Kepler’s Equal Areas Law

29 March 2017

Jim Stevenson

I have long been fascinated by Newton’s proof of Kepler’s Equal Areas Law and wanted to write about it. Of course, others have as well, but I wanted to emphasize an aspect of the proof that supported my philosophy of mathematics.

Before I get to Newton, however, I wanted to discuss how Kepler himself justified this law, since his approach has a number of fascinating historical aspects to it. I have previously discussed Kepler’s ellipse ([5]) and in the process of doing that research, I came across a number of articles about how Kepler arrived at his equal areas law. One notable result is that even though now we call the idea that a planet orbits the Sun in an elliptical path with the Sun at one focus, Kepler’s First Law, and the idea that the line from the Sun to the planet sweeps out equal areas in equal times, Kepler’s Second Law, Kepler actually discovered these laws in reverse order.

Kepler (1571-1630) presents his research, theorizing, and calculations in his marvelous and expansive book *Astronomia Nova* (1609), the *New Astronomy* (see [3] for a nice detailed visual explanation of the ideas in the book). What makes the book so fascinating (and equally hard to read — aside from the Latin) is that rather than just present his final results, as would happen in current research, Kepler describes the lengthy path of his reasoning which included all the blind allies and misconceptions. In a way, the book was a massive journal of his research. So he began with the assumptions of his times bequeathed from the Greeks over 2000 years before and codified by Ptolemy (100 – 170 AD) in his book, *The Almagest*. It basically claimed the earth was the center of the universe and all the planets, Sun, and stars rotated around it attached to concentric spheres, so that their motion was circular. Ptolemy acquiesced to the already observed irregularities in the motion of the planets (especially the retrograde motion) by including the idea of epicycles, circles that rotated on other circles. So he preserved the all-important circular motion, but at the cost of added complexity.

Thus Kepler began his researches with this dependence on circular motion. He did deviate in one very important respect: he followed the Copernican innovation of having the earth and the other planets orbit the Sun rather than having the Sun and the planets other than the earth orbit the earth. One of the other Greek principles was that the planets were supposed to orbit the circles at uniform speeds. But Kepler already knew that when a planet was close to the Sun it moved faster than when it was further away, so that the motion was not uniform after all. He began to consider all kinds of manipulations to try to reconcile these behaviors and Greek theories.

Kepler’s Derivation

Probably the clearest explanation of Kepler’s equal areas law is given by Peter Barker and Bernard R. Goldstein, which I excerpt here (Barker et al. [1] pp.67-70). I have omitted the original footnotes for readability, replaced the figures and figure numbers with my colored versions, and added my own emphasis for later discussion.

> … In chapter 32, Kepler presented a geometrical argument to prove that 'the swiftness at perihelion and the slowness at aphelion are proportioned approximately as the lines drawn from the centre of the world to the planet'. The centre of the world, for Kepler, is the Sun. Strictly, this proof is applicable only at perihelion and aphelion, as it depends upon similar triangles that cease to be similar for positions outside the apses. However this mathematical demonstration is a crucial step forward. Here Kepler demonstrated a result taken as an axiom by earlier writers. Although he

demonstrated the result only in a special case, he noted that the result applied as a good approximation throughout the orbit. Kepler saw the result not as an axiom but as the outcome of the operation of physical causes, which he proceeded to discuss.

The next six chapters of the *Astronomia Nova* introduce the idea of a *virtus motrix* generated by the Sun and responsible for the overall pattern, if not the details, of planetary motion. The nature of Kepler's *virtus motrix* remain controversial. …

Regardless of which of these readings is correct, near the solar equator the force diminishes as if confined to a plane. Taking the emanation from the Sun as a fixed quantity, its intensity will diminish in proportion to the circumference of the successively larger circles it crosses while it spreads itself into the surrounding ether. Hence, wherever the emanation encounters a planet it exerts a force inversely proportional to the planet's distance from the Sun.

For Kepler the application of a force creates a velocity, overcoming the natural tendency of an object to remain at rest. Continued motion requires continued application of force. On its own the solar *virtus motrix* would move the planets in circles, at constant speed, centred on the Sun. But Kepler has been at pains to establish that if the planets are taken to have circular orbits, they are eccentric to the Sun, and that the linear velocity of the planet varies throughout this motion. The distance-velocity relation may describe the variation in velocity. Kepler has shown in chapter 32 that it applies accurately at the apses. However, this relation does not explain why the planet should approach and recede from the Sun at different points in its motion. Kepler therefore introduced a second force or *vis insita,* located in the planet itself. This force explains the motion of the planet along the radius vector from the Sun, and combines with the solar *virtus motrix* to produce the planet's trajectory.

4. The area law

Kepler immediately applied these physical ideas to calculate planetary positions in the final chapter of Part III, chapter 40. … To clarify his discussion we introduce Figure 1: consider the area A_1 defined by lines drawn from the Sun to the ends of a small arc s, representing the motion of a planet. Note that the Sun is not the centre of the circle on which the planet is supposed to move (but rather it is the physical basis for its motion, through the *virtus motrix).* Then, at aphelion and perihelion, where the motion of the planet is perpendicular to the radius vector drawn from the Sun, the area of this small triangle A_1 will be proportional to the product of the arc s and the length of the radius vector d_1 [from the sun]. Strictly this relation will be valid only at the aphelion and perihelion (as Kepler was aware).

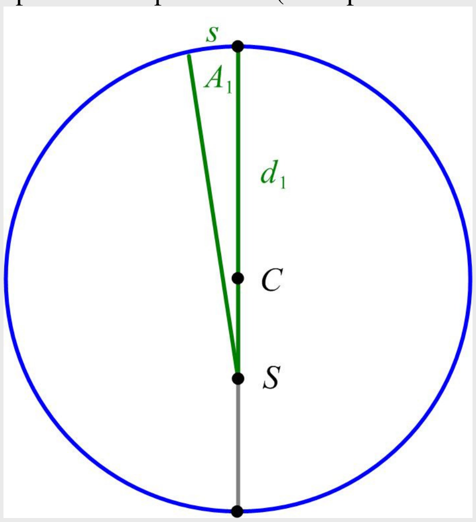


Figure 1. The area law at aphelion. Note that the Sun is located at point S.

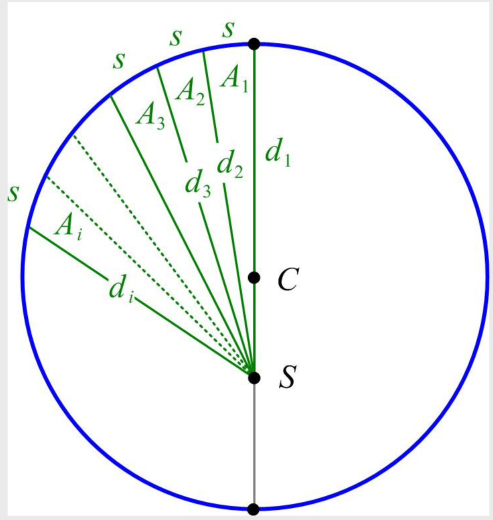


Figure 2. The derivation of the area law.

Next construct the motion of the planet around a portion of its orbit, by adding small segments like that already defined (see Figure 2). We then have i arcs of length s, and we seek the variable time intervals that correspond to each of these equal arc-lengths. Kepler calls these time intervals *morae:* in effect, they represent the time it takes a planet to move a unit distance along its trajectory. Following the proportionality already established, the sum of the areas $A_1, A_2, A_3, \ldots, A_i$ will be proportional to the sum of $sd_1; sd_2, sd_3; \ldots, sd_i$; where $d_1, d_2, d_3, \ldots, d_i$; are the lengths of the corresponding radial vectors, i.e.

$$A_1 + A_2 + \ldots + A_i \propto sd_1 + sd_2 + \ldots + sd_i \tag{1}$$

… Now Kepler believed, as we have already seen, that the linear velocity of a planet varies and is inversely proportional to its distance from the Sun. Hence, in each term in the above series we may replace the distance d_j (where $1 \leq j \leq i$) by the reciprocal of the corresponding velocity v_j producing a quotient s/v_j. Each of these quotients represents the distance travelled by the planet along a small portion of its orbit divided by the velocity with which it traverses that portion of the orbit, and thus defines the time taken to traverse that portion of the orbit. That is

$$A_1 + A_2 + \ldots + A_i \propto \frac{s}{v_1} + \frac{s}{v_2} + \ldots + \frac{s}{v_i} \tag{2}$$

$$\propto t_1 + t_2 + \ldots + t_i \tag{3}$$

Therefore, the ratio of the sum of the areas A_j making up a given segment of the orbit, to the area of the whole orbit (A), will be equal to the ratio of the sum of the corresponding times t_j, to the time required for the planet to complete one orbit, that is the period of the planet (T). That is

$$\frac{A_1 + A_2 + \ldots + A_i}{A} = \frac{t_1 + t_2 + \ldots + t_i}{T} \tag{4}$$

Now let us define

$$\alpha_i = A_1 + A_2 + \ldots + A_i$$

and

$$\tau_i = t_1 + t_2 + \ldots + t_i$$

Then, we can rewrite equation (4) as

$$\frac{\alpha_i}{A} = \frac{\tau_i}{T} \tag{5}$$

The correlation established here between areas and time intervals is the same one we recognize, for the case of an elliptical orbit with the Sun at one focus, as the Second Law of Planetary Motion.[1]

Now there are a number of remarkable things about this presentation.

No Law of Inertia

The sentences above, "For Kepler the application of a force creates a velocity, overcoming the natural tendency of an object to remain at rest. Continued motion requires continued application of force." means that at this point Kepler subscribed to the Scholastic idea of "impetus" and not to the "law of inertia", which was codified in Newton's First Law of Motion in 1687, after an earlier hint by Galileo (1564-1642) in 1613 and closer notion by Rene Descartes (1596-1650) in 1644. The Law of Inertia claims that a body at rest will remain at rest or a body moving along a straight line at constant

---

[1] Notice that equation (5) implies that $\Delta\alpha_i = (A/T)\Delta\tau_i$ so that equal areas are swept out in equal times. Since Kepler's argument did not make explicit use of the circular orbit, it is inferred that it would work for a non-circular orbit as well, such as an ellipse.

speed will continue to move along that line at that constant speed *unless* it is acted upon by a force, in which case it will change its state of rest or motion. The idea that objects move because they were *already* moving was a huge change of viewpoint. So bodies could be moving without experiencing a force, contrary to Kepler's statement.

Force between the Sun and Planets causes Planets to Move

As shown in Figure 2 above, at this stage of Kepler's thinking the planet moved in a circle (required by Ptolemy) but not centered on the Sun. The idea at the time was that the motion of planets was based on the center of their circular orbits. But assuming a heliocentric theory meant that the sun became the main actor and so it must be the agent of their motion. Kepler then posited physical forces to bring about this motion. This was quite an innovation by Kepler and is clearly a precursor to Newton's Theory of Gravitation.

Speed of Planets Inversely Proportional to Distance from Sun

By arguing that the force between the Sun and a planet was restricted to a plane, he could claim its intensity fell off as the inverse of the distance rather than the distance squared. Linking the force directly to the speed along the circle meant the speed varied inversely as the distance from the Sun. Of course, later Newton would tie the force directly to the acceleration of the planet, not its velocity, and arguing in three dimensions, the intensity of the force fell of as the square of the distance from the Sun (since the surface area of a sphere of constant intensity is proportional to the radius squared). Thus Newton had the acceleration of the planet be inversely proportional to the square of the distance from the Sun.

Kepler Anticipates Calculus

Kepler credits to Archimedes (c.287 – c.212 BC) the idea of taking small increments along the orbit of the planet, making a sequence of approximations, and then imagining the result when the increments are shrunk infinitesimally. This approach anticipates the integral calculus toward the end of the century.

There are, of course, numerous problems with Kepler's derivation, especially the approximations, but it is an impressive line of reasoning nevertheless.

Newton's Derivation

We turn now to the way Newton (1643-1727) established the equal areas law. Newton does employ ideas from the newly minted differential and integral calculus, but the heart of the demonstration is the gold standard at the time of Euclidean geometric reasoning.[2] We shall do the same, since it does make the result amazingly obvious.

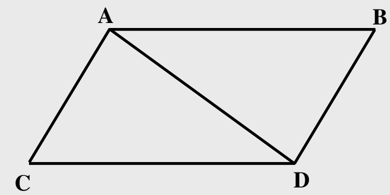
Just for the record, here is Newton's own proof from the *Principia* ([4] pp.83-104) (BOOK I PROPOSITION I). We have also included his statement of the law of inertia (LAW I), as well as the parallelogram law of the vector addition of forces (COROLLARY I).

> AXIOMS, OR LAWS OF MOTION.
>
> LAW I. [Law of Inertia]
>
> *Every body perseveres in its state of rest, or of uniform motion in a right line, unless it is compelled to change that state by forces impressed thereon. …*

[2] Based on Euclid's book *The Elements* written about 300 BC.

… CORROLLARY I.

A body by two forces conjoined will describe the diagonal of a parallelogram, in the same time that it would describe the sides, by those forces apart.



If a body in a given time, by the force M impressed apart in the place A, should with an uniform motion be carried from A to B ; and by the force N impressed apart in the same place, should be carried from A to C ; complete the parallelogram ABCD, and, by both forces acting together, it will in the same time be carried in the diagonal from A to D. For since the force N acts in the direction of the line AC, parallel to BD, this force (by the second law) will not at all alter the velocity generated by the other force M, by which the body is carried towards the line BD. The body therefore will arrive at the line BD in the same time, whether the force N be impressed or not ; and therefore at the end of that time it will he found somewhere in the line BD. By the same argument, at the end of the same time it AY ill be found somewhere in the line CD. Therefore it will be found in the point D, where both lines meet. But it will move in a right line from A to D, by Law I.[3] …

BOOK I. OF THE MOTION OF BODIES

SECTION II. Of the Invention of Centripetal Forces.

PROPOSITION I. THEOREM 1.

The areas, which revolving bodies describe by radii drawn to an immovable centre of force do lie in the same immovable planes, and are proportional to the times in which they are described.

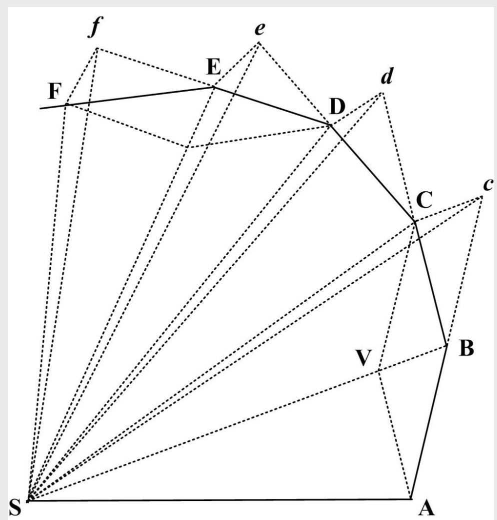


Figure 3 Original *Principia* Diagram

For suppose the time to be divided into equal parts, and in the first part of that time let the body by its innate force describe the right line AB In the second part of that time, the same would (by Law I.), if not hindered, proceed directly to c, along the line Bc equal to AB ; so that by the radii AS, BS, cS, drawn to the centre, the equal areas ASB, BSc, would be described. But when the body is arrived at B, suppose that a centripetal force acts at once with a great impulse, and, turning aside the body from the right line Bc, compels it afterwards to continue its motion along the right line BC. Draw cC parallel to BS meeting BC in C ; and at the end of the second part of the time, the body (by Cor. I. of the Laws) will be found in C, in the same plane with the triangle ASB. Join SC, and, because SB and Cc are parallel, the triangle SBC will be equal to the triangle SBc, and therefore also to the triangle SAB. By the like argument, if the centripetal force acts successively in C, D, E. &c., and makes the body, in each single particle of time, to describe the right lines CD, DE, EF, &c., they will all lie in the same plane : and the triangle SCD will be equal to the triangle SBC, and SDE to SCD, and SEF to SDE. And therefore, in equal times, equal areas are described in one immovable plane : and, by composition, any sums SADS, SAFS, of those areas, are one to the other as the times in which they are described. Now let the number of those triangles be augmented, and their

[3] The idea of the parallelogram law is similar to the effect of trying to row across a flowing river. The combination of the motion of the forward rowing and current of the stream results in landing on the opposite shore further downstream than originally intended.

breadth diminished *in infinitum* ; and (by Cor. 4, Lem. III.) their ultimate perimeter ADF will be a curve line : and therefore the centripetal force, by which the body is perpetually drawn back from the tangent of this curve, will act continually ; and any described areas SADS, SAFS, which are always proportional to the times of description, will, in this case also, be proportional to those times. Q.E.D.

Newton's exposition should be clear in itself, but we shall make it a bit more explicit. A number of others have already presented Newton's demonstration in all its simplicity. The earliest, and one of the best, I recall was by Richard Feynman in the Messenger Lectures he gave in 1964 at Cornell University. I was lucky enough to attend these lectures, which have now been made available on Youtube ([2]). They were also captured in a book, *The Character of Physical Law*. The particular section on Newton's proof of the equal area law is in the second lecture, "The Relation of Mathematics to Physics" ([2] pp.40-43).

The idea is to realize that if the Sun exerted no force on the moving planet, then the planet would proceed in a straight line at constant speed, according to the Law of Inertia (LAW I). As seen in Figure 4, the line joining the Sun to the planet would sweep out equal areas in equal times because all the triangular wedges have the same altitude and their bases, representing the distances traveled in equal intervals of time, are the same length.

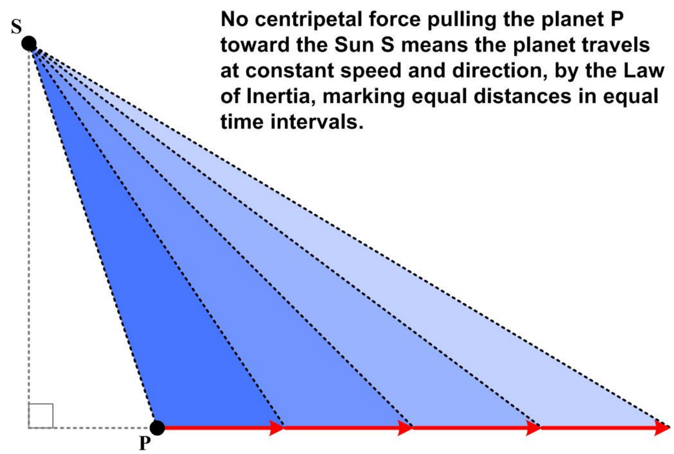


Figure 4 Equal Areas are swept out at equal time intervals with straight line uniform motion.

We now consider the effect of a centripetal force pulling the planet toward the Sun. With Newton we will consider the force acts in impulses at discrete, equally spaced small time intervals, between which times the planet travels according to the Law of Inertia. Figure 5, based on Newton's original figure (Figure 3), illustrates the situation. We have added (red) vectors to indicate the direction and intensity of the impulse forces. The planet is moving from A to B in the first interval of time and intends to move straight ahead according to the Law of Inertia to *c* in the next interval of time. But the planet experiences the impulse force at B, which, when combined with its motion to *c* according to the Parallelogram Law, causes the planet to move to C. At C the planet experiences another impulse force that, via the Parallelogram Law, redirects it to D, and so on. As the time intervals chosen are taken to be shorter and shorter, the broken line trajectory of the planet approaches the path of a smooth curve, the orbit of the planet under the effect of a continuous centripetal force.

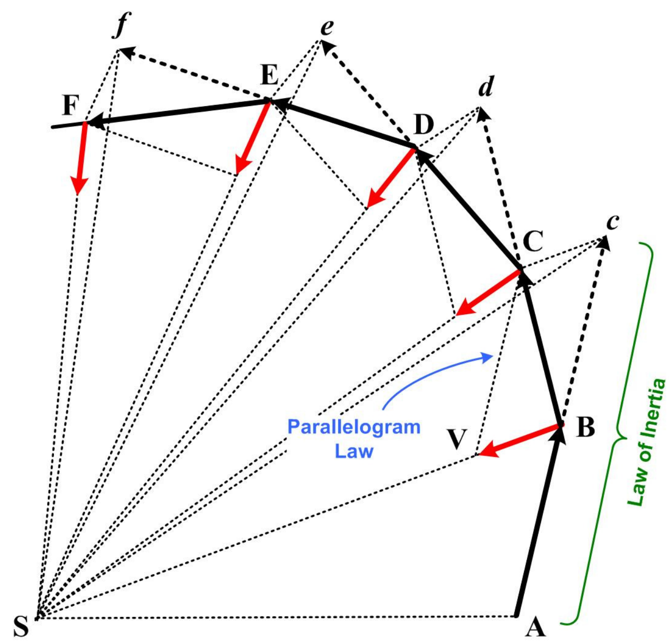


Figure 5 Curved Motion from Centripetal Force and Parallelogram Law

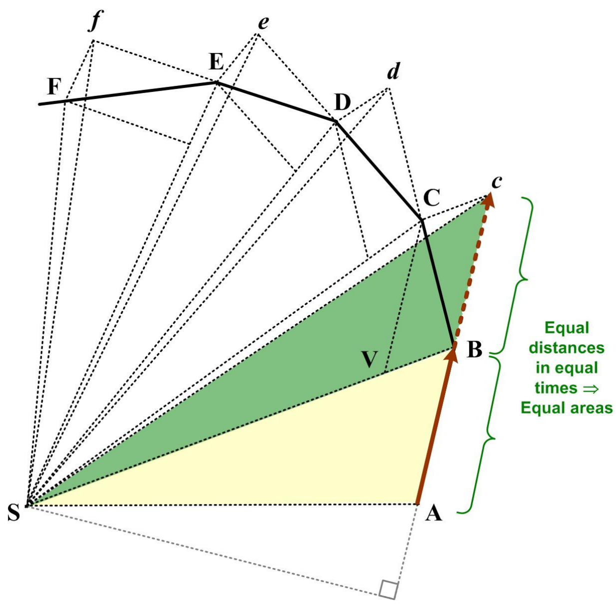


Figure 6 Constant Straight Line Motion Implies Equal Areas

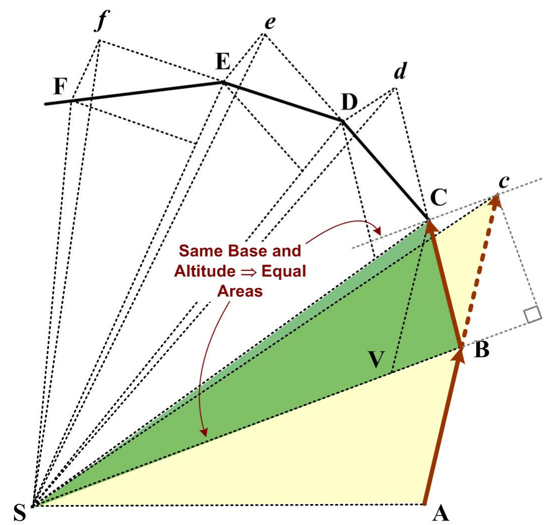


Figure 7 Same Base and Altitudes Implies Equal Areas

Now we consider what happens to the line from the Sun to the planet as it traverses its orbit. Again we consider the discrete equally spaced intervals of time as shown in Figure 6. Again the planet travels according to the Law of Inertia along a straight line from A to B. Suppose it continues as intended on to *c*. Then as shown in Figure 6 we have the situation represented in Figure 4 where the two triangular areas swept out by the line from the Sun to the planet have the same area. Figure 7 shows the situation after the planet experiences the impulse force at B and moves to C. The green triangle represents the area now swept out by the line from the Sun to the planet. But as shown in the figure, its altitude and base are the same as that of the previous green triangle (now yellow), and therefore so is its area. We continue in this manner throughout the rest of the path of the planet as shown in Figure 8. And so we see that the triangles swept out by the line from the Sun to the planet are all equal and for equal intervals of time. Again as the time intervals are taken to be smaller and smaller, the areas of the triangles approximate closer and closer the area continuously swept out by the line from the Sun to the planet moving along is smooth orbit, thus preserving the sweeping out of equal areas in equal time intervals.

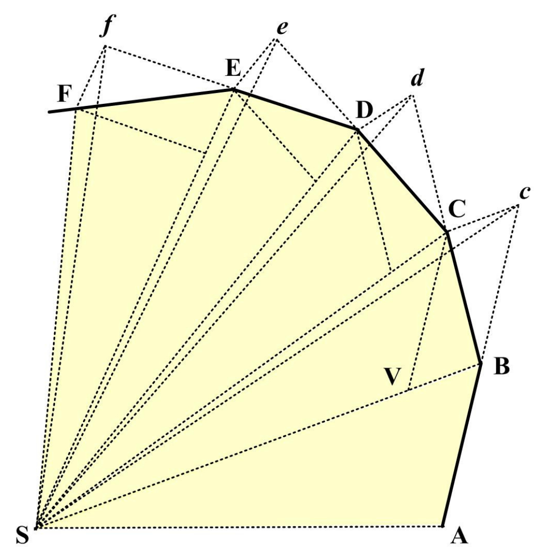


Figure 8

Implications

I hope the plethora of figures does not obscure the simplicity of Newton's proof of Kepler's equal areas law. The key is the use of the properties of triangles from the geometry of Euclid, namely that all triangles that have a base of the same length lying on the same line and with vertex opposite the base also lying on the same line parallel to the base have the same area. It is true that this geometric property is wedded to the parallelogram law deriving from physical observation, which combines the intent of the planet to move in a straight line according to the Law of Inertia with the constant pull of the planet toward the Sun. The resultant path of motion becomes the orbit of the planet.

The point I want to make here is that it looks like physical phenomena are behaving the way they do *because* of some simple mathematical relations (Euclid's geometry) that were developed independently of physical considerations. This is most mysterious. There are those who insist that all of mathematics is so dependent on physical reality that these connections should not be surprising. But I do not share this philosophy. Nor do I believe at the other extreme as some do, such as Max Tegmark, that what we interpret as physical reality is in fact just mathematics and mathematical objects. Of course I believe mathematics originated in our perception of physical reality and that it periodically gets stimulus from the conundrums of physical reality, but its modus operandi is totally different from the experimental approach of the physical sciences. The results of mathematics stem from the *rules of logical arguments* and not *physical causality*, which is our customary explanation given for physical behavior. So I find these seemingly independent connections surprising and wonderful. A classic essay on this subject is "The Unreasonable Effectiveness of Mathematics in the Natural Sciences" by Eugene Wigner ([6]).

Now it would not be quite fair to leave the subject with the idea that there is no physical explanation for the equal areas law, though "explanation" may depend on one's familiarity with concepts in physics. In fact, modern presentations of Kepler's equal areas law cite the concept of the conservation of angular momentum. This raises two points: why are there conservation laws, and what does angular momentum mean? The conservation laws are wonderful things in themselves and become quite mysterious in their seeming relationship to group theory from abstract algebra. So their physical causal purity is perhaps also somewhat tainted by mathematics. And then there is the idea of angular momentum, which I personally have always found a bit hard to fathom. Perhaps I will attempt another essay someday trying to establish an intuitive foundation for this phenomenon and how it can "explain" the equal area law. As of now, Newton's geometric proof is still my favorite.

References

1. Barker, Peter, and Bernard R. Goldstein, "Distance and Velocity in Kepler's Astronomy," *Annals of Science*, 51, 1994, pp.59-73 (http://dx.doi.org/10.1080/00033799400200131)

2. Feynman, Richard, *The Character of Physical Law,* MIT Press, 1965. First MIT Press Paperback Edition, March 1967, 12th printing, 1985 (people.virginia.edu/~ecd3m/1110/Fall2014/The_Character_of_Physical_Law.pdf)

   This is the book based on the wonderful Messenger Lectures Feynman gave in November 1964 at Cornell University, which I was blessed to attend when I was a graduate student there in mathematics. Feynman gives Newton's proof of the equal area law on pages 40-43 of the book. The video of the lectures is on YouTube. The part where Feynman demonstrates Newton's proof is in the second lecture beginning about 15 minutes into the talk (this can be seen on YouTube: https://www.youtube.com/watch?v=M9ZYEb0Vf8U)

3. Larouche Youth Movement (LYM): Johannes Kepler's *New Astronomy*, website with Flash animations, last updated 8 March 2008 (http://science.larouchepac.com/kepler/newastronomy/)

   This is a rather extraordinary website. Initially the Larouche economic theory intruded in the opening material, but then the actual discussion and explanation of Kepler's *Astronomia Nova* was pure science — and very elucidating. It was a bit reminiscent of some Russian popular expositions of mathematics during the Soviet period when the author had to reference the communist idea of dialectic materialism. It was easily isolated and did not intrude in the mathematical discussion.

4. Newton, Isaac, *Principia: The Mathematical Principles of Natural Philosophy*, Andrew Motte (d. 1730), trans., First American Edition, Revised And Corrected, With A Life Of The Author, By N. W. Chittenden, Published By Daniel Adee, 45 Liberty Street, New-York, 1846 (https://archive.org/details/newtonspmathema00newtrich)

5. Stevenson, James, "Kepler's Ellipse," in *Meditations on Mathematics*, 27 December 2016

6. Wigner, Eugene, "The Unreasonable Effectiveness of Mathematics in the Natural Sciences," in *Communications in Pure and Applied Mathematics*, vol. 13, No. I (February 1960). New York: John Wiley & Sons, Inc. (http://www.dartmouth.edu/~matc/MathDrama/reading/Wigner.html)