*CHAPTER 17 PLANETARY MOTION

Nature and Nature's Laws lay hid in night
God said "Let Newton be," and all was light.

Alexander Pope

Unlike Chapter 16, a short chapter diverging from the main stream of the book, this long chapter diverges from the main stream of the book to demonstrate that we are already in a position to do some real physics.

In 1609 Kepler published his first two laws of planetary motion. The first law describes the shape of planetary orbits:

The planets move in ellipses, with the sun at one focus.

The second law involves the area swept out by the segment from the sun to the planet (the 'radius vector from the sun to the planet') in various time intervals (Figure 1):

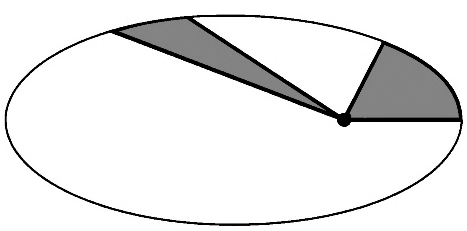

FIGURE 1

Equal areas are swept out by the radius vector in equal times. (Equivalently, the area swept out in time t is proportional to t.)

Kepler's third law, published in 1619, relates the motions of different planets. If a is the major axis of a planet's elliptical orbit and T is its period, the time it takes the planet to return to a given position, then:

The ratio a^3/T^2 is the same for all planets.

Newton's great accomplishment was to show (using his general law that the force on a body is its mass times its acceleration) that Kepler's laws follow from the assumption that the planets are attracted to the sun by a force (the gravitational force of the sun) always directed toward the sun, proportional to the mass of the planet, and satisfying an inverse square law; that is, by a force directed toward the sun whose magnitude varies inversely with the square of the distance from the sun to the planet and directly with the mass of the planet. Since force is mass times acceleration, this is equivalent simply to saying that the magnitude of the acceleration is a constant divided by the square of the distance from the sun.

Newton's analysis actually established three results that correlate with Kepler's individual laws. The first of Newton's results concerns Kepler's second law (which was actually discovered first, nicely preserving the symmetry of the situation):

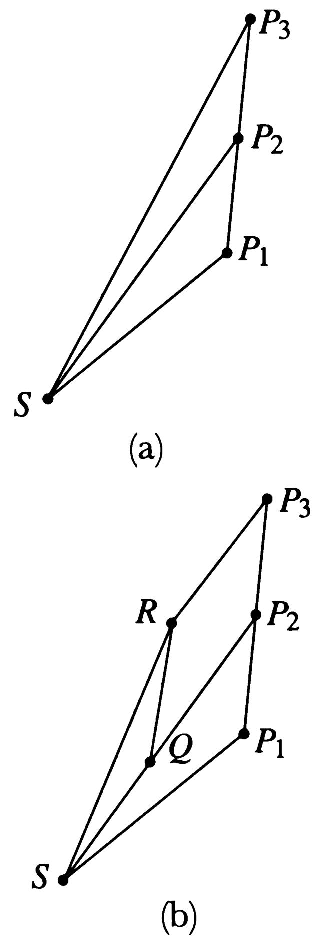

FIGURE 2

Kepler's second law is true precisely for 'central forces', i.e., if and only the force between the sun and the planet always lies along the line between the sun and the planet.

Although Newton is revered as the discoverer of calculus, and indeed invented calculus precisely in order to treat such problems, his derivation hardly seems to use calculus at all. Instead of considering a force that varies continuously as the planet moves, Newton first considers short equal time intervals and assumes that a momentary force is exerted at the ends of each of these intervals.

To be specific, let us imagine that during the first time interval the planet moves along the line P_1P_2, with uniform velocity (Figure 2a). If, during the next equal time interval, the planet continued to move along this line, it would end up at P_3, where the length of P_1P_2 equals the length of P_2P_3. This would imply that the triangle SP_1P_2 has the same area as the triangle SP_2P_3 (since they have equal bases, and the same height)—this just says that Kepler's law holds in the special case where the force is 0.

Now suppose (Figure 2b) that at the moment the planet arrives at P_2 it experiences a force exerted *along the line from S to P_2*, which by itself would cause the planet to move to the point Q. Combined with the motion that the planet already has, this causes the planet to move to R, the vertex opposite P_2 in the parallelogram whose sides are P_2P_3 and P_2Q.

Thus, the area swept out in the second time interval is actually the triangle SP_2R. But the area of triangle SP_2R is equal to the area of triangle SP_3P_2, since they have the same base SP_2, and the same heights (since RP_3 is parallel to SP_2). Hence, finally, the area of triangle SP_2R is the same as the area of the original triangle SP_1P_2! Conversely, if the triangle SRP_2 has the same area as SP_1P_2, and hence the same area as SP_3P_2, then RP_3 must be parallel to SP_2, and this implies that Q must lie along SP_2.

Of course, this isn't quite the sort of argument one would expect to find in a modern book, but in its own charming way it shows physically just *why* the result should be true.

To analyze planetary motion we will be using the material in the Appendix to Chapter 12, and the "determinant" det defined in Problem 4 of Appendix 1 to Chapter 4. We describe the motion of the planet by the parameterized curve

$$c(t) = r(t)(\cos\theta(t), \sin\theta(t)),$$

so that r always gives the length of the line from the sun to the planet, while θ gives the angle, which we will assume is increasing (the case where θ is decreasing then follows easily). It will be convenient to write this also as

$$(1) \qquad c(t) = r(t)\cdot\mathbf{e}(\theta(t)),$$

where

$$\mathbf{e}(t) = (\cos t, \sin t)$$

is just the parameterized curve that runs along the unit circle. Note that

$$\mathbf{e}'(t) = (-\sin t, \cos t)$$

is also a vector of unit length, but perpendicular to $\mathbf{e}(t)$, and that we also have

$$(2) \qquad \det(\mathbf{e}(t), \mathbf{e}'(t)) = 1.$$

Differentiating (1), using the formulas on page 247, we obtain

$$(3) \qquad c'(t) = r'(t) \cdot \mathbf{e}(\theta(t)) + r(t)\theta'(t) \cdot \mathbf{e}'(\theta(t)),$$

and combining with (1), together with the formulas in Problem 6 of Appendix 1 to Chapter 4, we get

$$\begin{aligned}\det(c(t), c'(t)) &= r(t)r'(t)\det(\mathbf{e}(\theta(t)), \mathbf{e}(\theta(t))) + r(t)^2\theta'(t)\det(\mathbf{e}(\theta(t)), \mathbf{e}'(\theta(t))) \\ &= r(t)^2\theta'(t)\det(\mathbf{e}(\theta(t)), \mathbf{e}'(\theta(t))),\end{aligned}$$

since $\det(v, v)$ is always 0. Using (2) we then get

$$(4) \qquad \det(c, c') = r^2\theta'.$$

As we will see, $r^2\theta'$ turns out to have another important interpretation.

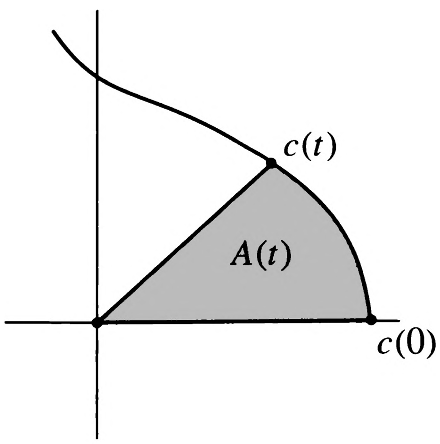

FIGURE 3

Suppose that $A(t)$ is the area swept out from time 0 to t (Figure 3). We want to get a formula for $A'(t)$, and, in the spirit of Newton, we'll begin by making an educated guess. Figure 4 shows $A(t+h) - A(t)$, together with a straight line segment between $c(t)$ and $c(t+h)$. It is easy to write down a formula for the area of the triangle $\Delta(h)$ with vertices O, $c(t)$, and $c(t+h)$: according to Problems 4 and 5 of Appendix 1 to Chapter 4, the area is

$$\text{area}(\Delta(h)) = \tfrac{1}{2}\det(c(t), c(t+h) - c(t)).$$

Since the triangle $\Delta(h)$ has practically the same area as the region $A(t+h) - A(t)$, this shows (or practically shows) that

$$\begin{aligned}A'(t) &= \lim_{h\to 0} \frac{A(t+h) - A(t)}{h} \\ &= \lim_{h\to 0} \frac{\text{area}\,\Delta(h)}{h} \\ &= \tfrac{1}{2}\det\left(c(t), \lim_{h\to 0} \frac{c(t+h) - c(t)}{h}\right) \\ &= \tfrac{1}{2}\det(c(t), c'(t)).\end{aligned}$$

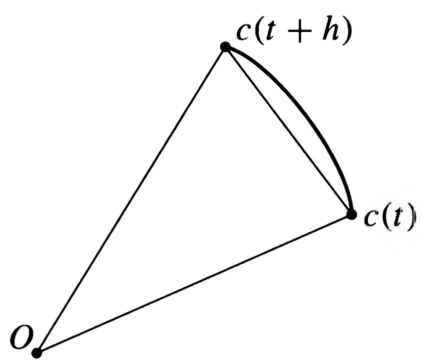

FIGURE 4

A rigorous derivation, establishing more in the process, can be made using Problem 13-24, which gives a formula for the area of a region determined by the graph of a function in polar coordinates. According to this Problem, we can write

$$(*) \qquad A(t) = \frac{1}{2}\int_{\theta(0)}^{\theta(t)} \rho(\phi)^2\, d\phi$$

if our parameterized curve $c(t) = r(t) \cdot \mathbf{e}(\theta(t))$ is the graph of the function ρ in polar coordinates (here we've used ϕ for the angular polar coordinate, to avoid confusion with the function θ used to describe the curve c).

Now the function ρ is just

$$\rho = r \circ \theta^{-1}$$

[for any particular angle ϕ, $\theta^{-1}(\phi)$ is the time at which the curve c has angular polar coordinate ϕ, so $r(\theta^{-1}(t))$ is the radius coordinate corresponding to ϕ]. Although the presence of the inverse function might look a bit forbidding, it's actually quite innocent: Applying the First Fundamental Theorem of Calculus and the Chain Rule to $(*)$ we immediately get

$$\begin{aligned} A'(t) &= \tfrac{1}{2}\rho(\theta(t))^2 \cdot \theta'(t) \\ &= \tfrac{1}{2}r(t)^2\theta'(t), \qquad \text{since } \rho = r \circ \theta^{-1}. \end{aligned}$$

Briefly,

$$A' = \tfrac{1}{2}r^2\theta'.$$

Combining with (4), we thus have

$$\boxed{A' = \tfrac{1}{2}\det(c, c') = \tfrac{1}{2}r^2\theta'.} \tag{5}$$

Now we're ready to consider Kepler's second law. Notice that *Kepler's second law is equivalent to saying that A' is constant*, and thus it is equivalent to $A'' = 0$. But

$$\begin{aligned} A'' = \tfrac{1}{2}\big[\det(c, c')\big]' &= \tfrac{1}{2}\det(c', c') + \tfrac{1}{2}\det(c, c'') \qquad \text{(see page 248)} \\ &= \tfrac{1}{2}\det(c, c''). \end{aligned}$$

So

$$\boxed{\text{Kepler's second law is equivalent to } \det(c, c'') = 0.}$$

Putting this all together we have:

THEOREM 1 Kepler's second law is true if and only if the force is central, and in this case each planetary path $c(t) = r(t) \cdot \mathbf{e}(\theta(t))$ satisfies the equation

$$r^2\theta' = \det(c, c') = \text{constant}. \tag{K_2}$$

PROOF Saying that the force is central just means that it always points along $c(t)$. Since $c''(t)$ is in the direction of the force, that is equivalent to saying that $c''(t)$ always points along $c(t)$. And this is equivalent to saying that we always have

$$\det(c, c'') = 0.$$

We've just seen that this is equivalent to Kepler's second law.

Moreover, this equation implies that $\big[\det(c, c')\big]' = 0$, which by (5) gives (K_2). ▮

Newton next showed that if the gravitational force of the sun is a central force and also satisfies an inverse square law, then the path of any object in it will be a conic section having the sun at one focus. Planets, of course, correspond to the case where the conic section is an ellipse, and this is also true for comets that visit the sun periodically; parabolas and hyperbolas represent objects that come from outside the solar system, and eventually continue on their merry way back outside the system.

THEOREM 2 If the gravitational force of the sun is a central force that satisfies an inverse square law, then the path of any body in it will be a conic section having the sun at one focus (more precisely, either an ellipse, parabola, or one branch of an hyperbola).

PROOF Notice that our conclusion specifies the shape of the path, not a particular parameterization. But this parameterization is essentially determined by Theorem 1: the hypothesis of a central force implies that the area $A(t)$ (Figure 5) is proportional to t, so determining $c(t)$ is essentially equivalent to determining A for arbitrary points on the ellipse. Unfortunately, the areas of such segments cannot be determined explicitly.* This means that we have to determine the *shape* of the path $c(t) = r(t) \cdot \mathbf{e}(\theta(t))$ without finding its parameterization! Since it is the function $r \circ \theta^{-1}$ which actually describes the shape of the path in polar coordinates, we shouldn't be surprised to find θ^{-1} entering into the proof.

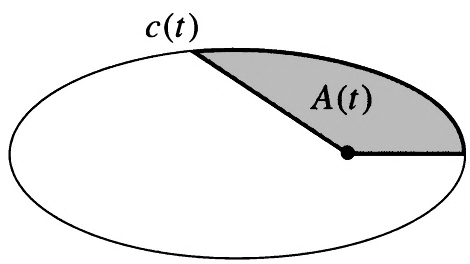

FIGURE 5

By Theorem 1, the hypothesis of a central force implies that

$$(K_2) \qquad r^2\theta' = \det(c, c') = M$$

for some constant M. The hypothesis of an inverse square law can be written

$$(*) \qquad c''(t) = -\frac{H}{r(t)^2}\,\mathbf{e}(\theta(t))$$

for some constant H. Using (K_2), this can be written

$$\frac{c''(t)}{\theta'(t)} = -\frac{H}{M}\,\mathbf{e}(\theta(t)).$$

Notice that the left-hand side of this equation is

$$[c' \circ \theta^{-1}]'(\theta(t)).$$

So if we let

$$D = c' \circ \theta^{-1}$$

(this is the main trick—"we consider c' as a function of θ"), then the equation can be written as

$$D'(\theta(t)) = -\frac{H}{M}\,\mathbf{e}(\theta(t)) = -\frac{H}{M}\big(\cos\theta(t), \sin\theta(t)\big),$$

*More precisely, we can't write down a solution in terms of familiar "standard functions," like sin, arcsin, etc.

and we can write this simply as

$$D'(u) = -\frac{H}{M}(\cos u, \sin u) = \left(-\frac{H}{M}\cos u, -\frac{H}{M}\sin u\right)$$

[for all u of the form $\theta(t)$ for some t], completely eliminating θ.

The equation that we have just obtained is simply a pair of equations, for the components of D, each of which we can easily solve individually; we thus find that

$$D(u) = \left(\frac{H \cdot \sin u}{-M} + A, \frac{H \cdot \cos u}{M} + B\right)$$

for two constants A and B. Letting $u = \theta(t)$ again we thus have an explicit formula for c':

$$c' = \left(\frac{H \cdot \sin\theta}{-M} + A, \frac{H \cdot \cos\theta}{M} + B\right).$$

[Here $\sin\theta$ really stands for $\sin\circ\theta$, etc., abbreviations that we will use throughout.]

Although we can't get an explicit formula for c itself, if we substitute this equation, together with $c = r(\cos\theta, \sin\theta)$, into the equation

$$\det(c, c') = M \qquad \text{(equation } (K_2)),$$

we get

$$r\left[\frac{H}{M}\cos^2\theta + B\cos\theta + \frac{H}{M}\sin^2\theta - A\sin\theta\right] = M,$$

which simplifies to

$$r\left[\frac{H}{M^2} + \frac{B}{M}\cos\theta - \frac{A}{M}\sin\theta\right] = 1.$$

Problem 15-8 shows that this can be written in the form

$$r(t)\left[\frac{H}{M^2} + C\cos(\theta(t) + D)\right] = 1,$$

for some constants C and D. We can let $D = 0$, since this simply amounts to rotating our polar coordinate system (choosing which ray corresponds to $\theta = 0$), so we can write, finally,

$$r[1 + \varepsilon\cos\theta] = \frac{M^2}{H} = \Lambda.$$

But this is the formula for a conic section derived in Appendix 3 of Chapter 4 (together with Problems 5, 6, and 7 of that Appendix). ▮

In terms of the constant M in the equation

$$r^2\theta' = M$$

and the constant Λ in the equation of the orbit

$$r[1 + \varepsilon\cos\theta] = \Lambda$$

the last equation in our proof shows that we can rewrite $(*)$ as

$$(**) \qquad c''(t) = -\frac{M^2}{\Lambda} \cdot \frac{1}{r(t)^2}\,\mathbf{e}(\theta(t)).$$

Recall (page 87) that the major axis a of the ellipse is given by

$$\text{(a)} \qquad a = \frac{\Lambda}{1-\varepsilon^2},$$

while the minor axis b is given by

$$\text{(b)} \qquad b = \frac{\Lambda}{\sqrt{1-\varepsilon^2}}.$$

Consequently,

$$\text{(c)} \qquad \frac{b^2}{\Lambda} = a.$$

Remember that equation (5) gives

$$A'(t) = \tfrac{1}{2}r^2\theta' = \tfrac{1}{2}M,$$

and thus

$$A(t) = \tfrac{1}{2}Mt.$$

We can therefore interpret M in terms of the period T of the orbit. This period T is, by definition, the value of t for which we have $\theta(t) = 2\pi$, so that we obtain the complete ellipse. Hence

$$\text{area of the ellipse} = A(T) = \tfrac{1}{2}MT,$$

or

$$M = \frac{2(\text{area of the ellipse})}{T} = \frac{2\pi ab}{T} \qquad \text{by Problem 13-17.}$$

Hence the constant M^2/Λ in $(**)$ is

$$\begin{aligned} \frac{M^2}{\Lambda} &= \frac{4\pi^2a^2b^2}{T^2\Lambda} \\ &= \frac{4\pi^2a^3}{T^2}, \qquad \text{using (c).} \end{aligned}$$

This completes the final step of Newton's analysis:

THEOREM 3 Kepler's third law is true if and only if the accelerations $c''(t)$ of the various planets, moving on ellipses, satisfy

$$c''(t) = -G \cdot \frac{1}{r^2}\,\mathbf{e}(\theta(t))$$

for a constant G that does not depend on the planet.

It should be mentioned that the converse of Theorem 2 is also true. To prove this, we first want to establish one further consequence of Kepler's second law. Recall that for

$$\mathbf{e}(t) = (\cos t, \sin t)$$

we have

$$\mathbf{e}'(t) = (-\sin t, \cos t).$$

Consequently,

$$\mathbf{e}''(t) = (-\cos t, -\sin t) = -\mathbf{e}(t).$$

Now differentiating (3) gives

$$\begin{aligned} c''(t) &= r''(t)\cdot\mathbf{e}(\theta(t)) + r'(t)\theta'(t)\cdot\mathbf{e}'(\theta(t)) \\ &\quad + r'(t)\theta'(t)\cdot\mathbf{e}'(\theta(t)) + r(t)\theta''(t)\cdot\mathbf{e}'(\theta(t)) + r(t)\theta'(t)\theta'(t)\cdot\mathbf{e}''(\theta(t)). \end{aligned}$$

Using $\mathbf{e}''(t) = -\mathbf{e}(t)$ we get

$$c''(t) = \left[r''(t) - r(t)\theta'(t)^2\right]\cdot\mathbf{e}(\theta(t)) + \left[2r'(t)\theta'(t) + r(t)\theta''(t)\right]\cdot\mathbf{e}'(\theta(t)).$$

Since Kepler's second law implies central forces, hence that $c''(t)$ is always a multiple of $c(t)$, and thus always a multiple of $\mathbf{e}(\theta(t))$, the coefficient of $\mathbf{e}'(\theta(t))$ must be 0 [as a matter of fact, we can see this directly by taking the derivative of formula (K_2)]. Thus Kepler's second law implies that

$$c''(t) = \left[r''(t) - r(t)\theta'(t)^2\right]\cdot\mathbf{e}(\theta(t)). \tag{6}$$

THEOREM 4 If the path of a planet moving under a central gravitational force lies on a conic section with the sun as focus, then the force must satisfy an inverse square law.

PROOF As in Theorem 2, notice that the hypothesis on the shape of the path, together with the hypothesis of a central force, which is equivalent to Kepler's second law, essentially determines the parameterization. But we can't write down an explicit solution, so we have to obtain information about the acceleration without actually knowing what it is.

Once again, the hypothesis of a central force implies that

$$r^2\theta' = M, \tag{K_2}$$

for some constant M, and the hypothesis that the path lies on a conic section with the sun as focus implies that it satisfies the equation

$$r[1 + \varepsilon\cos\theta] = \Lambda, \tag{A}$$

for some ε and Λ. For our (not especially illuminating) proof, we will keep differentiating and substituting from these two equations.

First we differentiate (A) to obtain

$$r'[1 + \varepsilon\cos\theta] - \varepsilon r\theta'\sin\theta = 0.$$

Multiplying by r this becomes

$$rr'[1 + \varepsilon\cos\theta] - \varepsilon r^2\theta'\sin\theta = 0.$$

Using both (A) and (K_2), this becomes

$$\Lambda r' - \varepsilon M \sin\theta = 0.$$

Differentiating again, we get

$$\Lambda r'' - \varepsilon M \theta' \cos\theta = 0.$$

Using (K_2) we get

$$\Lambda r'' - \frac{\varepsilon M^2}{r^2} \cos\theta = 0,$$

and then using (A) we get

$$\Lambda r'' - \frac{M^2}{r^2}\left[\frac{\Lambda}{r} - 1\right] = 0.$$

Substituting from (K_2) yet again, we get

$$\Lambda[r'' - r(\theta')^2] + \frac{M^2}{r^2} = 0,$$

or

$$r'' - r(\theta')^2 = -\frac{M^2}{\Lambda r^2}.$$

Comparing with (6), we obtain

$$c''(t) = -\frac{M^2}{\Lambda r^2}\,\mathbf{e}(\theta(t)),$$

which is precisely what we wanted to show: the force is inversely proportional to the square of the distance from the sun to the planet. ▮