www.iop.org/journals/physed

Teaching Kepler's laws as more than empirical statements

Ellis D Noll

Webb School of Knoxville, 9800 Webb School Drive, Knoxville, TN 37923, USA

E-mail: Ellis_Noll@webbschool.org

Abstract

At the pre-college and first-year college level of physics instruction, Kepler's laws are generally taught as empirical laws of nature. Introductory physics textbooks only derive Kepler's Second law of areas. It is possible to derive all of Kepler's laws mathematically from the conservation laws, employing only high-school algebra and geometry. Moreover, a treatment of Kepler's Third law can naturally proceed from the general elliptic orbit to the special case of a circular orbit. Consequently, a study of Kepler's Third law need not be restricted to circular orbits.

Johannes Kepler's three statements

Toward the end of the sixteenth century, the astronomer Tycho Brahe made precise observations of the motions of the planets. Johannes Kepler served as Tycho Brahe's assistant until the latter's death in 1601. On obtaining Brahe's star diaries at his death, one of Kepler's first problems was to plot the path of the Martian orbit. Copernicus had been forced to use five epicycles to account for the motion of the Red Planet.

Upon plotting the Martian orbit, Kepler abandoned the long obsession with circles as the only figures suitable to the perfection of the heavens. Kepler's thinking changed from classically circular to elliptic orbits. In 1609 his *Astronomia Nova* (*A New Astronomy*) appeared, which contained his first two laws: 'Planets move in elliptic orbits with the Sun as one of the foci, and a planet sweeps out equal areas in equal times.' The discovery of these laws would not have been possible without Brahe's precise data.

Just one year later, in 1610, Galileo's *Sidereus Nuncius* (*Starry Messenger)* appeared, in which Galileo announced that he had observed four satellites orbiting Jupiter, providing more evidence for a heliocentric system. In 1621 Kepler published *Epitome Astronomiae Copernicanae* (*Epitome of Copernican Astronomy*), which became the most influential introduction to heliocentric astronomy. All the epicycles previously required to explain planetary motion were discarded by Kepler in this most important and lasting work. In 1619 Kepler published *Harmonice Mundi* (*Harmony of the World*), in which he derived the heliocentric distances of the planets and their periods. In this work we find his third law, relating the periods of the planets to their mean orbital radii.

Without a doubt, the books of Kepler and Galileo made world-changing contributions, but the book that would change the world more than any other emerged from the press in 1687. In the *Principia* [1], Newton applied *a priori* reasoning and the power of mathematics to dynamics in three laws of motion and a universal law of gravitation. The full acceptance of the heliocentric model occurred when Newton discovered that Kepler's statements of planetary motion were a necessary consequence of these laws.

0031-9120/02/030245+06$30.00

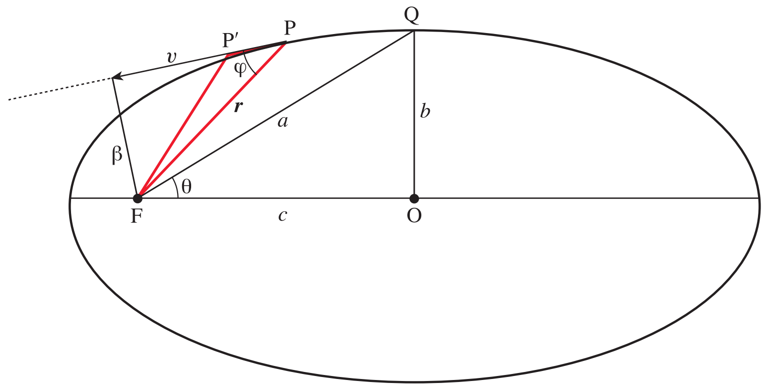

Figure 1. Elliptic orbit of a satellite around its focus F. The pedal distance, β, is the perpendicular distance from the focus to the tangent line of the velocity, v.

Rethinking the teaching of Kepler's laws

In the twenty-first century, without the benefit of original discovery, teaching elliptic orbits to physics students, especially first-year students, poses a challenge in how to proceed with a mathematical development of Kepler's laws. At this level of instruction, Kepler's laws are regarded as empirical laws of nature, and usually little attempt is made to provide a mathematical treatment. Most pre-college and first-year college physics textbooks offer little help in this matter, since their approach is also to nearly singularly emphasize the empirical nature of Kepler's laws. An occasional first-year college text [2] does provide a mathematical treatment of Kepler's Second law, usually employing calculus, but falls short in following through with the remaining laws of Kepler.

Ostensibly, the approach taken by textbook authors at this level of instruction is that students' mathematical skills are insufficient to derive mathematical expressions for elliptic orbits. This disposition is disconcerting in light of the pedagogical objective of supporting scientific law with both theory and experiment. A result of this perspective is the propensity for textbooks to state Kepler's laws, followed by a derivation of Kepler's Third law for circular orbits, completely disregarding a derivation of Kepler's Third law for elliptic orbits. Since circular orbits are a special case of elliptic orbits and very rare in our solar system, it is instructive to reverse the process and derive Kepler's Third law for elliptic orbits. Then his Third law for circular orbits can be shown to be a special case. In addition, the conservation laws can be organized into the pedal equation of an ellipse to arrive at Kepler's First law. Furthermore, it is not necessary for students to be familiar with the calculus to appreciate and understand how Kepler's laws are derived. Kepler's laws can be fully examined with high-school algebra, geometry and elementary physics.

Kepler's Second law:

A line joining any planet to the Sun sweeps out equal areas in equal times.

In figure 1 a satellite at point P is subject to the central force of gravity. Since for a central force there are no torques, the angular momentum of the satellite remains constant. As shown in figure 2, relative to a point in space, the angular momentum, $\boldsymbol{L}$, of a particle of mass, m, moving with a velocity, $\boldsymbol{v}$, at a perpendicular position, $r_\perp$, to the line of the velocity is defined as

$$L = mr_\perp v. \tag{1}$$

Thus, the angular momentum of the satellite at point P in figure 1 is

$$m\beta v = C_1 = \text{constant} \tag{2}$$

where β is the perpendicular distance from the force centre, taken as origin, to the line of the velocity.

Equation (2) establishes the second law of equal areas being swept out in equal times by the position vector. Let t_1 be the time for the satellite in figure 1 to travel from point P to point P′. Then the time rate of sweeping out of area A in triangle

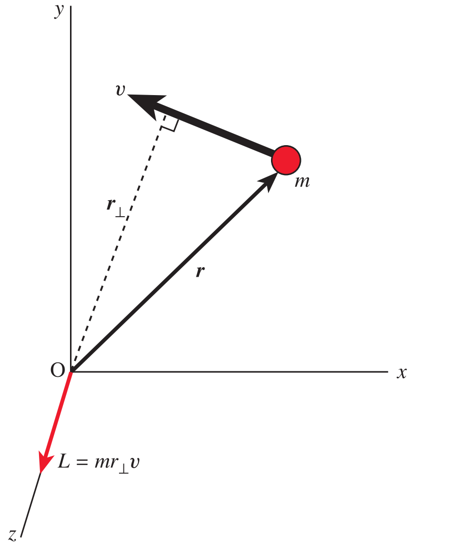

Figure 2. Angular momentum is defined relative to a point in space. In the figure, that point is the origin, O. The angular momentum vector, $\boldsymbol{L}$, is always perpendicular to the plane formed by the position vector, $\boldsymbol{r}$, and the velocity vector, $\boldsymbol{v}$. Thus, if the position vector, $\boldsymbol{r}$, drawn from the origin, and the velocity vector, $\boldsymbol{v}$, are in the x–y plane, as shown in the figure, then the angular momentum vector, $\boldsymbol{L}$, is along the z axis. The magnitude of $\boldsymbol{L}$ is given by $L = mr_\perp v$, where $r_\perp$ is the component of $\boldsymbol{r}$ perpendicular to the line of $\boldsymbol{v}$. In the absence of a net external torque, angular momentum, $\boldsymbol{L}$, is conserved.

FPP′ is constant. Thus,

$$A/t = \tfrac{1}{2}\beta \mathrm{PP}'/t_1 = \tfrac{1}{2}\beta v = C_1/2m = \text{constant} \qquad (3)$$

where β is the altitude of triangle FPP′ and path length PP′ is the triangle's base, assumed to be sufficiently small to require that in t_1 the satellite's average velocity, PP'/t_1, corresponds to its instantaneous velocity, v.

As shown in figure 3, the axes of an ellipse are the semi-major axis, a, and semi-minor axis, b. Since the area of an ellipse is πab, as shown in the Appendix, it follows from equation (3) that

$$\pi ab/T = C_1/2m \qquad (4)$$

where T is the period of the satellite.

Compare equations (3) and (4), obtaining

$$v = (2\pi ab/T)/\beta. \qquad (5)$$

Note that $2\pi ab/T$ in equation (5) is constant. When the satellite reaches its perigee or apogee, β in equation (5) corresponds to R_P at perigee and R_A at apogee. The distances R_P and R_A are directed from the focus perpendicular to the line of the velocity, as shown in figure 3. Since the perigee distance, R_P, is the satellite's distance of closest approach to the Earth and the apogee distance, R_A,

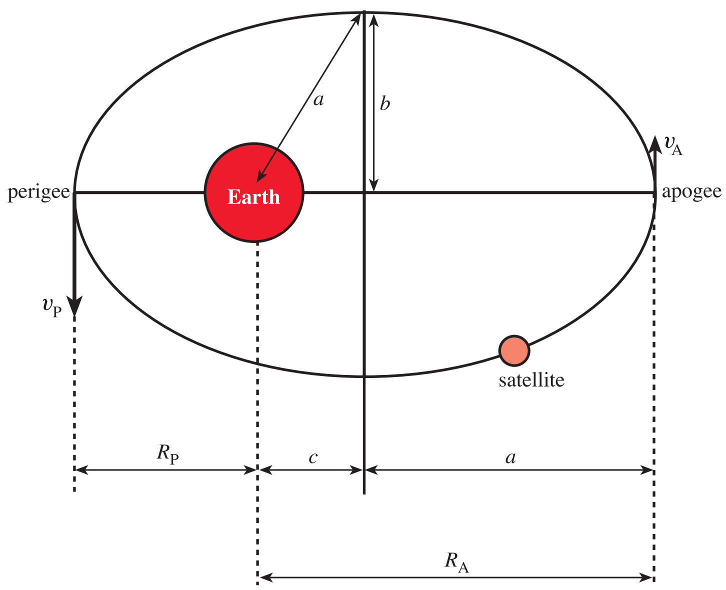

Figure 3. The Earth is located at one of the foci of an ellipse, with semi-major axis a, semi-minor axis b and eccentricity $e \equiv c/a$. The satellite's closest distance of approach to Earth (perigee) is $R_\mathrm{P} = a - c = a(1 - e)$. The satellite's farthest distance from Earth (apogee) is $R_\mathrm{A} = a + c = a(1 + e)$. The apogee to perigee distance ratio $R_\mathrm{A}/R_\mathrm{P} = v_\mathrm{P}/v_\mathrm{A}$. For $e = 0$, the orbit is a circle.

is the greatest distance between satellite and Earth, equation (5) establishes that the satellite's speed is greatest at perigee and least at apogee. It can also be ascertained from equation (5) that the perigee to apogee velocity ratio, v_P/v_A, is inversely related to the perigee to apogee distance ratio, R_P/R_A. Hence,

$$v_P/v_A = R_A/R_P. \qquad (6)$$

Kepler's Third law:

The square of the period of any planet is proportional to the cube of the semimajor axis of the ellipse.

The satellite's total energy C_2 can be determined from the conservation of energy principle. Taking the zero value of potential energy to be at infinity, then

$$\tfrac{1}{2}mv^2 - GMm/r = C_2 = \text{constant}. \qquad (7)$$

In equation (5), set β equal to R_P at perigee and R_A at apogee. Then

$$\tfrac{1}{2}m\frac{(2\pi ab/T)^2}{R_P^2} - \frac{GMm}{R_P} = \tfrac{1}{2}m\frac{(2\pi ab/T)^2}{R_A^2} - \frac{GMm}{R_A}. \qquad (8)$$

Rearrange equation (8), obtaining

$$(2\pi ab/T)^2\left(\frac{1}{R_A^2} - \frac{1}{R_P^2}\right) = 2GM\left(\frac{1}{R_A} - \frac{1}{R_P}\right). \qquad (9)$$

Simplify equation (9), obtaining

$$(2\pi ab/T)^2\frac{R_A + R_P}{R_A R_P} = 2GM. \qquad (10)$$

Observe from figure 3 that the sum of the *apogee* and *perigee* distances equals the length of the *major axis*:

$$R_A + R_P = 2a. \qquad (11)$$

From figure 3 and equation (A.4) in the Appendix, it can be seen that the product, $R_A R_P$, equals the square of the *semi-minor axis*:

$$R_A R_P = (a + c)(a - c) = b^2. \qquad (12)$$

Substitute equations (11) and (12) for the ratio in equation (10). Then

$$(4\pi^2a^2b^2/T^2)(2a/b^2) = 2GM. \qquad (13)$$

Simplify equation (13), obtaining

$$a^3/T^2 = GM/4\pi^2. \qquad (14)$$

Equation (14) is Kepler's Third law for elliptic orbits. In the special case of a circular orbit, the semi-major axis of the ellipse, a, is replaced by the circle's radius.

Let the satellite be located at point Q in figure 1. Then, $r = a$ and $\beta = b$, at point Q. Multiply the first term of equation (7) by (a/a), and compare equations (5) and (7) at point Q. Then

$$\tfrac{1}{2}ma(2\pi ab/T)^2/ab^2 - GMm/a = C_2. \qquad (15)$$

Rearrange equation (15), obtaining

$$\tfrac{1}{2}m(4\pi^2a^3/T^2)/a - GMm/a = C_2. \qquad (16)$$

Substitute GM for $4\pi^2a^3/T^2$ from equation (14) into equation (16) obtaining

$$GMm/2a - GMm/a = C_2 \qquad (17)$$

which simplifies to

$$C_2 = -GMm/2a \qquad (18)$$

where C_2 is the total constant energy of the satellite, provided the conservation of energy principle applies (i.e. no friction). Compare equations (7) and (18), obtaining

$$\tfrac{1}{2}mv^2 - GMm/r = -GMm/2a. \qquad (19)$$

Kepler's First law:

All planets move in elliptic orbits with the Sun at one focus.

Multiply equation (19) by the semi-major axis of the ellipse, a, and compare the first term of equation (19) with equation (5). Then

$$\tfrac{1}{2}m(4\pi^2a^3/T^2)b^2/\beta^2 - GMma/r = -GMm/2. \qquad (20)$$

From equation (14), $GM = 4\pi^2a^3/T^2$. Substituting this quantity into the first term of equation (20) and simplifying yields

$$\frac{b^2}{\beta^2} - \frac{2a}{r} = -1. \qquad (21)$$

Equation (21) is the pedal equation [3, 4] of an ellipse having semi-major axis a, semi-minor axis

b, with radial distance from the focus, r, and pedal (or perpendicular) distance, β.

From figure 1, φ is the angle between the line of velocity and r. Thus,

$$\beta = r\sin\varphi. \tag{22}$$

Compare equations (21) and (22), obtaining

$$\frac{b^2}{r^2\sin^2\varphi} - \frac{2a}{r} = -1. \tag{23}$$

Multiply equation (23) by r^2, generating the quadratic equation

$$r^2 - 2ar + \frac{b^2}{\sin^2\varphi} = 0. \tag{24}$$

That the pedal equation of the ellipse gives the results expected by the traditional ellipse equation can be seen by considering two special cases:

- Case I: $\varphi = 90°$. By applying the quadratic formula to equation (24) when $\varphi = 90°$, $r = a \pm c$, an expected result for an elliptical orbit at apogee (where $r = a + c$) and at perigee (where $r = a - c$).
- Case II: Consider a satellite located at point Q in figure 1, where $\varphi = \theta$ and $\sin\theta = b/r$. Substituting the latter quantity into equation (24) gives $r = a$, an expected result for point Q located on an elliptic orbit.

Compare equation (24) and equation (A.5) in the Appendix. Then,

$$r^2 - 2ar + \frac{a^2(1-e^2)}{\sin^2\varphi} = 0. \tag{25}$$

For the special case of a circular orbit, $\sin\varphi = 1$ and $a = r$. Substituting these quantities into equation (25) readily illustrates that the eccentricity, e, for a circle is zero.

Conclusion

Instead of simply stating Kepler's First law as an empirical statement, the conservation laws are employed to derive the pedal equation of an ellipse. Thus, a planetary body with negative total energy is clearly shown to orbit the force centre in an elliptic orbit. In the special case of zero eccentricity, the orbit is a circle. Kepler's Third law is derived for the general elliptic orbit instead of the customary circular orbit. This is pedagogically superior and intellectually more satisfying than deriving Kepler's Third law for a circular orbit and magically transforming the circle's orbital radius into a semi-major axis for the case of an elliptic orbit. The derivations presented illustrate all of Kepler's laws and are effective in illustrating other principles such as conservation of angular momentum and of mechanical energy. First-year physics students, unfamiliar with the calculus, can still appreciate the mathematics of elliptic orbits at a practical and realistic level. High-school algebra, geometry and elementary physics are sufficient prerequisites for a complete mathematical understanding of Kepler's laws.

Appendix. Fundamental properties of the ellipse

An ellipse is defined as a plane curve, such that the sum of the distances from any point on it to two fixed points is always the same. The two fixed points are called the *foci* of the ellipse. In figure A1, the two *foci* are the points $F_1(c, 0)$ and $F_2(-c, 0)$, and $P(x, y)$ is any point on the curve; then by definition the sum of the distances F_1P and F_2P is always the same. Let the sum be $2a$. Then

$$F_1P + F_2P = 2a. \tag{A.1}$$

The ellipse in figure A1 is symmetric with respect to each axis and to the origin, O. The line segment ST is called the *major axis* and its length is $2a$; and each half of it, OS or OT, is called a *semi-major axis*. The line segment PQ is called the *minor axis* and its length is $2b$; and each half of it, OP or OQ, is called a *semi-minor axis*.

The *eccentricity* e of an ellipse is defined as

$$e \equiv c/a \tag{A.2}$$

with c being the distance from the centre of the ellipse to its focus. Since $c < a$ for an ellipse, $e < 1$. For the ellipse shown in figure A1, the foci are located at $(\pm ae, 0)$. The standard equation of an ellipse whose *major* axis lies on the x-axis is given by

$$x^2/a^2 + y^2/b^2 = 1. \tag{A.3}$$

Apply the *Pythagorean theorem* to the right triangle F_1OP in figure A1, obtaining

$$a^2 - c^2 = b^2. \tag{A.4}$$

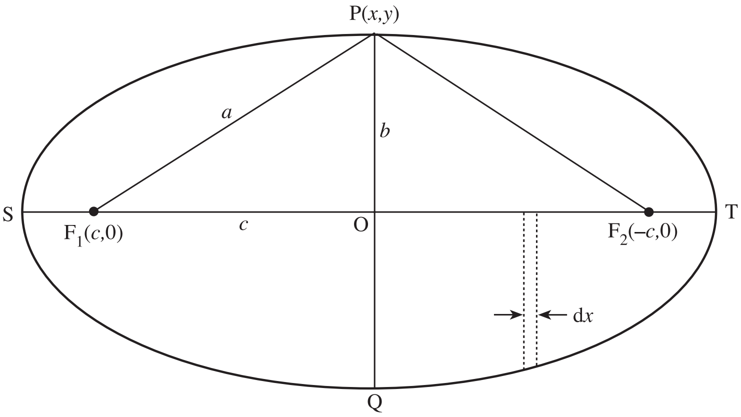

Figure A1. An ellipse is defined as a plane curve, such that the sum of the distances from any point on it to two fixed points, F_1 and F_2 in the figure, is always the same. The two fixed points are called the *foci* of the ellipse.

Compare equations (A.2) and (A.4), obtaining

$$b^2 = a^2(1 - e^2). \qquad \text{(A.5)}$$

When F_1 and F_2 are coincident in figure A1, b^2 in equation (A.5) approaches a^2 as e approaches zero.

Thus equation (A.3) has the limiting form

$$x^2 + y^2 = a^2 \qquad \text{(A.6)}$$

which is the equation of a circle with centre at the origin and radius a. *A circle is a limiting form of an ellipse whose eccentricity is zero.*

Consider the element of width dx and area $y\,\mathrm{d}x$ in the fourth quadrant of figure A1. Solve equation (A.3) for y and multiply by dx to obtain the area of the element given by

$$y\,\mathrm{d}x = \frac{b}{a}(a^2 - x^2)^{1/2}\,\mathrm{d}x. \qquad \text{(A.7)}$$

Consider one quadrant of the ellipse. The area, A_1, of one quadrant is given by

$$\begin{aligned} A_1 &= \frac{b}{a}\int_0^a (a^2 - x^2)^{1/2}\,\mathrm{d}x \\ &= \frac{b}{2a}\left[x(a^2 - x^2)^{1/2} + a^2\sin^{-1}(x/a)\right]_{x=0}^{x=a} \\ &= \pi ab/4. \end{aligned} \qquad \text{(A.8)}$$

Equation (A.8) must be multiplied by 4 to obtain the total area, A, of the ellipse. Thus,

$$A = \pi ab. \qquad \text{(A.9)}$$

Received 28 February 2001, in final form 29 January 2002
PII: S0031-9120(02)22337-0

References

[1] Newton I 1687 *Principia*. See F Cajori's revision of A Motte's translation (1729) (Norwalk, CT: Easton Press, 2000)
[2] Tipler P A 1999 *Physics for Scientists and Engineers* 4th edn, vol 1 (New York: Freeman/Worth) p 327
[3] Lockwood E H 1967 Pedal curves *A Book of Curves* (Cambridge: Cambridge University Press) ch 18, pp 152–5
[4] Yates R C 1952 Pedal curves *A Handbook on Curves and Their Properties* (Ann Arbor, MI: J W Edwards) pp 160–5

Ellis D Noll has taught college preparatory and advanced placement physics for 38 years. A Tandy Technology Scholar, Noll has been recognized by the AAPT for outstanding teaching. He is also active in science outreach programs and has authored national physics teaching and testing materials.