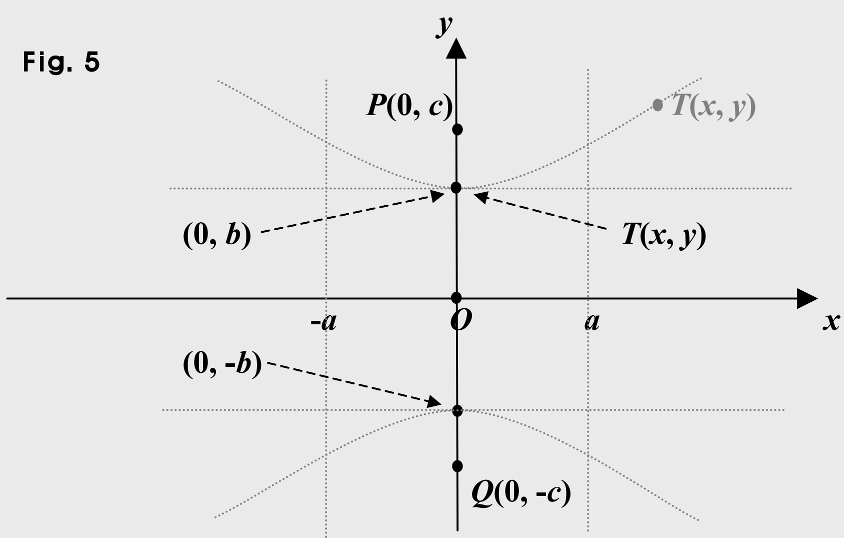
Fig. 4



Suppose next, the point ***T*** is moving now, along the hyperbola ***V***.
We know that the difference between two distances from each point in ***V*** to the two foci is constant, that is, the same. So this time, too, we are going to use the fact again.

Suppose thus, this time, that the point ***T*** is now at the point **(0, *b*)**.

Fig. 5



We can see that the length of ***TQ*** is $\boldsymbol{b - (-c) = b + c}$.

And the length of ***TP*** is $\boldsymbol{c - b}$. So we get $\boldsymbol{TQ - TP = (b + c) - (c - b) = 2b}$.

Therefore, no matter where the point T may be in the hyperbola V, the difference between two distances from the point T to the two foci is always $2b$, the transverse axis.

So we are going to get the difference between the two distances from $T(x, y)$ to the two foci, and then, set the difference equal to $2b$.

How then, can we get the two distances from $T(x, y)$ to the two foci?

We can use the distance formula, often called Pythagorean Theorem.
And using it, we can get an equation expressed in terms of x and y, that is, we get the connective equation between x and y, which are the coordinates of the arbitrary point $T(x, y)$ in the curve called the hyperbola V. How then do we call the equation?

The equation explains every point in the hyperbola V, so it indicates the hyperbola V.
The equation is thus, called the equation of the hyperbola V.
So let's now get the equation. How then can we apply the distance formula?

Assuming d is the distance between two points, Δx is the difference in x-coordinates, and Δy is the difference in y-coordinates, we can put the formula the way as follows.

$d^2 = (\Delta x)^2 + (\Delta y)^2$ And we have $P(0, c)$, $Q(0, -c)$, and $T(x, y)$.

So using the distance formula, and assuming $TP = p$, and $TQ = q$, we can get

$p^2 =(x - 0)^2 + (y - c)^2 = x^2 + (y - c)^2$, and $q^2 = (x - 0)^2 + \{y - (-c)\}^2 = x^2 + (y + c)^2$.

So we get $TP = \sqrt{x^2 + (y-c)^2}$, and $TQ = \sqrt{x^2 + (y+c)^2}$.

Thus, we get $TP - TQ = \sqrt{x^2 + (y-c)^2} - \sqrt{x^2 + (y+c)^2} = 2b$. Then, we get

$$\sqrt{x^2 + (y-c)^2} = 2b + \sqrt{x^2 + (y+c)^2} \Rightarrow x^2 + (y-c)^2 = (2b + \sqrt{x^2 + (y+c)^2})^2$$

$$\Rightarrow x^2 + (y-c)^2 = 4b^2 + 4b\sqrt{x^2 + (y+c)^2} + x^2 + (y+c)^2$$

$$\Rightarrow 4b\sqrt{x^2 + (y+c)^2} = (y-c)^2 - (y+c)^2 - 4b^2 = y^2 - 2cy + c^2 - y^2 - 2cy - c^2 - 4b^2$$

$$\Rightarrow 4b\sqrt{x^2+(y+c)^2} = -4b^2 - 4cy \Rightarrow b\sqrt{x^2+(y+c)^2} = -(b^2+cy)$$

$$\Rightarrow b^2\{x^2+(y+c)^2\} = (b^2+cy)^2 = b^4 + 2b^2cy + c^2y^2$$

$$\Rightarrow b^2(x^2+y^2+2cy+c^2) = b^2x^2 + b^2y^2 + 2b^2cy + b^2c^2 = a^4 + 2b^2cy + c^2y^2$$

$$\Rightarrow b^2x^2 + b^2y^2 + b^2c^2 - c^2y^2 = b^4 \Rightarrow b^2x^2 + (b^2-c^2)y^2 = b^4 - b^2c^2 = b^2(b^2-c^2)$$

$$\Rightarrow \frac{b^2x^2}{b^2-c^2} + y^2 = b^2 \Rightarrow \frac{x^2}{b^2-c^2} + \frac{y^2}{b^2} = 1.$$ What then about $b^2 - c^2$?

We can add to the graph a circle of radius c centered at the origin the way as follows.

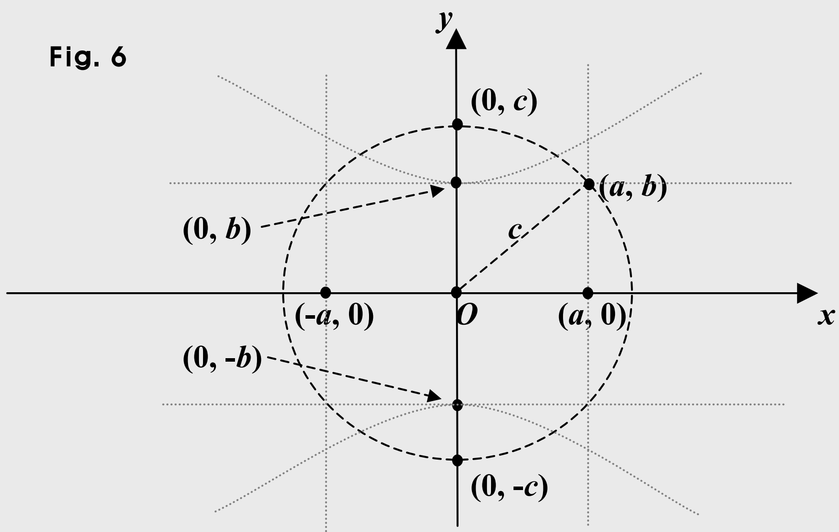


Fig. 6

Then, the circle's center is the center of the hyperbola.   So?

So the radius is the focal distance of the hyperbola.   What then?

Assuming $b > 0$, and putting a point (a, b) in the circle the way above, we can get a right triangle where the vertices are $(a, 0)$, (a, b), and the origin. So the three sides in the right triangle are a, b, and c, where c is the hypotenuse.   So?

Applying thus, the distance formula, we can get this $c^2 = b^2 + a^2 \Rightarrow -a^2 = b^2 - c^2$.

And we now have $\dfrac{x^2}{b^2 - c^2} + \dfrac{y^2}{b^2} = 1$. So we get $-\dfrac{x^2}{a^2} + \dfrac{y^2}{b^2} = 1 \Rightarrow \dfrac{y^2}{b^2} - \dfrac{x^2}{a^2} = 1$.

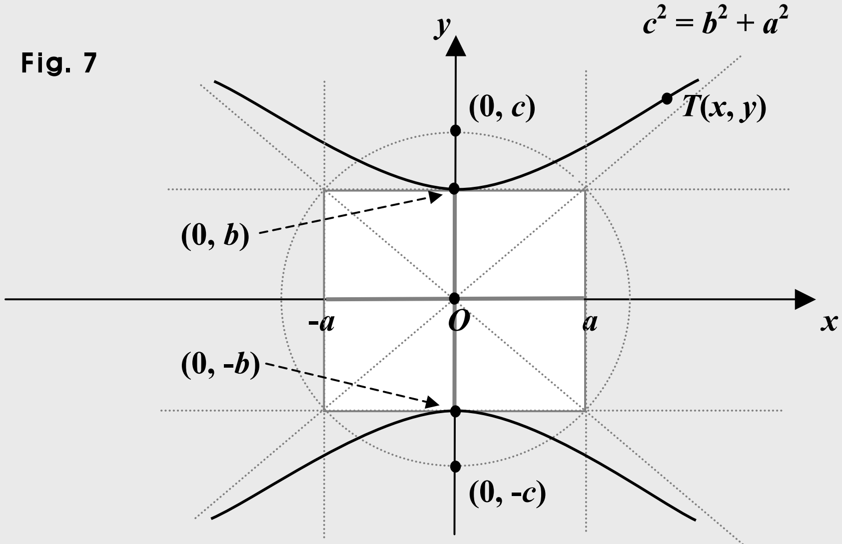
And the equation above is an equation expressed in terms of x and y. And we call such an equation a connective equation connecting x and y.

So in this case, the connective equation is the equation that connects the two variables used as the coordinates of the arbitrary point $T(x, y)$ in the curve called the hyperbola V.

Thus, the equation of the hyperbola V is $\dfrac{y^2}{b^2} - \dfrac{x^2}{a^2} = 1$.

And putting the hyperbola V in the x-y plane, we get

Fig. 7



So we can now say that the hyperbola V is vertical and centered at the origin, the major axis is $2b$, and is parallel to the y-axis, and the minor axis is $2a$.
That is, the transverse axis is $2b$, and is vertical, and the conjugate axis $2a$.

Then, putting the hyperbola V in the equation, we get $\dfrac{y^2}{b^2} - \dfrac{x^2}{a^2} = 1$.

What are then, the domain and the range?

The domain is $x \in R$ where R is a set of all real numbers, that is, the domain is a set of all real numbers. And the range is $y \leq -b$ or $y \geq b$, that is, $|y| \geq b$.

What then about the domain and the range of this hyperbola: $\frac{y^2}{2^2} - \frac{x^2}{3^2} = \mathbf{1}$?

The domain is $\boldsymbol{x} \in \boldsymbol{R}$ where $\boldsymbol{R}$ is a set of all real numbers, that is, the domain is a set of all real numbers. And the range is $\boldsymbol{y \le -2}$ or $\boldsymbol{y \ge 2}$, that is, $\boldsymbol{|y| \ge 2}$.

Note that the transverse axis is 4, and thus, is smaller than the conjugate axis, which is 6. So the transverse axis can be smaller than the conjugate axis.

What then about the domain and the range of this hyperbola: $\frac{(y-1)^2}{2^2} - \frac{(x-2)^2}{3^2} = \mathbf{1}$?

Translating the hyperbola $\frac{y^2}{2^2} - \frac{x^2}{3^2} = \mathbf{1}$ by 2 along the $\boldsymbol{x}$-axis, and by 1 along the $\boldsymbol{y}$-axis, we get the hyperbola $\frac{(y-1)^2}{2^2} - \frac{(x-2)^2}{3^2} = \mathbf{1}$.

And the range gets translated the way the hyperbola gets translated, too.
And getting the new range, we get
$\boldsymbol{y \le -2 + 1 = -1}$ or $\boldsymbol{y \ge 2 + 1 = 3}$, so the new range is $\boldsymbol{y \le -1}$ or $\boldsymbol{y \ge 3}$.

In this case however, the domain does not actually get translated, because it is a set of all real numbers. So the domain remains the same, and thus, is a set of all real numbers.

What then about the domain and the range of this hyperbola: $\frac{(y+1)^2}{2^2} - \frac{(x+2)^2}{3^2} = \mathbf{1}$?

Translating this time, the hyperbola $\frac{y^2}{2^2} - \frac{x^2}{3^2} = \mathbf{1}$ by -2 along the $\boldsymbol{x}$-axis, and by -1 along the $\boldsymbol{y}$-axis, we get this hyperbola: $\frac{(y+1)^2}{2^2} - \frac{(x+2)^2}{3^2} = \mathbf{1}$.

And the range gets translated the way the hyperbola gets translated, too. And getting the new range, we get $\boldsymbol{y \le -2 - 1 = -3}$ or $\boldsymbol{y \ge 2 - 1 = 1}$, so the new range is $\boldsymbol{y \le -3}$ or $\boldsymbol{y \ge 1}$.

In this case, too, though, the domain does not actually get translated, because it is a set of all real numbers. So the domain remains the same, thus, a set of all real numbers.

What then about this hyperbola: $\frac{(y-v)^2}{b^2}-\frac{(x-u)^2}{a^2}=1$?

Translating the hyperbola $\frac{y^2}{b^2}-\frac{x^2}{a^2}=1$ by u along the x-axis, and by v along the y-axis, we get this hyperbola: $\frac{(y-v)^2}{b^2}-\frac{(x-u)^2}{a^2}=1$.

And the range gets translated the way the hyperbola gets translated along the y-axis, too. We know that the range of $\frac{x^2}{a^2}-\frac{y^2}{b^2}=1$ is $y \le -b$ or $y \ge b$.

So getting the range of $\frac{(y-v)^2}{b^2}-\frac{(x-u)^2}{a^2}=1$, we get $y \le -b + v$, or $y \ge b + v$.

In this case, too, though, the domain does not actually get translated, because it is a set of all real numbers. So the range remains the same, and is a set of all real numbers.

(And note that unlike an ellipse, the major axis of a hyperbola, that is, the transverse axis does not have to be bigger than the minor axis, which is often called the conjugate axis. So the transverse axis does not have to be larger than the conjugate axis. That is, the conjugate axis can be larger than the transverse axis.)

So in sum, we have just covered two equations indicating hyperbolas in two kinds.

One is $\frac{x^2}{a^2}-\frac{y^2}{b^2}=1$, where $c^2 = b^2 + a^2$, where c is the focal distance.

The hyperbola above is horizontal, the center is at (0, 0), the transverse axis is $2a$, and is a part of (thus, parallel to) the x-axis, and the conjugate axis is $2b$. So the hyperbola is as follows.

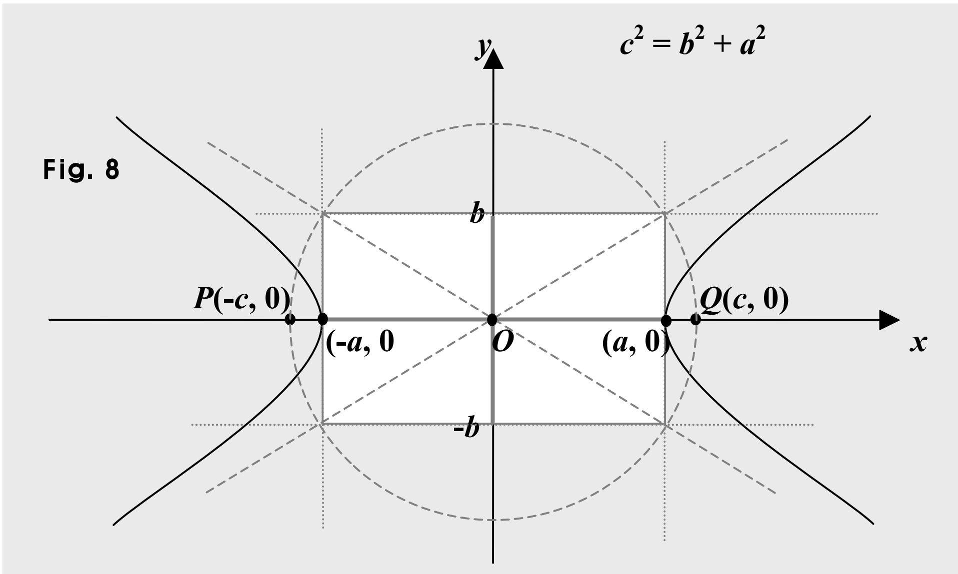


Fig. 8

And the other is $\frac{y^2}{b^2} - \frac{x^2}{a^2} = \mathbf{1}$, where $c^2 = b^2 + a^2$, and c is the focal distance.

The hyperbola above is vertical, the center is at (0, 0), the transverse axis is $\mathbf{2}b$, and is a part of (thus, parallel to) the y-axis, and the conjugate axis is $\mathbf{2}a$.

So the hyperbola is as follows.

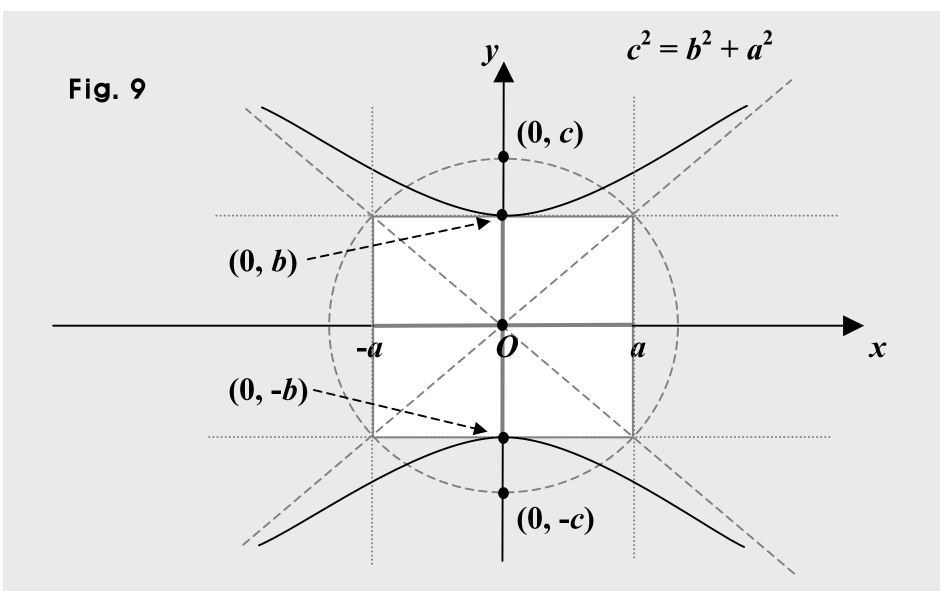
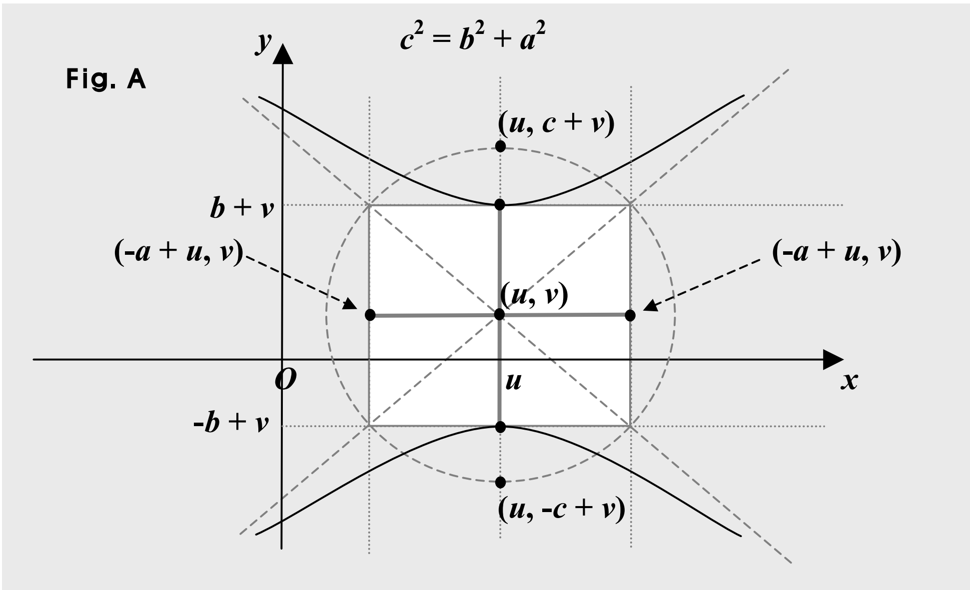


Fig. 9

What if the center is not at the origin as in the case as follows?

Fig. A



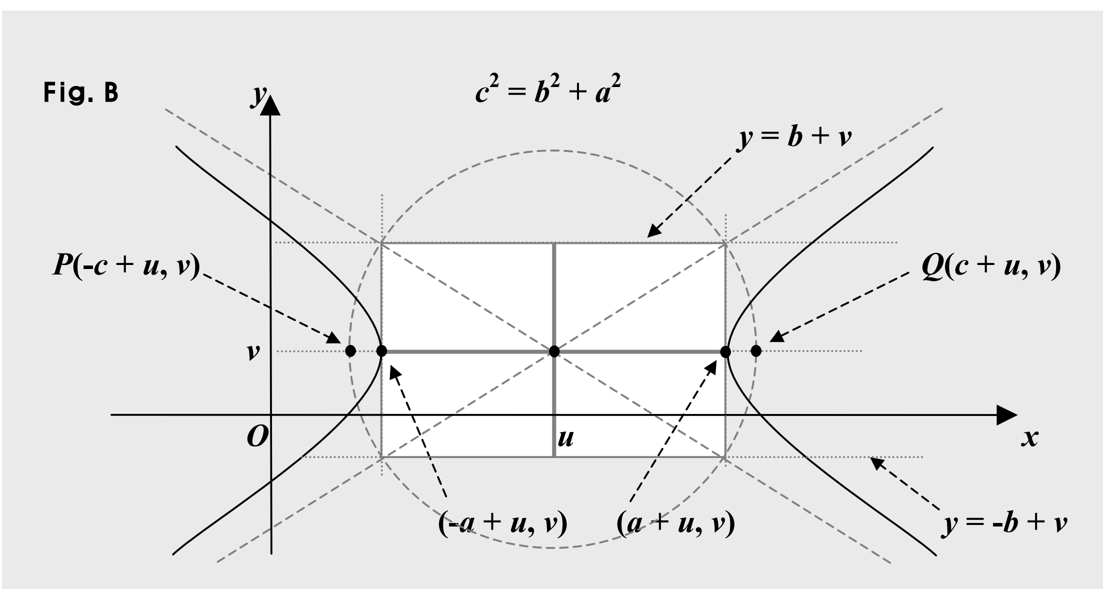
Then, the equation is $\dfrac{(y-v)^2}{b^2} - \dfrac{(x-u)^2}{a^2} = 1$, where $c^2 = b^2 + a^2$.

Note that we can get the hyperbola above translating a hyperbola in amount of u along the x-axis, and in the amount of v along the y-axis.

And the hyperbola to be translated is $\dfrac{y^2}{b^2} - \dfrac{x^2}{a^2} = 1$.

And in the case as follows,

Fig. B



The equation is $\dfrac{(x-u)^2}{a^2}-\dfrac{(y-v)^2}{b^2}=1$, where $c^2 = b^2 + a^2$.

Note that assuming K is the hyperbola above, we can get K translating a hyperbola in amount of u along the x-axis, and in the amount of v along the y-axis.

And the hyperbola to be translated is $\dfrac{x^2}{a^2}-\dfrac{y^2}{b^2}=1$.

Also, assuming L is the hyperbola above, we can get L translating the hyperbola K in amount of $-u$ along the x-axis, and in the amount of $-v$ along the y-axis.

4.3. Elements of a Hyperbola

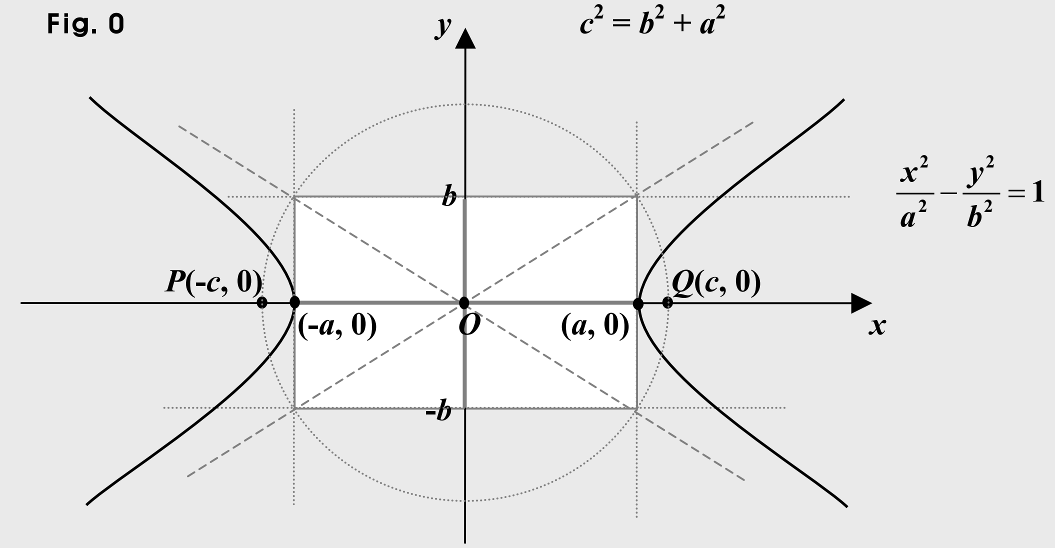
As in the case of an ellipse, working with a hyperbola, too, we often need to refer to some basic elements in it. And the elements are as follows.

center, foci, directrices, vertices, main axes, focal distance, eccentricity, and asymptotes.

So let's now take a close look at each of those elements.

To begin with, a hyperbola has two vertices, which are both ends of the transverse axis.

Fig. 0



So in the hyperbola above, the two vertices are **(-*a*, 0)** and **(*a*, 0)**. What about ***P*** and ***Q***?

The two points ***P*(-*c*, 0)** and ***Q*(*c*, 0)** are the foci, and the center is **(0, 0)**.

And the transverse axis is a part of a line called the transverse line, and we can notice that the transverse line has all the five points: the vertices, the foci, and the center.

So we can also notice that if a hyperbola is horizontal, the ***y***-coordinates are the same at all those five points, which share thus, the same ***y***-coordinate.

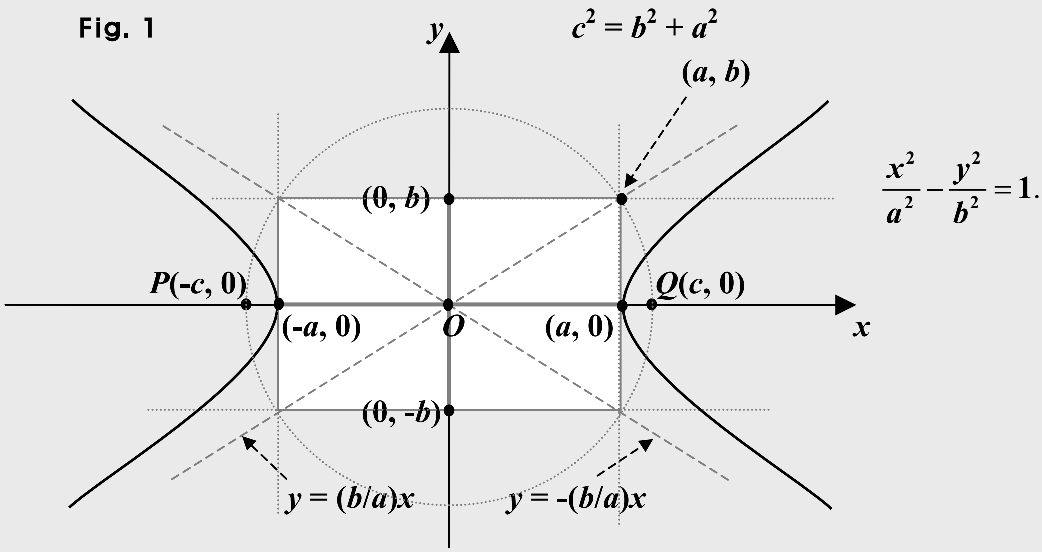
Then, we can notice also that if a hyperbola is vertical, the ***x***-coordinates are the same at all the five points: the vertices, foci, and center, which share thus, the same ***x***-coordinate.

Next, a hyperbola is symmetric, and has two axes of symmetry, called main axes, too.

One of the two is called the major axis, and the other is called the minor axis. Usually though, the major is called the transverse axis, and the minor is called the conjugate axis.

And in the case of the hyperbola below, the line segment connecting the vertices is the transverse axis, the length of which is $\mathbf{2}\boldsymbol{a}$, and the line segment connecting the two points $\mathbf{(0,}\ \boldsymbol{b}\mathbf{)}$ and $\mathbf{(0,\ -}\boldsymbol{b}\mathbf{)}$ is the conjugate axis, the length of which is $\mathbf{2}\boldsymbol{b}$. Usually though, we just call $\mathbf{2}\boldsymbol{a}$ the transverse axis, and call $\mathbf{2}\boldsymbol{b}$ the conjugate axis.

Fig. 1



What then about $\boldsymbol{a}$ and $\boldsymbol{b}$?

We call $\boldsymbol{a}$ the semi major axis, and $\boldsymbol{b}$ is called the semi minor axis.

And a hyperbola has two lines called asymptotes, each of which has a slope. We can get the two slopes using the semis, $\boldsymbol{a}$ and $\boldsymbol{b}$. Note however, in cases of asymptotes, it doesn't matter which of $\boldsymbol{a}$ and $\boldsymbol{b}$ is the semi major axis. We just use the definition for slopes, rise over run. So in the case of the asymptotes above, one slope is $\frac{b}{a}$, and the other is $-\frac{b}{a}$.

And since both asymptotes pass through the origin, they are the two lines $\boldsymbol{y = \pm\frac{b}{a}x}$.

And as we can see in the equation of the hyperbola above, if a hyperbola is horizontal, the sign of $\boldsymbol{x^2}$-term is positive, and the sign of $\boldsymbol{y^2}$-term is negative.

And if it is vertical, the sign of $\boldsymbol{x^2}$-term is negative, and the sign of $\boldsymbol{y^2}$-term is positive.

Next, the focal distance is the distance from a focus to the center.

The two foci are symmetric about the center. So the focal distance is half the distance between the foci. So in the hyperbola above, $\boldsymbol{c}$ is the focal distance.

Next, in the case of the hyperbola above, the two foci are $\boldsymbol{(-c, 0)}$ and $\boldsymbol{(c, 0)}$.

The difference between two distances from every point in a hyperbola to its two foci is constant, that is, the same. And in the case of the hyperbola above, the difference is $\boldsymbol{2a}$, which is (the length of) the transverse axis.

What then about this: $\boldsymbol{c^2 = b^2 + a^2}$?

It can be called the connective equation between the focal distance and the semi axes. And we can get the equation from the three facts as follows.

1. One is that $\boldsymbol{c}$ is the focal distance, that is, the distance from one focus to the center.

2. Another is that the focal distance is the radius of the circle that is centered at the center of the hyperbola, and is passing through the foci of the hyperbola.

3. And the other fact is that a right triangle is made by three points, one is a point where the circle meets an asymptote, another is a vertex, and the other is the center.

And in the case of the hyperbola above, the three points can be $\boldsymbol{(a, b)}$, $\boldsymbol{(a, 0)}$, and $\boldsymbol{(0, 0)}$.

So the three points above determine a right triangle, where the hypotenuse is $\boldsymbol{c}$, which is the focal distance. Using thus, the distance formula, we can get $\boldsymbol{c^2 = b^2 + a^2}$.

So the focal distance is $\boldsymbol{c = \sqrt{a^2 + b^2}}$.

And in the case of the horizontal hyperbola above, the two foci are **(-*c*, 0)** and **(*c*, 0)**.

So we can put them this way, too: $(-\sqrt{a^2+b^2},\ 0)$ and $(\sqrt{a^2+b^2},\ 0)$.
What if a hyperbola is vertical, and the semis are ***a*** and ***b***?

We still get $c^2 = b^2 + a^2$. So we get $c = \sqrt{a^2+b^2}$, too.

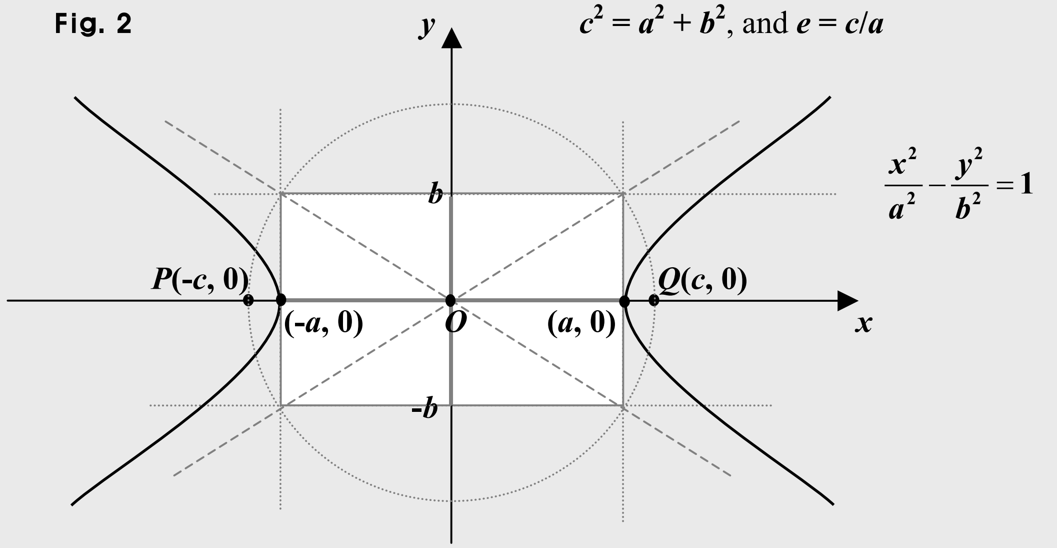
Therefore, in the case of a vertical hyperbola, too, the focal distance is $c = \sqrt{a^2+b^2}$.

And in the case of a hyperbola vertical and centered at the origin, the two foci are **(0, *c*)** and **(0, -*c*)**, because the transverse line is the ***y***-axis.

So in that case, we can put the foci this way, too: $(0,\ \sqrt{a^2+b^2})$ and $(0,\ -\sqrt{a^2+b^2})$.

Next, a hyperbola has a special number called an eccentricity, which is a ratio of a focal distance to the semi major, i.e., the focal distance over half the transverse axis.

Fig. 2



So of a hyperbola horizontal, the eccentricity is $e = \dfrac{c}{a}$, and specifies the degree, to which the hyperbola bends. It can tell us thus, how flat or sharp the branches are.
Usually, an eccentricity is denoted by ***e***, and we have $e > 1$ for a hyperbola.

And if the eccentricity ***e*** were 1, each branch would be a ray.

So if the eccentricity is close to 1, each branch is very sharp at the vertex.

And if $\boldsymbol{e}$ is very large, the branches are very flat, and look like two parallel lines.

So the bigger the eccentricity, the flatter the branches.

Now, the eccentricity $\boldsymbol{e}$ is a ratio, which is the focal distance over half the transverse.

So in the case of a hyperbola horizontal, we get $e = \dfrac{c}{a}$. because the transverse is $\mathbf{2}a$.

And in the case of a hyperbola vertical, we get $e = \dfrac{c}{b}$, because the transverse is $\mathbf{2}b$.

So we can put the focal distance c the ways below, too:

If horizontal, since $e = \dfrac{c}{a}$, we get $c = ae$, and if vertical, since $e = \dfrac{c}{b}$, we get $c = be$.

How though, would each branch be a ray if it were the case where $e = 1$?

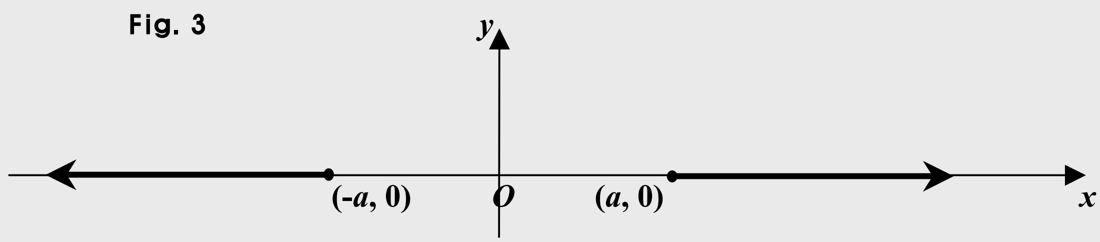
If a hyperbola is horizontal, we have $e = \dfrac{c}{a}$. So we get $c = a$ if $e = 1$.

And next, we have $c = \sqrt{a^2 + b^2}$, so if $c = a$, we would get $b = 0$.

Thus, if $e = 1$, the invisible rectangle would be a line segment of length $\mathbf{2}a$.

So if it were the case where $e = 1$, the hyperbola would be as follows.

Fig. 3



Therefore, if $e = 1$, each branch would be a ray beginning at the vertex.

Let's take another look at the eccentricity, but this time, a bit differently.

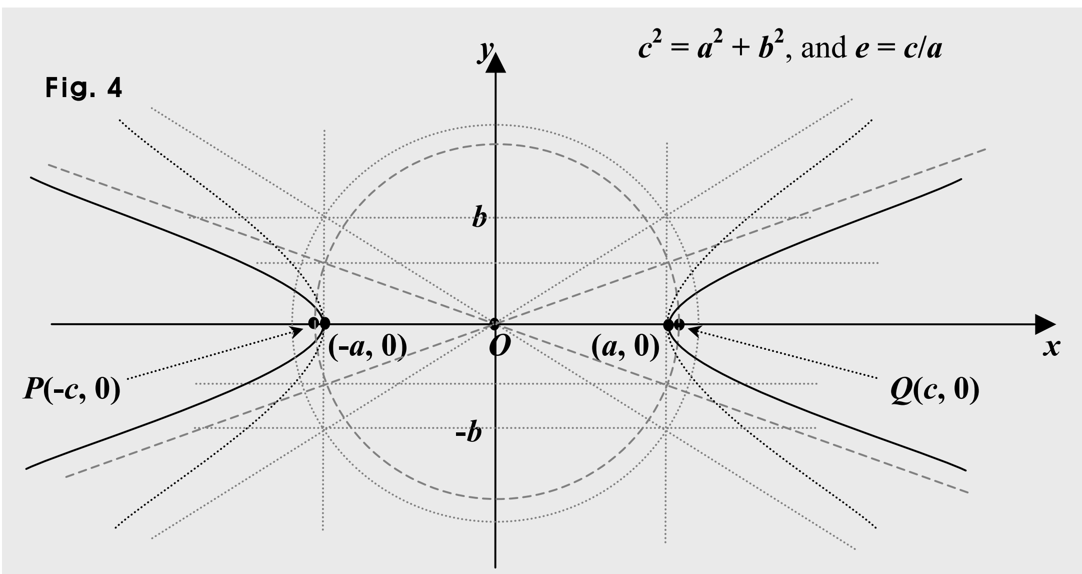
If a hyperbola is horizontal, we have $\boldsymbol{e} = \dfrac{\boldsymbol{c}}{\boldsymbol{a}}$, and $\boldsymbol{c} = \sqrt{\boldsymbol{a}^2 + \boldsymbol{b}^2}$.

Then, we can put the eccentricity ***e*** this way, too: $\boldsymbol{e} = \dfrac{\sqrt{\boldsymbol{a}^2 + \boldsymbol{b}^2}}{\boldsymbol{a}} = \sqrt{\dfrac{\boldsymbol{a}^2 + \boldsymbol{b}^2}{\boldsymbol{a}^2}} = \sqrt{\mathbf{1} + \dfrac{\boldsymbol{b}^2}{\boldsymbol{a}^2}}$.

Then, as ***b*** changes and approaches 0, the eccentricity ***e*** changes and approaches 1. And since $\boldsymbol{c} = \sqrt{\boldsymbol{a}^2 + \boldsymbol{b}^2}$, the focal distance ***c*** approaches ***a***. So the foci approach the vertices. Thus, as ***b*** changes the way above, each branch changes and approaches a ray beginning at its vertex.

In the figure below, we can see that as ***b*** gets smaller, the branches get sharper at the vertices. So both branches would be two rays beginning at the vertices if ***b*** were 0.

Fig. 4



So as ***b*** changes the way above, each branch approaches a ray beginning at its vertex, and the foci approach the vertices. And the same is true for a hyperbola vertical, too

If a hyperbola is vertical, we have $\boldsymbol{e} = \dfrac{\boldsymbol{c}}{\boldsymbol{b}}$, and $\boldsymbol{c} = \sqrt{\boldsymbol{a}^2 + \boldsymbol{b}^2}$.

Then, we can put the eccentricity ***e*** this way, too: $\boldsymbol{e} = \dfrac{\sqrt{\boldsymbol{a}^2 + \boldsymbol{b}^2}}{\boldsymbol{b}} = \sqrt{\dfrac{\boldsymbol{b}^2 + \boldsymbol{a}^2}{\boldsymbol{b}^2}} = \sqrt{\mathbf{1} + \dfrac{\boldsymbol{a}^2}{\boldsymbol{b}^2}}$.

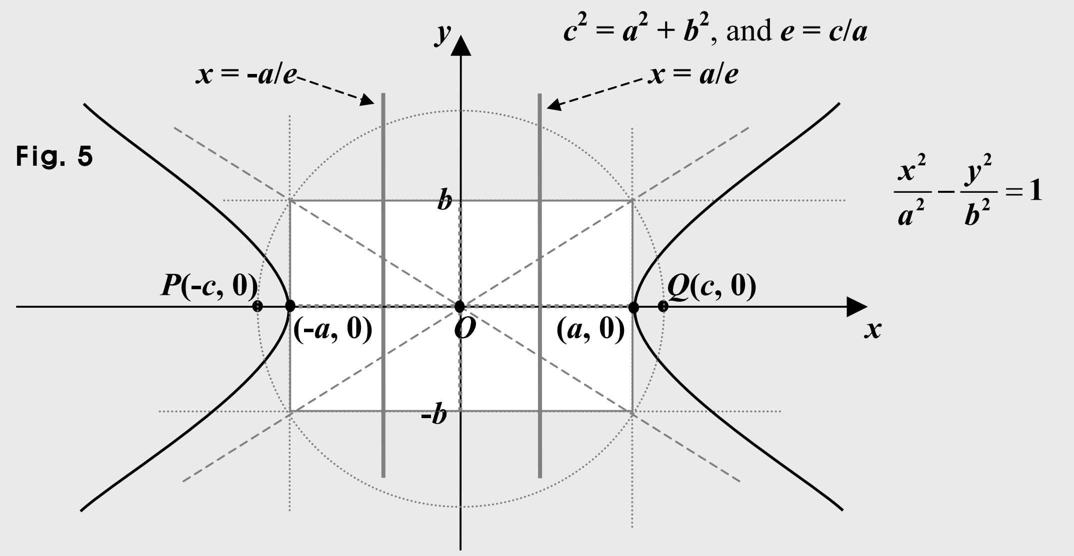
Then, as ***a*** changes and approaches 0, the eccentricity ***e*** changes and approaches 1. And since $\boldsymbol{c} = \sqrt{\boldsymbol{a}^2 + \boldsymbol{b}^2}$, the focal distance ***c*** approaches ***b***. So the foci approach the vertices.

Thus, as a changes the way above, each branch changes and approaches a ray beginning at its vertex.

And next, as in the case of an ellipse, a hyperbola has two ***directrices***, too.

In the figure below, the two vertical lines in gray are the two directrices, one is with the left branch, and the other is with the right branch.

Fig. 5



So if a hyperbola is horizontal, directrices are vertical lines, that is, lines perpendicular to the x-axis. And the focus on the left is said to correspond to the directrix on the left, and the focus on the right is said to correspond to the directrix on the right.

So in the figure above, the vertical line $x = -a/e$ is the directrix corresponding to the focus P, and the other line $x = a/e$ is the directrix corresponding to the focus Q.

How are the two directrices though, the two lines $x = -\frac{a}{e}$, and $x = \frac{a}{e}$?

To begin with, in an ellipse, the distance from a point to a focus is less than the distance from the point to the directrix corresponding to the focus, since $0 < e < 1$ for an ellipse.

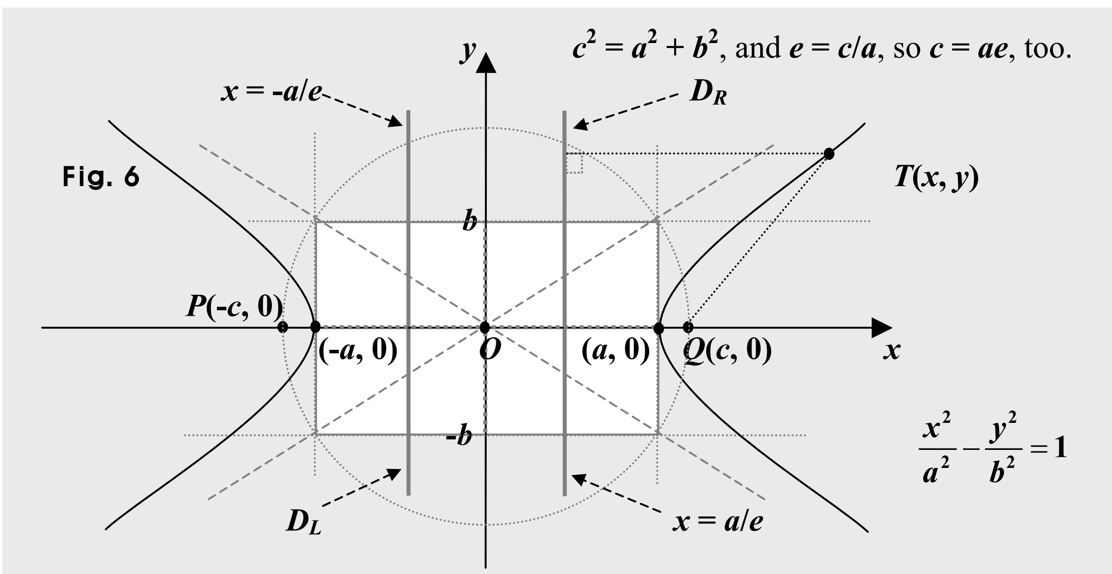
However, in a hyperbola, the distance from a point to a focus is greater than the distance from the point to the directrix corresponding to the focus, because $e > 1$ for a hyperbola.

So next, we can use the fact above the way as follows.

Suppose d is the distance from a point in a hyperbola to a focus, and D is the distance from the point to the directrix corresponding to the focus.
Then, we get $d = eD$ where e is the eccentricity. In other words, we get $\frac{d}{D} = e$.
And the same is true for the other directrix, too. What then are the two directrices?

Assuming H is the hyperbola below, we can say that the directrices of H are two vertical lines in gray, one is D_L, which is $x = -\frac{a}{e}$, and the other is D_R, which is $x = \frac{a}{e}$.

Fig. 6



Let's now check to see if the directrices of H are $x = -\frac{a}{e}$, and $x = \frac{a}{e}$.
To begin with, we know the fact that $d = eD$, where e is the eccentricity, d is the distance from a point in a hyperbola to a focus, and D is the distance from the point to the directrix corresponding to the focus. So we are going to use the fact above.

Suppose first, $x = m$ is the equation of the directrix D_R shown in the figure above, and next, the arbitrary point $T(x, y)$ is now in the right branch.

Then first, we can see that $(x - m)$ is the distance from $T(x, y)$ to the directrix D_R.

And next, assuming q is the distance from $T(x, y)$ to the focus $Q(c, 0)$, we get

$q^2 = (x - c)^2 + (y - 0)^2 = (x - c)^2 + y^2 \Rightarrow q^2 = (x - c)^2 + y^2$.

And we have $\dfrac{x^2}{a^2} - \dfrac{y^2}{b^2} = 1$, and $c^2 = b^2 + a^2$.

So we get $\dfrac{x^2}{a^2} - \dfrac{y^2}{b^2} = \dfrac{x^2}{a^2} - \dfrac{y^2}{c^2 - a^2} = \dfrac{x^2}{a^2} + \dfrac{y^2}{a^2 - c^2} = 1 \Rightarrow \dfrac{y^2}{a^2 - c^2} = 1 - \dfrac{x^2}{a^2} = \dfrac{a^2 - x^2}{a^2}$

$\Rightarrow y^2 = \dfrac{(a^2 - c^2)(a^2 - x^2)}{a^2}$. Thus now, getting back to this: $q^2 = (x - c)^2 + y^2$, we get

$$\Rightarrow q^2 = (x - c)^2 + \frac{(a^2 - c^2)(a^2 - x^2)}{a^2} = \frac{a^2(x - c)^2 + (a^2 - c^2)(a^2 - x^2)}{a^2}.$$

Meanwhile, $a^2(x - c)^2 + (a^2 - c^2)(a^2 - x^2) = a^2(x^2 - 2cx + c^2) + a^4 - a^2x^2 - a^2c^2 + c^2x^2$
$= a^2x^2 - 2a^2cx + a^2c^2 + a^4 - a^2x^2 - a^2c^2 + c^2x^2 = (a^4 - 2a^2cx + c^2x^2) = (a^2 - cx)^2$.

So we get $q^2 = \dfrac{(a^2 - cx)^2}{a^2} = \left(\dfrac{a^2 - cx}{a}\right)^2 = \left(a - \dfrac{cx}{a}\right)^2$.

And we know q is a distance. So we get $q = \left|a - \dfrac{cx}{a}\right|$.

And we know $T(x, y)$ is now in the right branch. So we now have $x \geq a$, and $c > a$.

Thus, we get $x \geq a \Rightarrow \dfrac{x}{a} \geq 1 \Rightarrow \dfrac{cx}{a} \geq c \Rightarrow -\dfrac{cx}{a} \leq -c \Rightarrow a - \dfrac{cx}{a} \leq a - c < 0$, since $c > a$.

So we get $a - \dfrac{cx}{a} < 0 \Rightarrow \dfrac{cx}{a} - a > 0$. And we have $q = \left|a - \dfrac{cx}{a}\right|$.

Thus, we get $q = \dfrac{cx}{a} - a$. And we have $e = \dfrac{c}{a}$. So we get $c = ae$, too.

Thus, we get $q = \frac{cx}{a} - a = \frac{aex}{a} - a = ex - a \Rightarrow q = ex - a$.

Next, we have $d = eD$, where d is the distance from a point in a hyperbola to a focus, and D is the distance from the point to the directrix corresponding to the focus.

And in this case, d is the distance from the arbitrary point $T(x, y)$ to the focus $Q(c, 0)$, and thus, is q, which is $ex - a$. So the distance d is $ex - a$.

And D is the distance from the point $T(x, y)$ to the right directrix corresponding to the focus Q, and the right directrix is D_R, which is the line $x = m$. So the distance D is $x - m$.

Thus, in sum, we have $d = ex - a$, and $D = x - m$. So we get

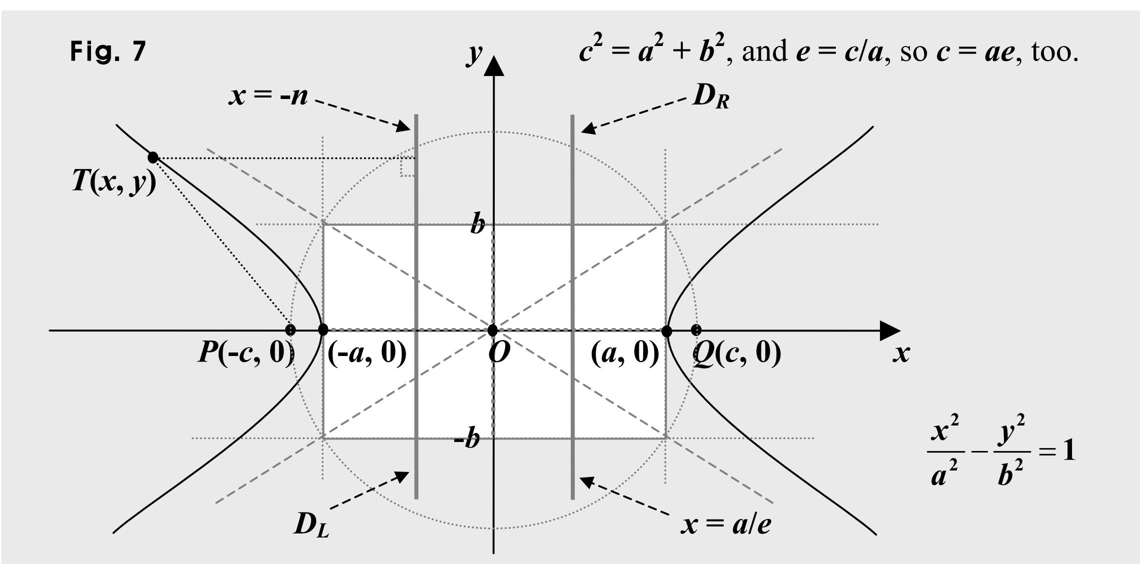
$$d = eD \Rightarrow ex - a = e(x - m) = ex - em \Rightarrow ex - a = ex - em \Rightarrow a = em \Rightarrow m = a/e.$$

And the directrix is D_R, which is the line $x = m$.

So the directrix D_R is the line $x = \frac{a}{e}$.

Suppose this time, $x = -n$ is the equation of the directrix D_L as shown in the figure below, and the arbitrary point $T(x, y)$ is now in the left branch.

Fig. 7



Then first, we can see that ($-n - x$) is the distance from $T(x, y)$ to the directrix D_L.
And next, assuming p is the distance from $T(x, y)$ to the focus $P(-c, 0)$, we get
$p^2 = (x + c)^2 + (y - 0)^2 = (x + c)^2 + y^2 \Rightarrow p^2 = (x + c)^2 + y^2$.

And we know that since we have $\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1$, and $c^2 = b^2 + a^2$, we can get

$y^2 = \frac{(a^2 - c^2)(a^2 - x^2)}{a^2}$.

So we get $p^2 = (x + c)^2 + \frac{(a^2 - c^2)(a^2 - x^2)}{a^2} = \frac{a^2(x + c)^2 + (a^2 - c^2)(a^2 - x^2)}{a^2}$.

Meanwhile, $a^2(x + c)^2 + (a^2 - c^2)(a^2 - x^2) = a^2(x^2 + 2cx + c^2) + a^4 - a^2x^2 - a^2c^2 + c^2x^2$
$= a^2x^2 + 2a^2cx + a^2c^2 + a^4 - a^2x^2 - a^2c^2 + c^2x^2 = (a^4 + 2a^2cx + c^2x^2) = (a^2 + cx)^2$.

So we get $p^2 = \frac{(a^2 + cx)^2}{a^2} = \left(\frac{a^2 + cx}{a}\right)^2 = \left(a + \frac{cx}{a}\right)^2$.

And we know p is a distance. So we get $p = |a + \frac{cx}{a}|$.

And we know $T(x, y)$ is now in the left branch. So we now have $x \leq -a$, and $c > a$.

Thus, we get $x \leq -a \Rightarrow \frac{x}{a} \leq -1 \Rightarrow \frac{cx}{a} \leq -c \Rightarrow a + \frac{cx}{a} \leq a - c < 0$, since $c > a$.

So we get $a + \frac{cx}{a} < 0 \Rightarrow -a - \frac{cx}{a} > 0$. And we have $p = |a + \frac{cx}{a}|$.

Thus, we get $p = -a - \frac{cx}{a}$. And we have $e = \frac{c}{a}$. So we get $c = ae$, too.

Thus, we get $p = -a - \frac{cx}{a} = -a - \frac{aex}{a} = -a - ex \Rightarrow p = -(a + ex)$.

Next, we have $d = eD$, where d is the distance from a point in a hyperbola to a focus, and D is the distance from the point to the directrix corresponding to the focus.
And in this case, d is the distance from the arbitrary point $T(x, y)$ to the focus $P(-c, 0)$, and thus, is p, which is $-(a + ex)$. So the distance d is $-(a + ex)$.

And $\boldsymbol{D}$ is the distance from the point $\boldsymbol{T(x, y)}$ to the left directrix corresponding to the focus $\boldsymbol{P}$, and the left directrix is $\boldsymbol{D_L}$, which is the line $\boldsymbol{x = -n}$. So the distance $\boldsymbol{D}$ is $\boldsymbol{-n - x}$.

Thus, in sum, we have $\boldsymbol{d = -(a + ex)}$, and $\boldsymbol{D = -(n + x)}$. So we get

$\boldsymbol{d = eD \Rightarrow -(a + ex) = -e(n + x) \Rightarrow a + ex = e(n + x) = en + ex \Rightarrow a = en \Rightarrow n = a/e}$.

And the left directrix is $\boldsymbol{D_L}$, which is the line $\boldsymbol{x = -n}$.
So the directrix $\boldsymbol{D_L}$ is the line $\boldsymbol{x = -\frac{a}{e}}$.

And let's now find for instance, the directrices of this hyperbola: $\boldsymbol{\frac{x^2}{4^2} - \frac{y^2}{2^2} = 1}$.

We have $\boldsymbol{x = \pm\frac{a}{e}}$, $\boldsymbol{e = \frac{c}{a}}$, and $\boldsymbol{c = \sqrt{a^2 + b^2}}$. So first, we can get $\boldsymbol{x = \pm\frac{a^2}{c}}$.

And we have $\boldsymbol{a = 4}$, and $\boldsymbol{b = 2}$.

So finding first, the focal distance, we get $\boldsymbol{c = \sqrt{a^2 + b^2} = \sqrt{4^2 + 2^2} = \sqrt{20} = 2\sqrt{5}}$.

Thus, next, we can get $\boldsymbol{\frac{a^2}{c} = \frac{16}{2\sqrt{5}} = \frac{8}{\sqrt{5}} = \frac{8\sqrt{5}}{5}}$.

So the directrices are $\boldsymbol{x = \pm\frac{8\sqrt{5}}{5} \approx \pm 3.5777}$.

What then about the directrices of this hyperbola: $\boldsymbol{\frac{(x-u)^2}{a^2} - \frac{(y-v)^2}{b^2} = 1}$?

We know that the hyperbola $\boldsymbol{H}$ is $\boldsymbol{\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1}$.

So assuming $\boldsymbol{G}$ is the hyperbola in question, we can get $\boldsymbol{G}$ translating the hyperbola $\boldsymbol{H}$ in the amount of $\boldsymbol{u}$ in the direction of the $\boldsymbol{x}$-axis, and in the amount of $\boldsymbol{v}$ along the $\boldsymbol{y}$-axis.

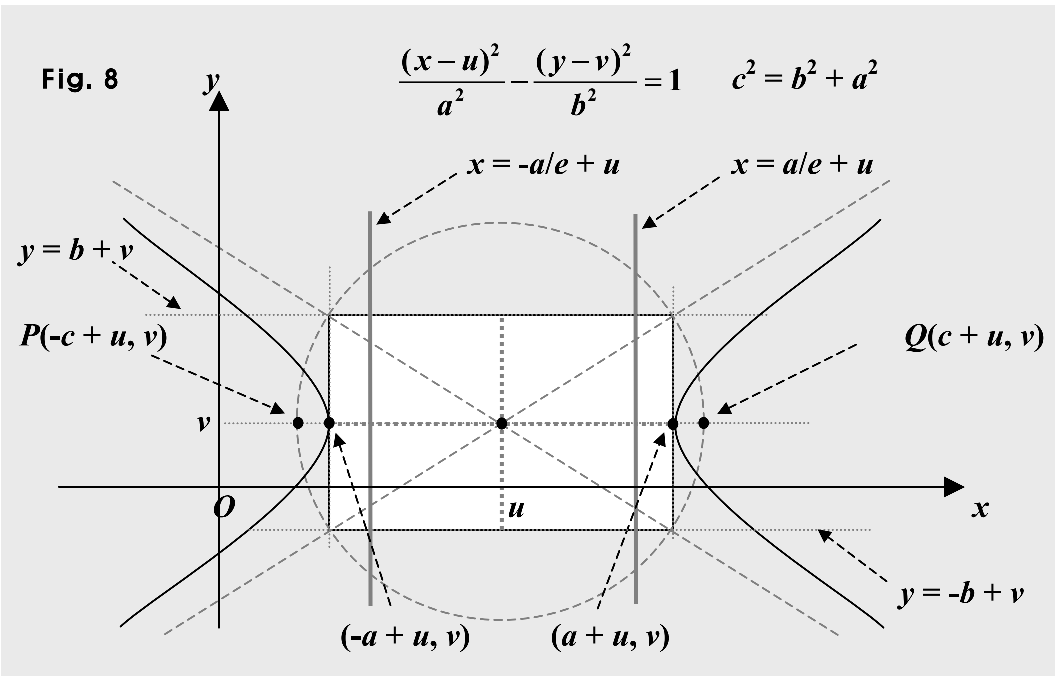
And we know that the center of the hyperbola $\boldsymbol{G}$ is $\boldsymbol{(u, v)}$.
So in short, moving the center of the hyperbola $\boldsymbol{H}$ to the point $\boldsymbol{(u, v)}$, we can get $\boldsymbol{G}$.
Then, the directrices of $\boldsymbol{H}$ will be translated exactly the way the hyperbola $\boldsymbol{H}$ gets translated.

And we know that the directrices of $\boldsymbol{H}$ are $\boldsymbol{x} = \pm\dfrac{\boldsymbol{a}}{\boldsymbol{e}}$.

So the directrices of $\boldsymbol{G}$ will be as follows $\boldsymbol{x} = \pm\dfrac{\boldsymbol{a}}{\boldsymbol{e}} + \boldsymbol{u}$.

That is, the two directrices are $\boldsymbol{x} = -\dfrac{\boldsymbol{a}}{\boldsymbol{e}} + \boldsymbol{u}$, and $\boldsymbol{x} = \dfrac{\boldsymbol{a}}{\boldsymbol{e}} + \boldsymbol{u}$.

Fig. 8



Let's now, for instance, find the directrices of this hyperbola: $\dfrac{(\boldsymbol{x}-\mathbf{2})^2}{\mathbf{4}^2} - \dfrac{(\boldsymbol{y}-\mathbf{1})^2}{\mathbf{2}^2} = \mathbf{1}$.

First, we've already found the directrices of $\dfrac{\boldsymbol{x}^2}{\mathbf{2}^2} - \dfrac{\boldsymbol{y}^2}{\mathbf{4}^2} = \mathbf{1}$, and they are $\boldsymbol{x} = \pm\dfrac{\mathbf{8}\sqrt{\mathbf{5}}}{\mathbf{5}}$.

So next, translating the directrices above by 2 along the $\boldsymbol{x}$-axis, and by 1 along the $\boldsymbol{y}$-axis, we get the directrices of the hyperbola we want to find the directrices of.

We know however, translating a vertical line along the $\boldsymbol{y}$-axis, we see no change in the line. So assuming (2, 1) is a new center, we just add the $\boldsymbol{x}$-coordinate of the new center to the directrices of the hyperbola centered at the origin.

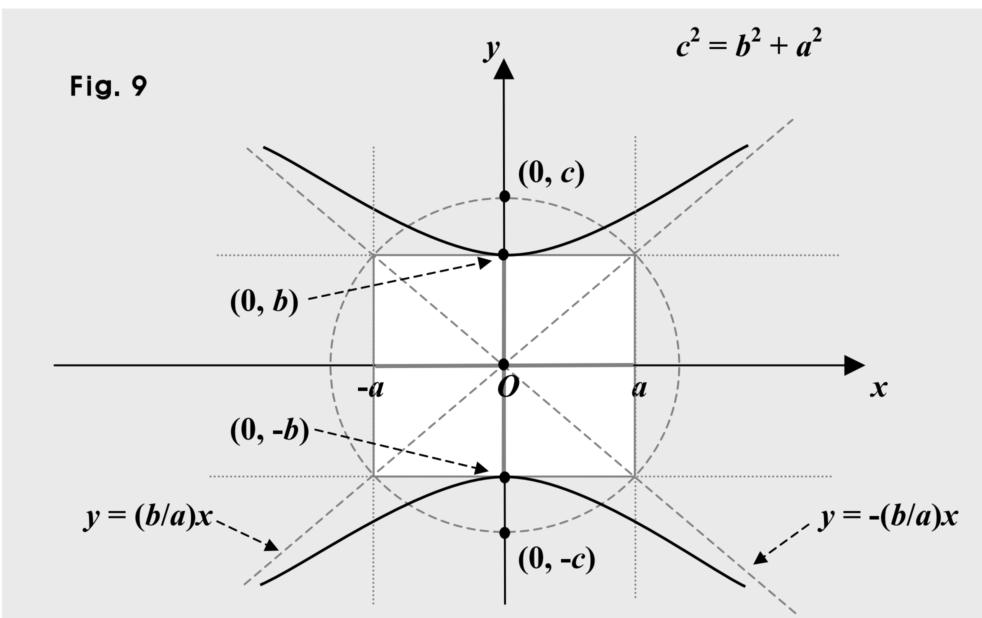
Then, we get $x = \pm\frac{8\sqrt{5}}{5} + 2$, which are the directrices we want. And more specifically,

one is $x = -\frac{8\sqrt{5}}{5} + 2 = \frac{10 - 8\sqrt{5}}{5} \approx -1.58$, which corresponds to the left focus.

And the other is $x = \frac{8\sqrt{5}}{5} + 2 = \frac{10 + 8\sqrt{5}}{5} \approx 5.58$, which corresponds to the right focus.

What if the hyperbola is vertical, that is, if it is $\frac{y^2}{b^2} - \frac{x^2}{a^2} = 1$?

Putting first, the hyperbola above in a graph, we can put it the way as follows.



To begin with, we know the fact that if in a hyperbola, d is the distance from a point to a focus, and D is the distance from the point to the directrix corresponding to the focus, we get $d = eD$, where e is the eccentricity. In other words, we get $\frac{d}{D} = e$.

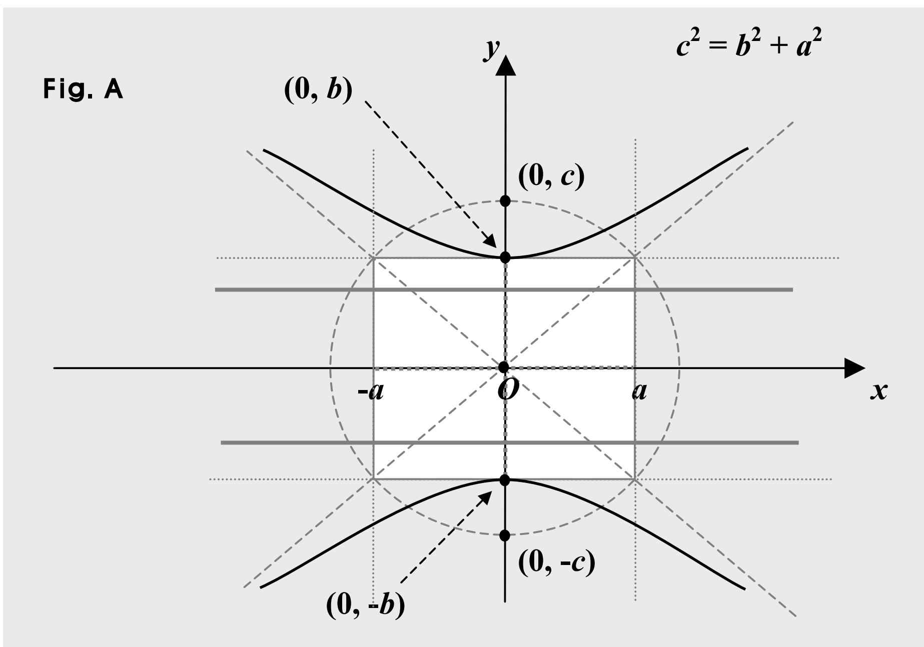
Next, the directrices are perpendicular to the transverse axis.

And we know if a hyperbola is horizontal, the transverse axis is parallel to the x-axis.
So the directrices of a hyperbola horizontal are perpendicular to the x-axis.
What then about the directrices of a hyperbola vertical?

If the hyperbola is vertical, the transverse axis is parallel to the $\boldsymbol{y}$-axis.
So the directrices are horizontal, that is, perpendicular to the $\boldsymbol{y}$-axis.

And next, the directrices of $\boldsymbol{H}$ are $\boldsymbol{x} = \pm\frac{\boldsymbol{a}}{\boldsymbol{e}}$, where $\boldsymbol{a}$ is half the transverse axis, and $\boldsymbol{e}$ is the eccentricity. What then are the directrices of the hyperbola vertical?

If the vertical hyperbola is $\boldsymbol{K}$, the directrices of $\boldsymbol{K}$ are $\boldsymbol{y} = \pm\frac{\boldsymbol{b}}{\boldsymbol{e}}$, where $\boldsymbol{b}$ is half the transverse axis.



So for instance, what are the directrices of this hyperbola: $\frac{\boldsymbol{y}^2}{\boldsymbol{4}^2} - \frac{\boldsymbol{x}^2}{\boldsymbol{2}^2} = \boldsymbol{1}$?

We know finding the directrices of $\frac{\boldsymbol{x}^2}{\boldsymbol{4}^2} - \frac{\boldsymbol{y}^2}{\boldsymbol{2}^2} = \boldsymbol{1}$, we get $\boldsymbol{x} = \pm\frac{\boldsymbol{8\sqrt{5}}}{\boldsymbol{5}}$.

So the directrices of $\frac{\boldsymbol{y}^2}{\boldsymbol{4}^2} - \frac{\boldsymbol{x}^2}{\boldsymbol{2}^2} = \boldsymbol{1}$ are $\boldsymbol{y} = \pm\frac{\boldsymbol{8\sqrt{5}}}{\boldsymbol{5}}$. How?

We have $\boldsymbol{y} = \pm\frac{\boldsymbol{b}}{\boldsymbol{e}}$, $\boldsymbol{e} = \frac{\boldsymbol{c}}{\boldsymbol{b}}$, and $\boldsymbol{c} = \sqrt{\boldsymbol{b}^2 + \boldsymbol{a}^2}$. So first, we can get $\boldsymbol{y} = \pm\frac{\boldsymbol{b}^2}{\boldsymbol{c}}$.

And we have $\boldsymbol{a} = \mathbf{2}$, and $\boldsymbol{b} = \mathbf{4}$.

So finding first, the focal distance, we get $c = \sqrt{a^2 + b^2} = \sqrt{4^2 + 2^2} = \sqrt{20} = 2\sqrt{5}$.

Thus, the directrices are $y = \pm\frac{b^2}{c} = \pm\frac{16}{2\sqrt{5}} = \pm\frac{8}{\sqrt{5}} = \pm\frac{8\sqrt{5}}{5}$. That is, $y = \pm\frac{8\sqrt{5}}{5}$.

What then about the directrices of this hyperbola: $\frac{(y-v)^2}{b^2} - \frac{(x-u)^2}{a^2} = 1$?

Assuming the hyperbola above is $\boldsymbol{J}$, we can get $\boldsymbol{J}$ translating the hyperbola $\boldsymbol{K}$ in the amount of $\boldsymbol{u}$ along the $\boldsymbol{x}$-axis, and in the amount of $\boldsymbol{v}$ in the direction of the $\boldsymbol{y}$-axis.

And we know that the center of the hyperbola $\boldsymbol{J}$ is $(\boldsymbol{u}, \boldsymbol{v})$.
So in short, moving the center of the hyperbola $\boldsymbol{K}$ to the point $(\boldsymbol{u}, \boldsymbol{v})$, we can get $\boldsymbol{J}$.

Then, the directrices of $\boldsymbol{K}$ will be translated exactly the way the hyperbola $\boldsymbol{K}$ gets translated.
And we know that the directrices of $\boldsymbol{K}$ are $y = \pm\frac{b}{e}$.

So the directrices of $\boldsymbol{J}$ will be as follows. $y = \pm\frac{b}{e} + v$

That is, the two directrices are $y = -\frac{b}{e} + v$, and $y = \frac{b}{e} + v$.

Let's now, for instance, find the directrices of this hyperbola $\frac{(y+5)^2}{4^2} - \frac{(x-3)^2}{2^2} = 1$.

First, we've already found the directrices of $\frac{y^2}{4^2} - \frac{x^2}{2^2} = 1$, and they are $y = \pm\frac{8\sqrt{5}}{5}$.
So next, translating those directrices by 3 along the $\boldsymbol{x}$-axis, and by -5 along the $\boldsymbol{y}$-axis, we get the directrices of the hyperbola we want to find the directrices of.

We know however, the directrices above are horizontal, and translating horizontal lines along the $\boldsymbol{x}$-axis, we see no change in the lines.

So assuming (3, -5) is a new center, we just add the $\boldsymbol{y}$-coordinate of the new center to the directrices above.

Note however, we can do so, because the hyperbola getting translated is centered at the origin.

Then, we get $\boldsymbol{y}=\pm\frac{\mathbf{8}\sqrt{\mathbf{5}}}{\mathbf{5}}-\mathbf{5}$, which are the directrices we want. And more specifically,

one is $\boldsymbol{y}=-\frac{\mathbf{8}\sqrt{\mathbf{5}}}{\mathbf{5}}-\mathbf{5}\approx\mathbf{-8.58}$, which corresponds to the lower focus.

And the other is $\boldsymbol{y}=\frac{\mathbf{8}\sqrt{\mathbf{5}}}{\mathbf{5}}-\mathbf{5}\approx\mathbf{-2.58}$, which corresponds to the upper focus.

And translating the curve of $\frac{y^2}{b^2}-\frac{x^2}{a^2}=1$ in the amount of $\boldsymbol{u}$ along the $\boldsymbol{x}$-axis, and in the amount of $\boldsymbol{v}$ along the $\boldsymbol{y}$-axis, we get the curve of $\frac{(y-v)^2}{b^2}-\frac{(x-u)^2}{a^2}=1$.

Besides, the equations of the asymptotes are as follows.

$y-v=\pm\frac{b}{a}(x-u)$, which are the two equations of the two asymptotes we get if we translate the asymptotes $y=\pm\frac{b}{a}x$ by $\boldsymbol{u}$ along the $\boldsymbol{x}$-axis and by $\boldsymbol{v}$ along the $\boldsymbol{y}$-axis.

And putting in graphs the two curves above along with their elements, we can put them the way as follows.

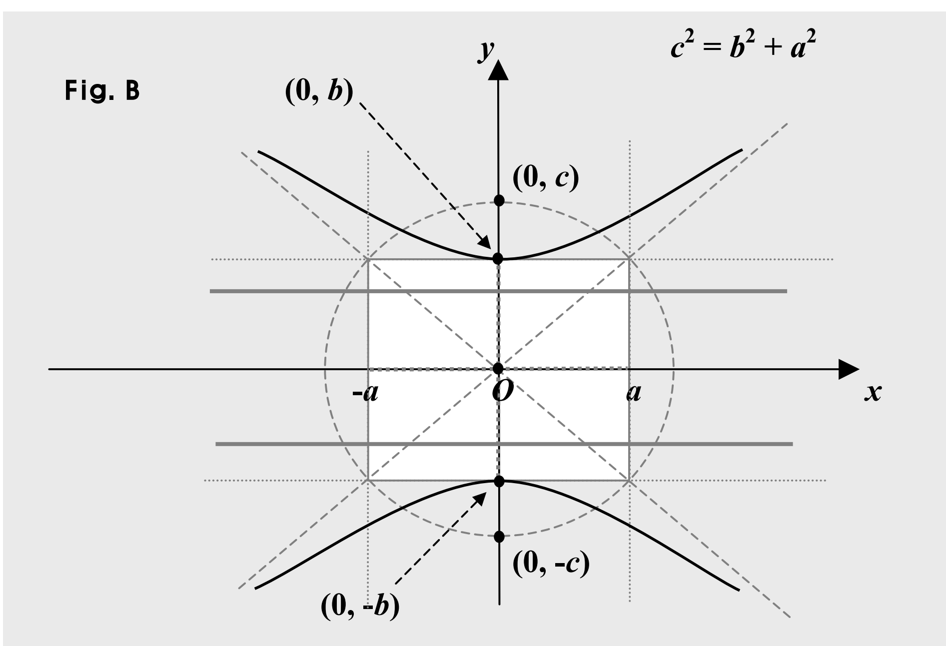
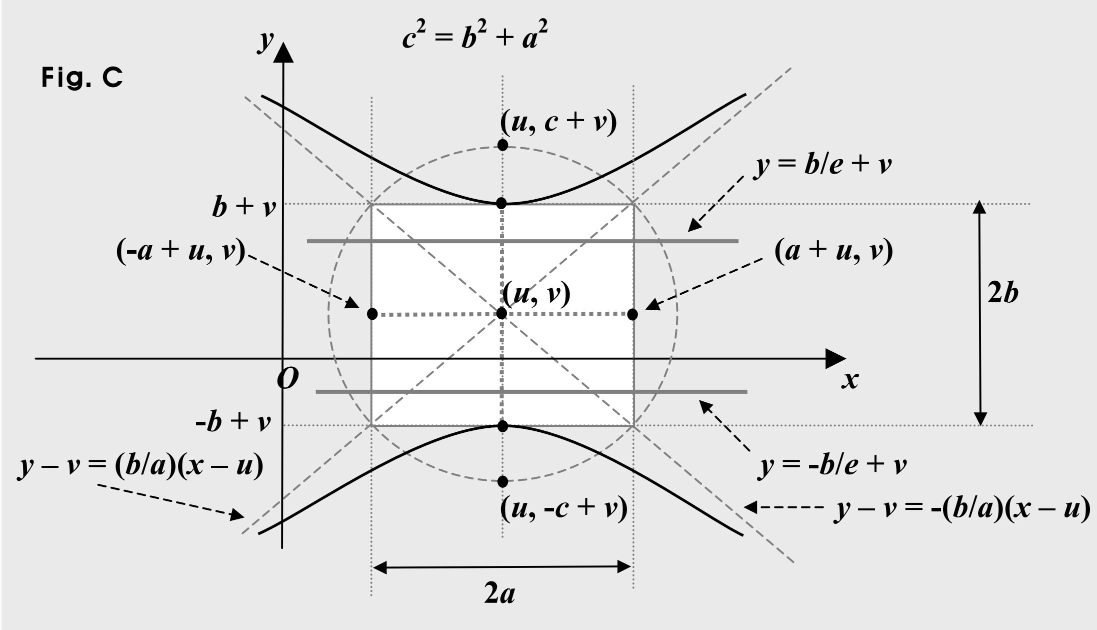


Fig. C



$c^2 = b^2 + a^2$
y
x
O
$(u, c + v)$
$y = b/e + v$
$b + v$
$(-a + u, v)$
$(a + u, v)$
(u, v)
2b
$-b + v$
$y - v = (b/a)(x - u)$
$y = -b/e + v$
$(u, -c + v)$
$y - v = -(b/a)(x - u)$
2a

4.4. Summary on Hyperbolas

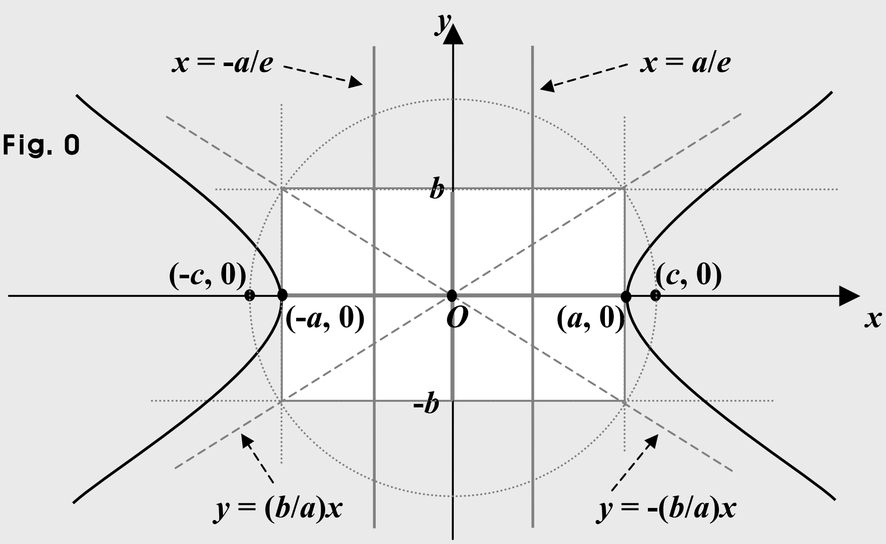
Now, summing up, we can put together the ideas on hyperbolas the way as follows.

To begin with, putting a hyperbola in a standard equation, we can get $\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1$.

What hyperbola then is it?

If H is the hyperbola, H is centered at $(0, 0)$, the origin, and since the coefficient of x^2 is positive, it is horizontal, so the transverse axis is parallel to the x-axis, and is $2a$, and the conjugate axis is $2b$. And assuming c is the focal distance, we get $c^2 = b^2 + a^2$.

Fig. 0



So next, since the semi major axis is a, assuming e is the eccentricity, we get $e = \frac{c}{a}$, where c is the focal distance, of course. And we have $e > 1$, since $c > a$.

And next, since H is horizontal, the two directrices are two lines parallel to the y-axis.

The distance from the center to each directrix is $\frac{\boldsymbol{a}}{\boldsymbol{e}}$, where $\boldsymbol{a}$ is half the transverse axis, and $\boldsymbol{e}$ is the eccentricity. So since the center is **(0, 0)**, the two directrices are $\boldsymbol{x} = \pm\frac{\boldsymbol{a}}{\boldsymbol{e}}$.

And we know $\boldsymbol{e} = \frac{\boldsymbol{c}}{\boldsymbol{a}}$. So we can put the two directrices this way, too: $\boldsymbol{x} = \pm\frac{\boldsymbol{a}^2}{\boldsymbol{c}}$.

And since the center is **(0, 0)**, the two asymptotes are $\boldsymbol{y} = \pm\frac{\boldsymbol{b}}{\boldsymbol{a}}\boldsymbol{x}$.

Next, since the hyperbola ***H*** is horizontal, the center and foci have the same ***y***-coordinate. And the center is the midpoint between the foci. So since the center is **(0, 0)**, and the focal distance is ***c***, the two foci are **(-*c*, 0)** and **(*c*, 0)**.

Also, since ***H*** is horizontal, the center and the vertices share the same ***y***-coordinate, too.

And the center is the midpoint between the vertices. The distance from the center to a vertex is the semi major axis, which is half the transverse axis. So since half the transverse axis is ***a***, and the center is **(0, 0)**, the two vertices are **(-*a*, 0)** and **(*a*, 0)**.

So for instance, putting a horizontal hyperbola in an equation, we can get $\frac{\boldsymbol{x}^2}{\mathbf{4}^2} - \frac{\boldsymbol{y}^2}{\mathbf{3}^2} = \mathbf{1}$.

What hyperbola then is it?

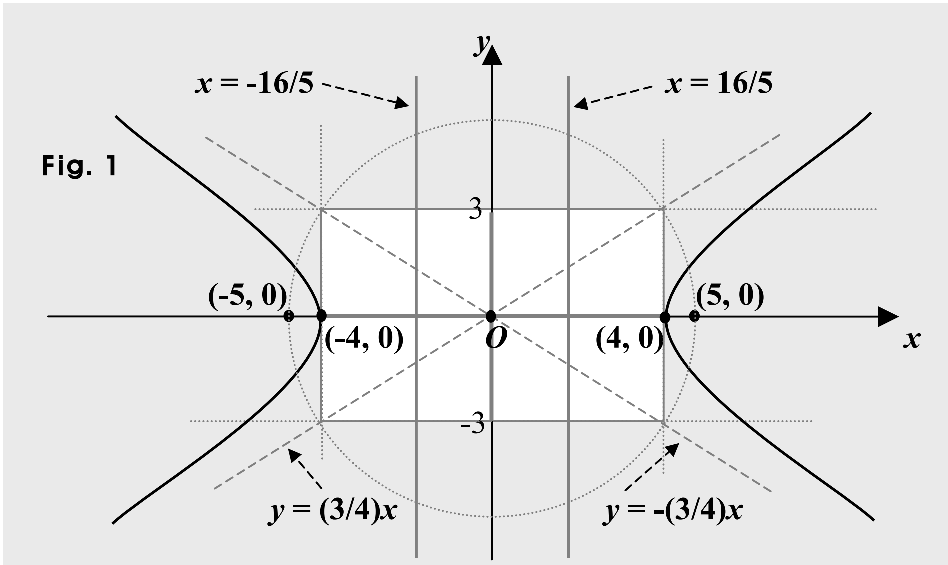
If ***K*** is the hyperbola, ***K*** is centered at (0, 0), the origin, and is horizontal, so the transverse axis is parallel to the ***x***-axis, and is 8, and the conjugate axis is 6.

And the focal distance is 5, because if ***c*** is the focal distance, we get $\boldsymbol{c}^2 = \boldsymbol{b}^2 + \boldsymbol{a}^2$, where $\boldsymbol{a} = \mathbf{4}$, and $\boldsymbol{b} = \mathbf{3}$, where ***a*** is the semi major axis, and ***b*** is the semi minor axis.

And since the center is **(0, 0)**, the two asymptotes are $\boldsymbol{y} = \pm\frac{\boldsymbol{b}}{\boldsymbol{a}}\boldsymbol{x} = \pm\frac{3}{4}\boldsymbol{x}$.

Next, assuming ***e*** is the eccentricity, we get $\boldsymbol{e} = \frac{\boldsymbol{c}}{\boldsymbol{a}} = \frac{5}{4}$, where ***a*** is the semi major axis, which is half the transverse axis, and ***c*** is the focal distance.

Fig. 1



And since the center is **(0, 0)**, the two directrices are $\boldsymbol{x} = \pm\frac{16}{5}$.

Next, since the hyperbola ***K*** is horizontal, the center is (0, 0), and the focal distance is 5, the two foci are (-5, 0) and (5, 0). (Note that horizontal means the same ***y***-coordinate.)

Also, since ***K*** is horizontal, half the transverse axis, that is, the semi major axis is 4, and the center is (0, 0), the two vertices are (-4, 0) and (4, 0).

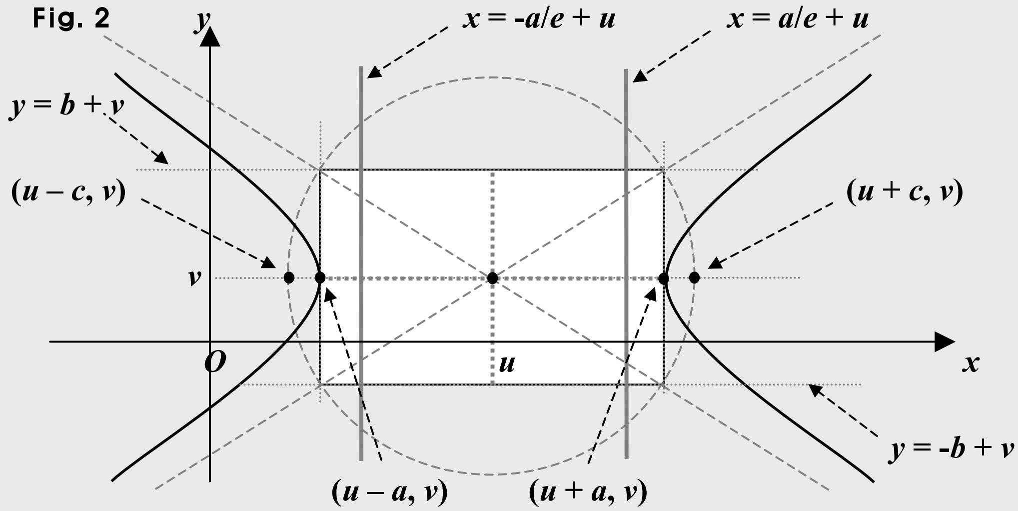
What then about the hyperbola as follows? $\frac{(x-u)^2}{a^2} - \frac{(y-v)^2}{b^2} = 1$

If ***G*** is the hyperbola, the hyperbola ***G*** is centered at **(*u*, *v*)**, and since the coefficient of x^2 is positive, ***G*** is horizontal, so the transverse axis is parallel to the ***x***-axis, and is **2*a***, and the conjugate axis is parallel to the ***y***-axis, and is **2*b***.

And assuming ***c*** is the focal distance, we get $c^2 = b^2 + a^2$, where ***a*** is half the transverse axis, and ***b*** is half the conjugate axis.

So next, since half the transverse axis is ***a***, assuming ***e*** is the eccentricity, we get $e = \frac{c}{a}$, where ***c*** is the focal distance, of course. Also, we have $e > 1$, too, since $c > a$.

Fig. 2



And since the center is (u, v), the two asymptotes are $y - v = \pm\frac{b}{a}(x - u)$.

Also, since the center is (u, v), the two directrices are $x = \pm\frac{a}{e} + u = \pm\frac{a^2}{c} + u$.

Next, since the hyperbola G is horizontal, the center is at (u, v), and the focal distance is c, the two foci are $(u - c, v)$ and $(u + c, v)$.

Also, since G is horizontal, the center is (u, v), and the semi major axis, that is, half the transverse axis is a, the two vertices are $(u - a, v)$ and $(u + a, v)$.

And notice that translating the hyperbola H in the amount of u along the x-axis, and in the amount of v along the y-axis, we get the hyperbola G.

Also, of course, translating the hyperbola G in the amount of $-u$ along the x-axis, and in the amount of $-v$ along the y-axis, we get the hyperbola H.

So for instance, putting a horizontal hyperbola in an equation, we can get

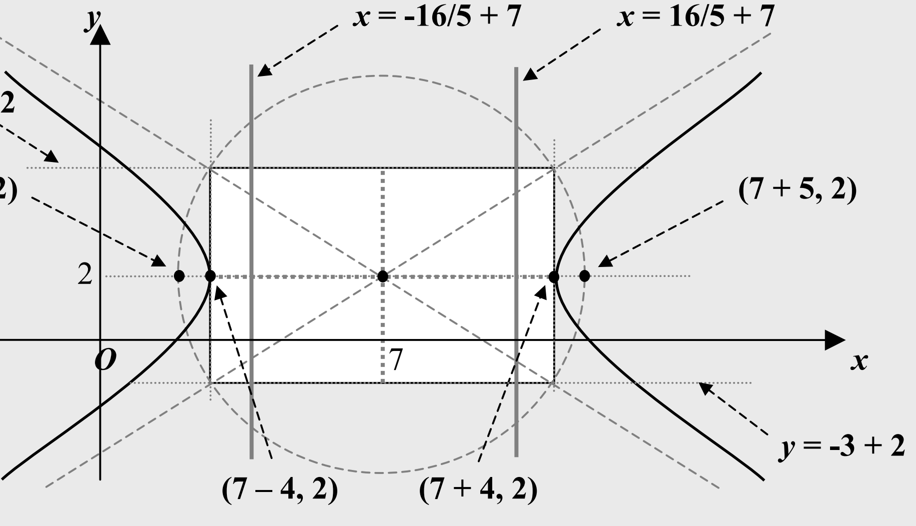
$$\frac{(x-7)^2}{4^2} - \frac{(y-2)^2}{3^2} = 1 \qquad \text{What hyperbola then is it?}$$

If L is the hyperbola, the hyperbola L is centered at (7, 2), and is horizontal, so the transverse axis is parallel to the x-axis, and is 8, and the conjugate axis is 6.

And assuming c is the focal distance, we get $c^2 = b^2 + a^2$, where $a = 4$, and $b = 3$, where a is half the transverse axis, and b is half the conjugate axis. So the focal distance is 5.

Thus next, assuming e is the eccentricity, we get $e = \dfrac{c}{a} = \dfrac{5}{4}$, where c is the focal distance, and a is half the transverse axis.

Fig. 3



And since the center is **(7, 2)**, the two asymptotes are $y - 2 = \pm\dfrac{3}{4}(x - 7)$.

Also, since the center is **(7, 2)**, the two directrices are $x = \pm\dfrac{16}{5} + 7$.

Next, since the hyperbola L is horizontal, the focal distance is 5 and the center is (7, 2), the two foci are (7 – 5, 2) and (7 + 5, 2).

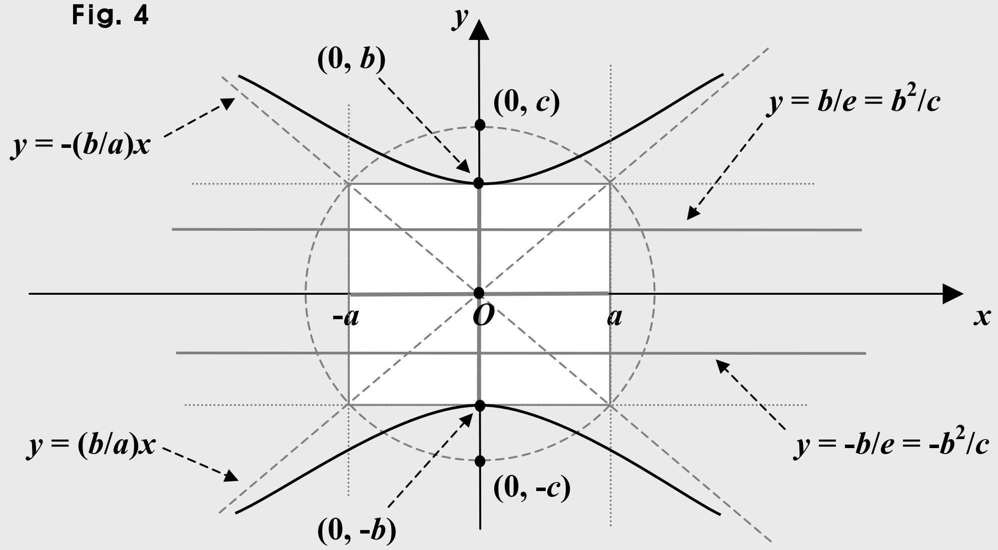
Also, since L is horizontal, the center is (7, 2), and the semi major axis, that is, half the transverse axis is 4, the two vertices are (7 – 4, 2) and (7 + 4, 2).

Next, putting a vertical hyperbola in a standard equation, we can get $\dfrac{y^2}{b^2} - \dfrac{x^2}{a^2} = 1$.

What hyperbola then is it?

If V is the hyperbola above, V is centered at **(0, 0)**, the origin, and since the coefficient of y^2 is positive, it is vertical, so the transverse axis is parallel to the y-axis, and is $2b$, and the conjugate axis is $2a$. And assuming c is the focal distance, we get $c^2 = b^2 + a^2$.

Fig. 4



So next, assuming e is the eccentricity, we get $e = \dfrac{c}{b}$, where c is the focal distance, and b is half the transverse axis. Also, we have $e > 1$, since $c > b$.

And since the center is **(0, 0)**, the two asymptotes are $y = \pm\dfrac{b}{a}x$.

Also, since the center is **(0, 0)**, the two directrices are $y = \pm\dfrac{b}{e} = \pm\dfrac{b^2}{c}$.

Next, since the hyperbola V is vertical, the center and foci share the same x-coordinate. So since the center is **(0, 0)**, and the focal distance is c, the two foci are **(0, *c*)** and **(0, -*c*)**.

Also, since V is vertical, the center and the vertices share the same x-coordinate, too. So since the semi major axis, that is, half the transverse axis is b, and the center is **(0, 0)**, the two vertices are **(0, *b*)** and **(0, -*b*)**.

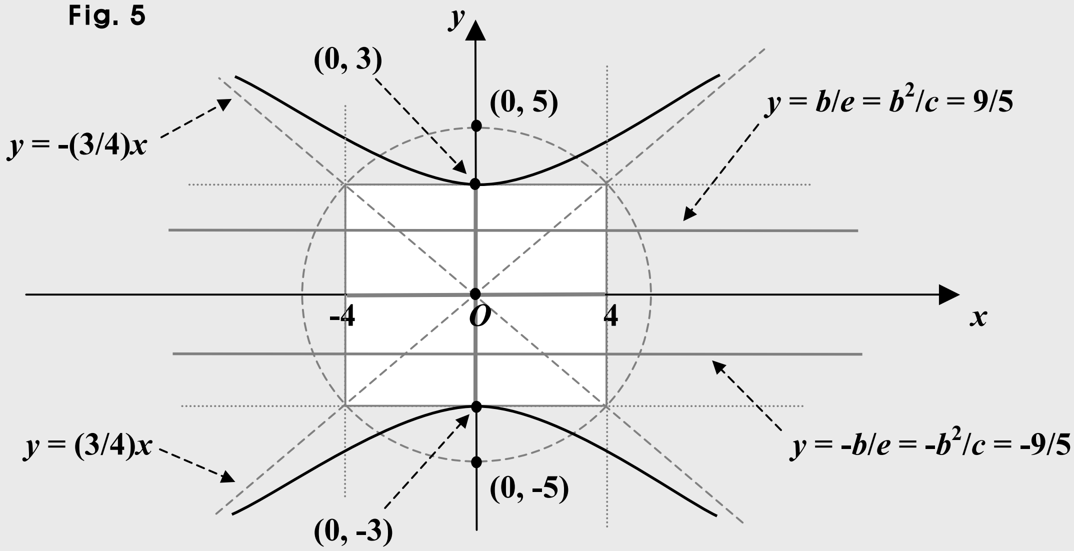
So for instance, putting a vertical hyperbola in an equation, we can get $\dfrac{y^2}{3^2} - \dfrac{x^2}{4^2} = 1$.

What hyperbola then is it?

If M is the hyperbola, M is centered at (0, 0), the origin, and is vertical, so the transverse axis is parallel to the y-axis, and is 6, and the conjugate axis is 8.
And assuming c is the focal distance, we get $c^2 = b^2 + a^2$, where $b = 3$, and $a = 4$, where b is half the transverse axis, and a is half the conjugate axis. So the focal distance is 5.

So next, assuming e is the eccentricity, we get $e = \frac{c}{b} = \frac{5}{3}$, where c is the focal distance, and b is half the transverse axis.

Fig. 5



And since the center is **(0, 0)**, the two asymptotes are $y = \pm\frac{b}{a}x = \pm\frac{3}{4}x$.

Also, since the center is **(0, 0)**, the two directrices are $y = \pm\frac{b}{e} = \pm\frac{b^2}{c} = \pm\frac{9}{5}$.

Next, since the hyperbola M is vertical, the center is (0, 0), and the focal distance is 5, the two foci are (0, 5) and (0, -5). (Note that vertical means the same x-coordinate.)

Also, since M is vertical, the center is (0, 0), and the semi major axis, that is, half the transverse axis is 3 and, the two vertices are (0, 3) and (0, -3).

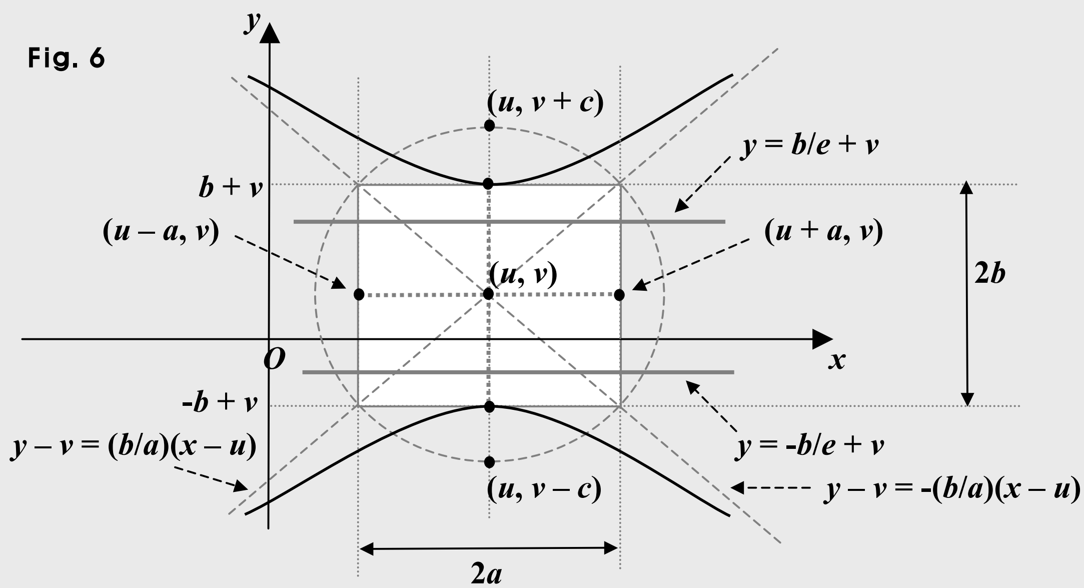
What then about the hyperbola as follows? $\frac{(y-v)^2}{b^2} - \frac{(x-u)^2}{a^2} = 1$

If J is the hyperbola, J is centered at (u, v), and is vertical, so the transverse axis is parallel to the y-axis, and is $2b$, and the conjugate axis is parallel to the x-axis, and is $2a$.

And assuming c is the focal distance, we get $c^2 = b^2 + a^2$.

So next, assuming e is the eccentricity, we get $e = \frac{c}{b}$, where c is the focal distance, and half the transverse axis is b. Also, we have $e > 1$, too, since $c > b$.

Fig. 6



And since the center is (u, v), the two asymptotes are $y - v = \pm\frac{b}{a}(x - u)$.

Also, since the center is (u, v), the two directrices are $y = \pm\frac{b}{e} + v = \pm\frac{b^2}{c} + v$.

Next, since J is vertical, the center and foci share the same x-coordinate. So since the center is (u, v), and the focal distance is c, the two foci are $(u, v + c)$ and $(u, v - c)$.

Also, since J is vertical, the center and the vertices share the same x-coordinate, too. So since the major radius, that is, the semi major axis is b, and the center is (u, v), the two vertices are $(u, v + b)$ and $(u, v - b)$.

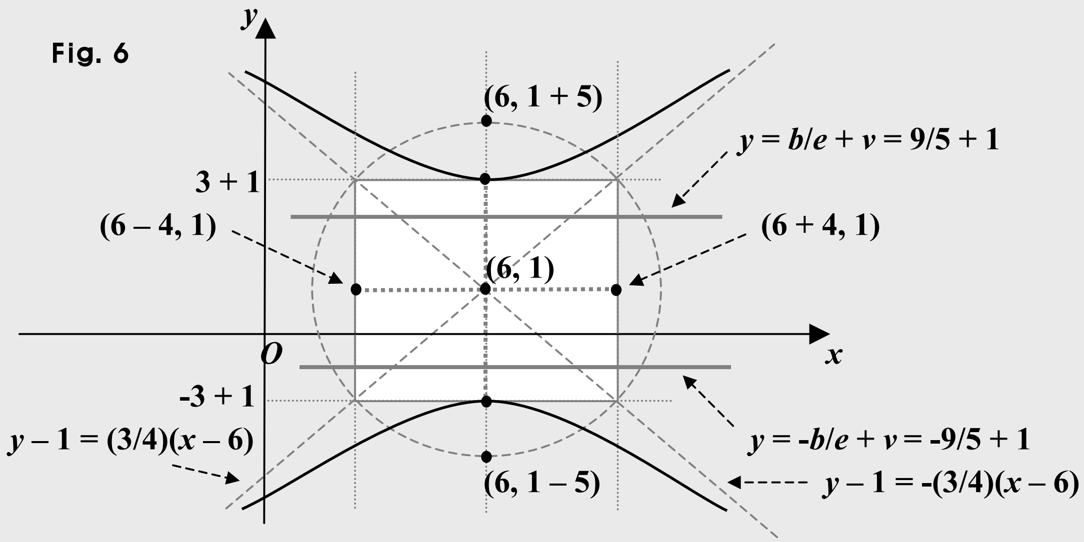
So for instance, putting a vertical hyperbola in an equation, we can get this:

$$\frac{(y-1)^2}{3^2} - \frac{(x-6)^2}{4^2} = 1. \qquad \text{What hyperbola then is it?}$$

If N is the hyperbola above, N is centered at (6, 1), and is vertical, so the transverse axis is parallel to the y-axis, and is 6, and the conjugate axis is 8.

And assuming c is the focal distance, we get $c^2 = b^2 + a^2$, where $b = 3$, and $a = 4$, where b is half the transverse axis, and a is half the conjugate axis. So the focal distance is 5.

Fig. 6



And since the center is **(6, 1)**, the two asymptotes are $\boldsymbol{y - 1 = \pm\frac{3}{4}(x - 6)}$.

Also, since the center is **(6, 1)**, the two directrices are $\boldsymbol{y = \pm\frac{b}{e} + v = \pm\frac{9}{5} + 1}$.

Next, since the hyperbola ***N*** is vertical, the center and foci share the same ***x***-coordinate. So since the center is (5, 1), and the focal distance is 5, the two foci are (6, 6) and (6, -4).

Also, since ***N*** is vertical, the center and the vertices share the same ***x***-coordinate. So since the center is (6, 1), and the semi major axis, that is, half the transverse axis is 3, the two vertices are (6, 3 + 1) and (6, -3 + 1).

And we call all the equations above standard equations for hyperbolas, and the hyperbolas above can be said to be perpendicular, because each main axis is perpendicular to a coordinate axis.
What then about an equation that can indicate an hyperbola perpendicular or not?

Putting an hyperbola in such an equation, we can put it the way as follows.

$$\boldsymbol{ax^2 + by^2 + cxy + ux + vy + w = 0}, \text{ where } \boldsymbol{4ab < c^2}$$

Note that ***a*** and ***b*** can be the same. And of course, ***a***, ***b***, ***c***, ***u***, ***v***, and ***w*** are constant.

Note however, for some values of the constants, the equation does not indicate an hyperbola, even if $4ab < c^2$.

For instance, $x^2 + 3y^2 + 4xy + 3x + 7y + 2 = 0$ is not an equation of a hyperbola but an equation of two lines. One is $x + 3y + 1 = 0$, and the other is $x + y + 2 = 0$.

That's because the expression $x^2 + 3y^2 + 4xy + 3x + 7y + 2$ can be factorized into the product of the two expressions $x + 3y + 1$ and $x + y + 2$.

That is to say that we can get $x^2 + 3y^2 + 4xy + 3x + 7y + 2 = (x + 3y + 1)(x + y + 2) = 0$. So we can get $x + 3y + 1 = 0$, or $x + y + 2 = 0$.

And doing the factorization, we can take $x^2 + 3y^2 + 4xy + 3x + 7y + 2 = 0$ as a quadratic equation with respect to x or y, and then, apply the quadratic formula.
What then, about an equation that is in the general form and indicates an hyperbola perpendicular only?

Putting an hyperbola in such an equation, we can put it the way as follows.

$$ax^2 + by^2 + ux + vy + w = 0, \text{ where } ab < 0$$

So the equation above indicates a hyperbola with a main axis perpendicular to the x-axis. Notice that it does not have the xy-term, that is, cxy in the equation shown earlier.

Note however, for some values of the constants, the equation does not indicate a hyperbola, even if $ab < 0$.

For instance, $x^2 - y^2 - 2x + 2y - 1 = 0$ and of course, $x^2 - 9y^2 - 1 = 0$ are hyperbola equations, but $x^2 - y^2 - 2x + 2y = 0$ and $x^2 - 9y^2 = 0$ are not.

That's because $x^2 - 9y^2 = 0 \Rightarrow (x - 3y)(x + 3y) = 0$ which indicates not a hyperbola but two lines, one is $x - 3y = 0$, and the other is $x + 3y = 0$.

And next, moving on to the next equation, we get

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

$= (x - 1)^2 - (y - 1)^2 = 0 \Rightarrow \{x - 1 - (y - 1)\}\{x - 1 + (y - 1)\} = (x - y)(x + y - 2) = 0,$

which indicates two lines, one is $x - y = 0$, and the other is $x + y - 2 = 0$.

So putting more specifically a perpendicular hyperbola in a general equation, we can put it the way as follows.

$ax^2 + by^2 + ux + vy + w = 0$, where $a > 0$, $b > 0$, and $bu^2 + av^2 \neq 4abw$

That's simply because, putting the equation into the sum of perfect squares, we get

$ax^2 + by^2 + ux + vy + w = ax^2 + ux + by^2 + vy + w$

$$= a(x^2 + \frac{u}{a}x) + b(y^2 + \frac{v}{b}y) + w = a(x^2 + \frac{u}{a}x + \frac{u^2}{4a^2} - \frac{u^2}{4a^2}) + b(y^2 + \frac{v}{b}y + \frac{v^2}{4b^2} - \frac{v^2}{4b^2}) + w$$

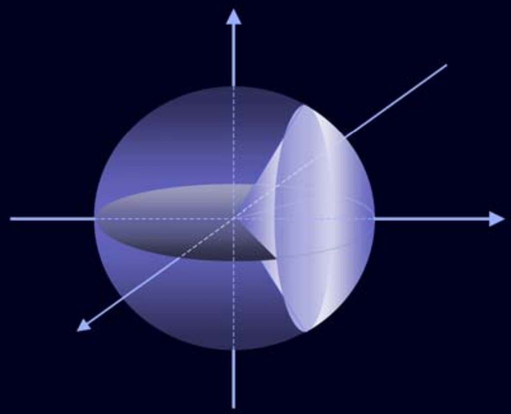
$$= a(x + \frac{u}{2a})^2 - \frac{u^2}{4a} + b(y + \frac{v}{2b})^2 - \frac{v^2}{4b} + w = 0$$

$$\Rightarrow a(x + \frac{u}{2a})^2 + b(y + \frac{v}{2b})^2 = \frac{u^2}{4a} + \frac{v^2}{4b} - w \neq 0.$$

Meanwhile, $\frac{u^2}{4a} + \frac{v^2}{4b} - w = \frac{bu^2 + av^2 - 4abw}{4ab}$.

So we get $\frac{u^2}{4a} + \frac{v^2}{4b} - w \neq 0 \Rightarrow \frac{bu^2 + av^2 - 4abw}{4ab} \neq 0 \Rightarrow bu^2 + av^2 - 4abw \neq 0$, for $ab \neq 0$.

Thus, we get $bu^2 + av^2 \neq 4abw$.



원뿔 곡선
Conic Sections

www.runmath.com