Paths of the Planets

Rachel W. Hall* Nigel Higson

December 9, 1998

Abstract

Assuming only Newton's Laws of Motion, we use plane geometry to argue that the paths of the planets are elliptical. The proof here follows a lecture of Richard Feynman, as related in *Feynman's Lost Lecture*, by Goodstein and Goodstein [1]. This is intended as a series of two or three lectures for honors calculus; extensive exercises are provided which students can work through on their own.

Kepler, Newton, and Feynman

Johanes Kepler's *Astronomia Nova*, published in 1609, contained two startling observations about the motion of the planets around the sun, which later became known as Kepler's first and second laws. Kepler stated that

1. the orbits of the planets are ellipses with the sun at one focus, and
2. the time it takes a planet to travel from one position in its orbit to another is proportional to the area swept out by a planet in that time.

Ten years later he published a third law:

3. the time it takes a planet to complete an entire orbit is proportional to the three-halves power of the longer axis of the ellipse.

Kepler's discoveries in celestial mechanics came a mere 66 years after the publication of Copernicus' *On the Revolutions of the Celestial Spheres*, which corrected the Aristotelean earth-centered universe. Copernicus proposed a sun-centered system in which the planets travel in circles. However, as more sophisticated instruments (these included giant measuring tools–the telescope was not invented until 1610!) and data became available in the later 16th century, Copernicus was due for a challenge. Kepler's new laws reconciled observation and theory.

Although Kepler's observations were correct, he was unable to explain *why* the planets behaved as they did. This task was completed by the scientific giant Isaac Newton as part of his *Principia*, published in 1687, in which he proposed his groundbreaking theory of motion. As a crowning touch, Newton used his laws of motion to deduce that the paths of the planets are elliptical. Moreover, he did this using only plane geometry, for although Newton had invented the powerful tools of calculus, he needed to explain his discoveries in a way that most scientists of his day would understand.

When preparing a lecture for a freshman physics class in 1964, the physicist Richard Feynman decided to prove the law of ellipses as Newton had–without refering to calculus. The proof here follows Feynman's lecture, as related in *Feynman's Lost Lecture*, by Goodstein and Goodstein [1]. Following a nineteenth-century argument by James Clerk Maxwell, Feynman deviates somewhat from Newton's proof, as Newton used some arcane properties of conic sections which are not known today, but still manages a completely geometrical proof.

In following Feynman's proof, some knowledge of conic sections, vector addition, Newton's laws of motion and gravitation, and Kepler's laws will be helpful.

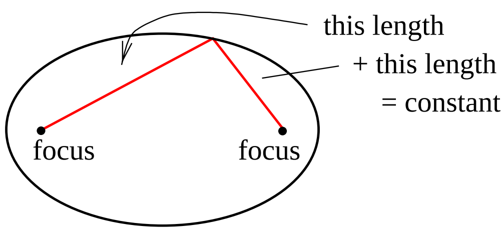
*Supported by a grant from the Fund for Excellence in Learning and Teaching

Law of Ellipses

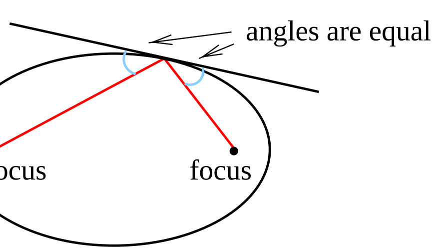
Three Definitions of an Ellipse

Conic sections were a serious subject of study for mathematicians of Newton's day–in fact, their knowledge of the subject was much more sophisticated than that of mathematicians today! We will start out with three of the basic definitions of an ellipse.

1. **Tack-and-String Definition** Pick any two points. These will be called the foci of the ellipse. The set of all points at which the sum of the distances to the two foci is some fixed number is an ellipse. Equivalently, affix two tacks to a board, tie each end of a string to a tack, and draw the curve created by a pencil which stretches the string taut.



2. **Reflection Property** Again, pick any two points as the foci. The curve whose tangent at any point forms equal angles with the lines to each focus will be an ellipse. Because of this property, a large elliptical wall forms a whispering gallery–anything spoken at one focus is reflected to the other focus.



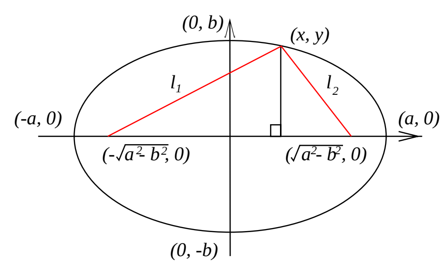
3. **Equation** For any fixed real numbers a and b, the set of points (x, y) in the plane which satisfy

$$\boxed{\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1}$$

will be an ellipse.

Exercise 1 Starting with the equation of the ellipse, we will find the coordinates of the foci, and verify the tack-and-string definition and the reflection property.

1. Graph the ellipse $\frac{x^2}{4} + \frac{y^2}{9} = 1$.
2. The semimajor axis of an ellipse connects the center with a point on the ellipse farthest from the center, while the semiminor axis connects the center with a point on the ellipse closest to the center. What do a and b represent in the equation $\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$?
3. Now let's find the coordinates of the foci.
   (a) Argue that, if the tack-and-string definition is true, the foci of the ellipse must lie on the semimajor axes.
   (b) Suppose $a > b$ and the foci of the ellipse $\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$ are located at $(c, 0)$ and $(-c, 0)$. Show that the length of the string in the tack-and-string definition must be $2a$.
   (c) Show that $c = \sqrt{a^2 - b^2}$.
   (d) What if $b > a$? Find the coordinates of the foci in this case.
4. Now we will verify the tack-and-string definition–that is, we'll show that if (x, y) is any point at which $\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$, then, assuming $a > b$, the distance from (x, y) to the focus $(-\sqrt{a^2 - b^2}, 0)$ plus the distance from (x, y) to the focus $(\sqrt{a^2 - b^2}, 0)$ is a fixed number (in fact, it is equal to $2a$).
   (a) Suppose l_1 is the distance from (x, y) to $(-\sqrt{a^2 - b^2}, 0)$ and l_2 is the distance from (x, y) to $(\sqrt{a^2 - b^2}, 0)$.

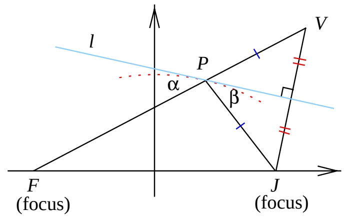


   Use the Pythagorean Theorem to show that

$$\begin{aligned} l_1^2 &= (\frac{x}{a}\sqrt{a^2 - b^2} + a)^2, \text{ and} \\ l_2^2 &= (\frac{x}{a}\sqrt{a^2 - b^2} - a)^2. \end{aligned}$$

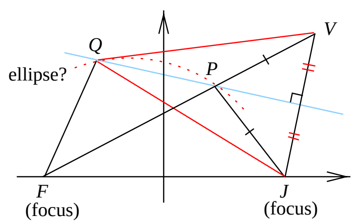
   (b) Conclude that $l_1 + l_2 = 2a$. Hint: l_2 is the positive square root of $(\frac{x}{a}\sqrt{a^2 - b^2} - a)^2$.
5. Starting with the tack-and-string definition we'll prove the reflection property–that is, we'll show that a line intersecting the ellipse which forms equal angles with the lines to the foci is a tangent to the ellipse.

(a) First, we will construct the line described by the reflection property.



Suppose point P is on an ellipse with foci F and J. Draw lines $\overline{PF}$ and $\overline{PJ}$. Now extend line $\overline{PF}$ by a distance equal to the length of $\overline{PJ}$, and label point V as shown. Construct the perpendicular bisector to $\overline{VJ}$ and call this line l. Argue that the angles labelled α and β are equal. Hence, l is the line described by the reflection property.

(b) A *tangent* to an ellipse is defined to be a line which intersects the ellipse in only one point. We will show that line l is a tangent.



i. Suppose line l intersects the ellipse in another point, say, point Q. Use the diagram above to argue that the lengths of $\overline{QV}$ and $\overline{QJ}$ are equal.

ii. The tack-and-string definition says that the sum of the lengths of $\overline{PF}$ plus $\overline{PJ}$ equals the sum of the lengths of $\overline{QF}$ plus $\overline{QJ}$. Explain, with reference to the diagram, that that cannot happen (unless, of course, P and Q coincide). Conclude that Q is actually *outside* the ellipse and l is tangent.

6. Is line l the same tangent defined by calculus? We'll show that it is.

(a) Use calculus to show that the slope of the tangent line to the ellipse at the point (x_0, y_0) is $-\frac{b^2 x_0}{a^2 y_0}$.

(b) (Hard) Show that the intersection of the line passing through (x_0, y_0) with slope $-\frac{b^2 x_0}{a^2 y_0}$ with the ellipse $\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$ is exactly at (x_0, y_0)–and no other points. The calculations are somewhat easier if you use the parametric form of the equation for an ellipse;

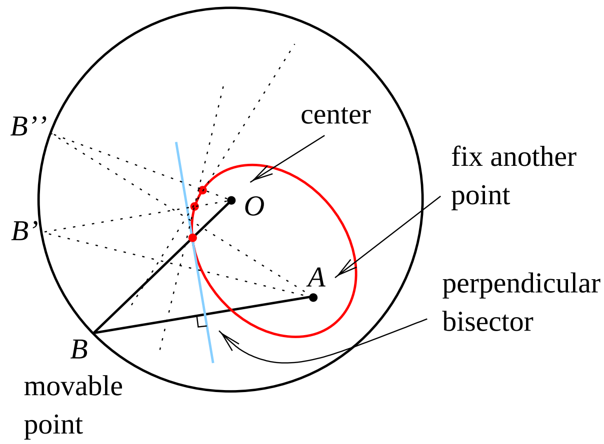
$$x = a\cos t, \text{ and } y = b\sin t.$$

(Why is the parametric form equivalent?)

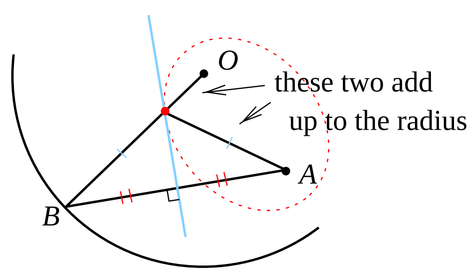
The Circle Construction of an Ellipse

We will show that the following construction is equivalent to the above definitions of an ellipse.

Draw a circle with center O. Fix a point A inside the circle which is not the center. Pick any point B on the circle and connect it to points O and A. Find the intersection of the perpendicular bisector to the line $\overline{AB}$ with the line $\overline{OB}$. Now allow B to move around the circle. The set of all such intersection points will form an ellipse.



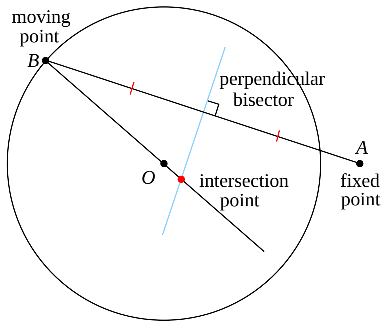
Why? As the diagram below shows, the circle construction is equivalent to the tack-and-string definition. The radius of the circle is equal to the length of the string. Notice also that the perpendicular bisector satisfies the reflection property and hence is tangent to the ellipse.



Exercise 2 We will investigate how the location of point A affects the figure.

1. First, try moving point A.

(a) With a ruler and compass, draw something like this:



Now let the point B move around the circle and find the intersection point for each B. Find enough points to be reasonably confident of the figure.

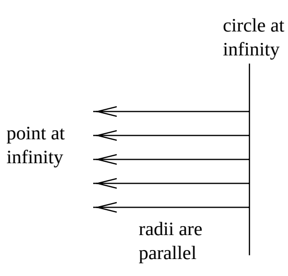
(b) Find the analog of the tack-and-string definition in this case. That is, describe the relationship between the lengths of the lines from the intersection point to points O and A.

(c) What can you say about the reflection property? Show that a ray of light emanating from the center of the circle will be reflected so that it appears to come from point A.

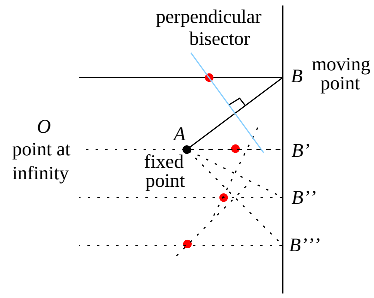
(d) What happens if point A is on the circle? At the center? Construct diagrams in each case.

2. So far, we've used this construction to come up with four conic sections: the line, circle, hyperbola, and ellipse. How can we get the parabola?

(a) Picture the circle getting bigger and bigger. A piece of an extremely large circle will appear to be a straight line, and lines connected to the faraway center of the circle will look parallel. We can think of the *circle at infinity* as a straight line. Its center is the *point at infinity* and the radii of the circle will be parallel lines.



Now suppose that O is the point at infinity, fix A, and construct the diagram. Here's a start.



(b) Describe how a ray of light emanating from point A will be reflected in the parabola.

(c) What if A were on the other side of the circle?

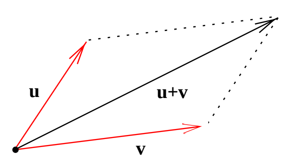
Newton's Dynamics[1]

Now that we have developed some of the properties of conic sections we'll need later, let's look at planetary orbit, as Newton described it. Newton's first two laws of motion are

1. (Principle of Inertia) If no forces are acting on a body, it will either stay at rest or continue travelling in a straight line at constant speed, and

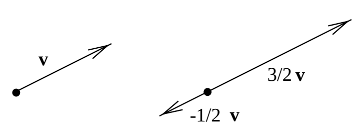
[1] We'll need to use vectors in this section. If you haven't seen vectors before, here's a brief introduction. A vector is a directed line segment, often used to represent some physical quantity such as force, with the following properties:

- Two vectors $\mathbf{u}$ and $\mathbf{v}$ emanating from the same point may be added by completing the parallelogram they describe, as shown.



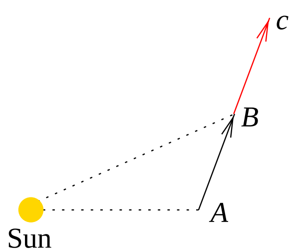
The vector $\mathbf{u}+\mathbf{v}$ is the diagonal emanating from the same point.

- The length of $\mathbf{v}$ is denoted $\|\mathbf{v}\|$.
- Multiplication of $\mathbf{v}$ by a positive real number k results in a vector $k\mathbf{v}$ which has the same direction as $\mathbf{v}$ and length $k\|\mathbf{v}\|$. If k is negative, $k\mathbf{v}$ has the opposite direction to $\mathbf{v}$ and length $|k|\|\mathbf{v}\|$.

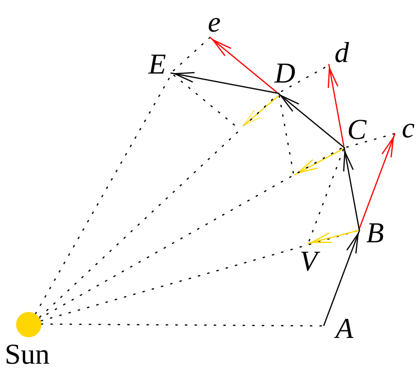


2. the change in motion of a body is proportional to and in the direction of any force acting on the body.

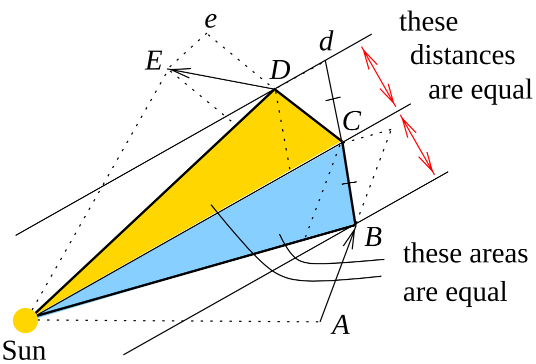
Using these two laws, Newton showed why Kepler's observation that the planets sweep out equal areas in equal times is true. Suppose a planet travels from A to B in a certain unit of time.



If no force were acting on the planet, it would continue on to point c after another time unit has elapsed. However, the gravitational force directed towards the sun pulls the planet towards point V. The sum of the intertial force directing the planet towards c plus gravity directing it to V causes the planet to arrive at point C. We can continue this process to find points D, E, etc. Here's a diagram similar to the one Newton drew.



Let's see why the areas covered in each unit of time are equal. Take any two successive triangles, extend their common side, and draw parallels as shown:

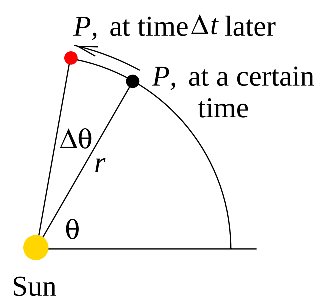


Since C is the midpoint of $\overline{Bd}$, the distances between the parallel lines are equal. The two shaded triangles have the same base and height, and hence the same area. Since each triangle represents the area swept out by the planet in one unit of time, Newton's mechanics have proved Kepler's second law!

The Inverse Square Law

Note that we haven't used the fact that the gravitational pull of the sun is inversely proportional to the square of the radius yet. We'll use it now, to compute the changes in the velocity vector.

The diagrams in the last section represent the orbit as a succession of straight lines, rather than a smooth curve. If we let the unit of time Δt go to zero, the diagram looks something like this.



P, the position of the planet, is represented in polar coordinates by the pair (r, θ), which are both functions of time, t. $\Delta\theta$ represents the angle traversed in time Δt.

At this point, Feynman deviates from Newton's argument. Instead of breaking the orbit up into equal-time pieces, he breaks it into equal-*angle* pieces. If the angles are small, the areas swept out are approximately proportional to the square of the radius, that is,

$$\text{area} \approx \text{constant} \cdot r^2.$$

Let $\mathbf{v}$ be the velocity vector, and use $\Delta\mathbf{v}$ to denote the change in $\mathbf{v}$ during the time Δt. Note that Δt is now the time taken to traverse the angle $\Delta\theta$. $\Delta\mathbf{v}$ will be a vector pointing towards the sun. The length of $\Delta\mathbf{v}$, denoted $\|\Delta\mathbf{v}\|$ is the total change in the planet's velocity on the interval Δt. Newton's inverse square law tells us that

$$\|\Delta\mathbf{v}\| = \text{constant} \cdot \frac{1}{r^2}\,\Delta t.$$

Since r^2 is proportional to area,

$$\|\Delta\mathbf{v}\| \approx \text{constant} \cdot \frac{\Delta t}{\text{area swept out in } \Delta t}.$$

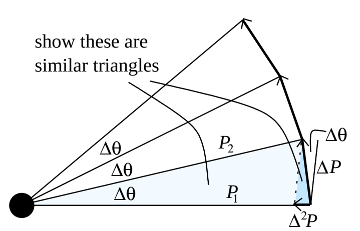
But the area swept out in Δt is just a constant multiple of Δt, by Kepler's second law. Therefore we can cancel area against time, and

$$\|\Delta\mathbf{v}\| \approx \text{constant}$$

that is, the change in velocity with respect to change in *angle* is constant.

Exercise 3 We don't need to take the inverse-square law of gravitation for granted–we can derive it, as Newton did, from Kepler's third law! We're going to prove the inverse square law for circular orbits.

1. Assuming the action of gravity does not vary on a circular orbit, the planet will travel with constant speed. The orbit can be approximated by a regular polygon. Let the vector $\mathbf{P}$ represent position, $\Delta\mathbf{P}$ change in position, and $\Delta^2\mathbf{P}$ change in $\Delta\mathbf{P}$.



Show that the exterior angle marked is equal to $\Delta\theta$, and use that to argue that the shaded triangles are similar. Conclude that

$$\frac{\|\Delta^2\mathbf{P}\|}{\|\Delta\mathbf{P}\|} = \frac{\|\Delta\mathbf{P}\|}{\|\mathbf{P}\|}.$$

2. Assume $\Delta\theta$ is small. Let r be the radius of the circle, and T the time taken to complete one revolution, and approximate $\|\mathbf{P}\|$ and $\|\Delta\mathbf{P}\|$ in terms of r , T, and Δt. Show that

$$\|\Delta^2\mathbf{P}\| \approx \frac{4\pi^2 r(\Delta t)^2}{T^2}.$$

Since Kepler's third law states that

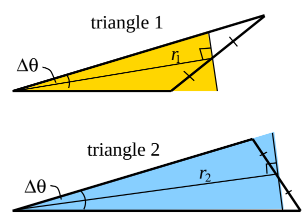
$$T = \text{constant} \cdot r^{\frac{3}{2}},$$

then

$$\|\Delta^2\mathbf{P}\| \approx \text{constant} \cdot \frac{(\Delta t)^2}{r^2}.$$

3. Remember that we're actually trying to find the length of $\Delta\mathbf{v}$. The velocity vector $\mathbf{v}$ equals $\frac{1}{\Delta t}\Delta\mathbf{P}$, so $\Delta\mathbf{v}$ is $\frac{1}{\Delta t}\Delta^2\mathbf{P}$. Conclude that acceleration due to gravity (that is, $\frac{\|\Delta\mathbf{v}\|}{\Delta t}$) is inversely proportional to r^2.

Exercise 4 We'll prove the assertion that area swept out over equal angles is proportional to the square of the radius. Assume that $\Delta\theta$ is small enough that it makes sense to approximate the areas by triangles.



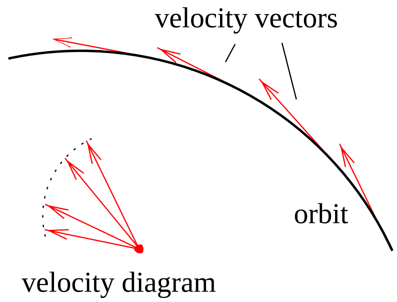
Here are two pieces of the orbit. The radius is the length of the bisector of $\Delta\theta$. Construct perpendiculars to the radius, as shown, and argue that the shaded triangles are similar. Moreover, show they have the same area as the original triangles. Then show

$$\frac{\text{area of triangle 1}}{\text{area of triangle 2}} = \frac{r_1^2}{r_2^2}$$

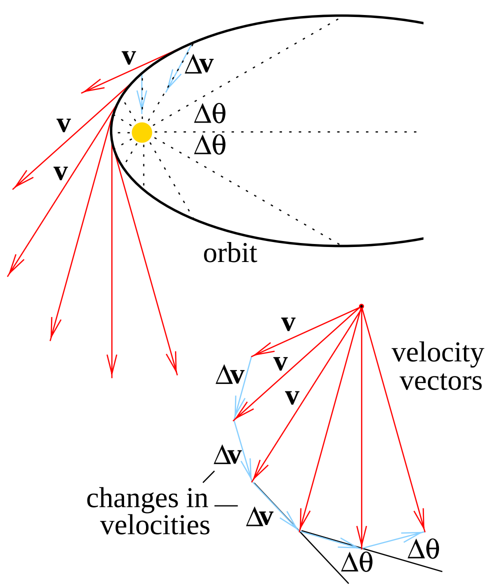
and conclude that area is proportional to the square of the radius.

The Velocity Diagram (Hodograph)

Any point along the orbit corresponds to a velocity vector which is tangent to the orbit and whose length signifies the speed at which the planet is travelling. Translate all the velocity vectors to a point. This is called a velocity diagram or hodograph.



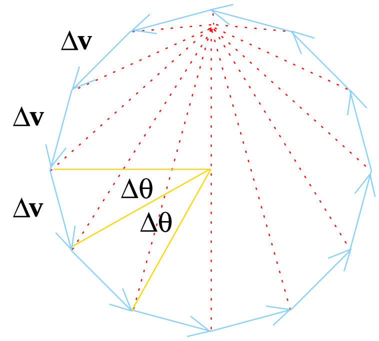
What is the shape of the figure formed by the ends of the arrows in the velocity diagram? Let's break the orbit up into equal-*angle* pieces, and compute the velocity vector after each $\Delta\theta$.



Note that

- Each $\Delta\mathbf{v}$ has the same length, since $\|\Delta\mathbf{v}\| =$ constant.
- Each $\Delta\mathbf{v}$ points $\Delta\theta$ beyond the previous one, since $\Delta\mathbf{v}$ points towards the sun in the original diagram.

So, if the orbit is closed, then the figure created by the $\Delta\mathbf{v}$'s is a regular polygon with exactly $\frac{360^\circ}{\Delta\theta}$ sides.



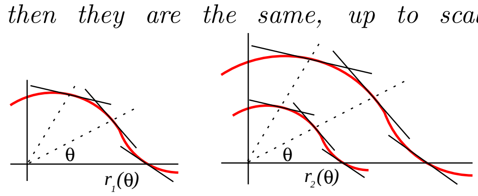
Notice that the angles created by the lines between the ends of the velocity vectors and the center of the polygon are equal to $\Delta\theta$. This means that the angle swept out from the *center* of the velocity diagram is equal to the angle swept out from the sun. As $\Delta\theta \to 0$, we get a circle.

Exercise 5 If the orbit of the planet is *not* closed, is the velocity diagram still a circle? Suppose the orbit is a hyperbola. For some values of θ, a ray from the sun will not intersect the orbit (Why?). If we break the orbit up into equal angle pieces, does the relationship $\|\mathbf{v}\| =$ constant still hold? Sketch the velocity diagram for a hyperbolic orbit.

The Shape of the Orbit

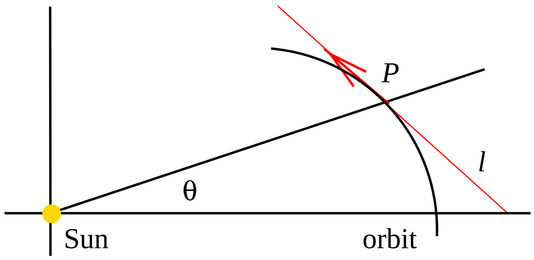
Now we know the shape of the velocity diagram, but we actually started out trying to find the shape of the orbit. We're going to need the following:

The Tangent Principle *If two curves (in polar coordinates) $r_1(\theta)$ and $r_2(\theta)$ have the same tangent at every θ, then they are the same, up to scaling.*

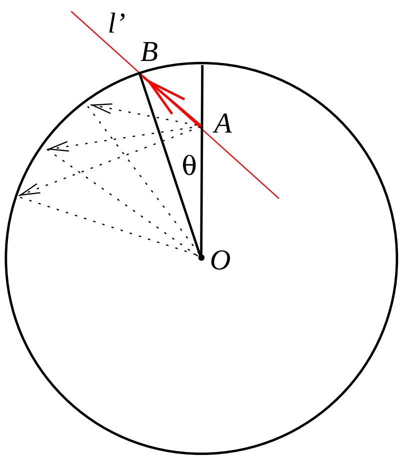


So, if we can show that the *tangent* to the orbit at every point is the same as the tangent to an ellipse, then the orbit itself is an ellipse.

Here's a planet orbiting the sun.

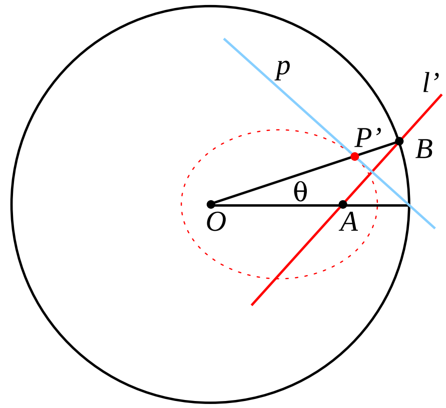


Draw the velocity diagram for the orbit by translating all velocity vectors to a point A.



Suppose l is the tangent line to the orbit at point P. Since the velocity vector at P is also tangent to the orbit, the line l' in the velocity diagram is parallel to line l. Let B be the intersection of line l' with the circle, and O the center of the circle. As we have remarked before, the angle swept out in the orbit is equal to the angle swept out from the *center* of the velocity diagram. Hence angle AOB will be equal to θ.

Now rotate the entire diagram clockwise by $90°$.



Construct a perpendicular bisector p to the segment $\overline{AB}$. Call the intersection of p with $\overline{OB}$ P'. Notice that what we have just done is exactly the circle construction of an ellipse! So P' is a point on an ellipse and p is the tangent to the ellipse. Since line l' is *perpendicular* to our original line l, p is parallel to it. That is, *the tangent to the ellipse constructed above and the tangent to the orbit agree at every* θ. By the tangent principle, the orbit of the planet must be an ellipse.

References

[1] David L. Goodstein and Judith R. Goodstein. *Feynman's Lost Lecture : the motion of planets around the sun.* W. W. Norton & Company, New York, 1996.