2 Emergence of Modern Astronomy

Modern astronomy has deep historical roots. The main path of development for astronomy begins with the ancient Babylonians. Greek astronomers built on the observations of the Babylonians, creating a science of astronomy that was mathematical and deductive in nature. Ancient knowledge about the heavens was preserved and expanded during medieval times by Arabic scientists. During the Renaissance, the heliocentric theory of Copernicus led to additional advances by scientists such as Galileo and Kepler. This lineage, Babylonians to Greeks to Arabs to Europeans, is a great oversimplification of the rich history of astronomy. However, in a single chapter, we have only enough space for a broad overview of how modern astronomy evolved.

2.1 ▪ EARLY GREEK ASTRONOMY

Of the nine muses of classical mythology, eight dealt with various forms of music, dance, and poetry; the ninth muse, Urania, was the Muse of Astronomy. This is indicative of the ancient Greek approach to astronomy: the motions of Sun, Moon, and planets were regarded as a type of cosmic dance, revealing an underlying rhythm and harmony. A main goal of ancient Greek astronomers was to build, using deductive reasoning and mathematical computations, a conceptual model for the universe that explained the (sometimes complicated) motions of celestial bodies. To provide a bit of clarification, when historians of science talk about "ancient Greek astronomy," they aren't talking solely about developments in the geographical region currently called Greece. Rather, they embrace the entire Greek-speaking world, which in Hellenistic times, after the conquests of Alexander the Great, embraced much of the Mediterranean basin and the Near East.

Our knowledge of Greek astronomy, particularly in the time prior to Aristotle, is sadly fragmentary, due to the incompleteness of the written record. Many early Greek astronomical works are lost and are known to have existed only because they were cited by later writers. Some general aspects of Greek astronomy are well established, however. For instance, the Greeks were the first known culture to realize that the sky is three-dimensional; that is, it has a significant depth. Earlier societies, such as the Babylonians and Egyptians, thought that the sky was a thin, solid dome, arching over a flat Earth. The most famous written description of such a domed universe is in the first book of Genesis:

"God made the firmament, and divided the waters which were under the firmament from the waters which were above the firmament; and it was so. And God called the firmament heaven."[1] Greek astronomers, however, realized that the Sun and Moon, instead of being disks stuck to a domed sky, were spherical objects, at different distances from the Earth.

The realization that space was three-dimensional led Greek astronomers to an understanding of various celestial effects. For instance, they correctly explained the causes of the **phases of the Moon**. During the course of 29.5 days, the Moon appears to change in shape against the sky (see Figure 4.10b, for instance). The Moon wanes from a full circle on the sky (the full Moon) through gibbous and crescent phases until it seems to disappear (the new Moon). It then waxes through the crescent and gibbous phases until it reaches full Moon again, 29.5 days after the previous full Moon. The ancient Greeks realized that the phases occur because the Moon is an opaque sphere illuminated by the Sun. As the Moon orbits the Earth, we see different fractions of the illuminated hemisphere of the Moon, causing the apparent change in shape.[2]

The Greeks also realized the cause of **eclipses**. During a lunar eclipse, the Moon darkens dramatically; this is because the Moon passes through the Earth's shadow, depriving it of the sunlight that usually illuminates the Moon's surface. During a solar eclipse, the Sun darkens dramatically; this is because the opaque Moon passes between the Earth and the Sun, blocking the sunlight that usually reaches the Earth's surface. Thus, Greek astronomers realized that the Sun is farther away from us than the Moon is.[3]

Aristotle (384–322 BC) was one of the great philosophers and scientists of the Greek world. In his work *On the Heavens*, written around 350 BC, he turned his attention to astronomy. In this work, Aristotle pointed out that the Earth was spherical and gave four physical reasons, based on observation, why this must be true. His first reason was based on his observations of how gravity works: since gravity draws dense materials toward the center of the Earth, the resulting compression must squeeze the Earth's substance into as compact a form as possible—which is a sphere. His second reason was based on observations of partial lunar eclipses: when the edge of the Earth's shadow falls on the Moon, it always forms an arc of a circle. The only object that *always* casts a circular shadow is a sphere; thus, the Earth must be spherical.

His third reason was based on observing that new stars appear above the horizon when you head south toward the equator: on a spherical Earth, observers at a latitude ℓ north of the equator cannot see stars with declination $\delta < -90^\circ + \ell$. To take an example known in ancient times, the star Canopus ($\delta \approx -53^\circ$) is invisible from Athens ($\ell \approx 38^\circ$ N) but is visible from Alexandria, in Egypt ($\ell \approx 31^\circ$ N). This showed that the Earth was curved in the north–south direction, as a sphere would be. His fourth reason was based on observing elephants: since elephants existed both in Morocco, the westernmost region known to Aristotle, and in India, the easternmost region known to him, the two regions

[1] The image portrayed is more graphic in the original Hebrew; the word translated as "firmament" in the King James translation is *raqia*, which means a metal sheet or bowl that has been hammered out of a solid ingot.

[2] The Moon's phases are discussed in more detail in Section 4.4.

[3] Eclipses are discussed in more detail in Section 4.6.

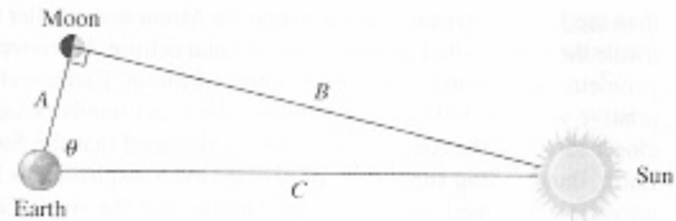


FIGURE 2.1 The geometrical method of Aristarchus for determining the relative distances to the Moon and to the Sun. (Not to scale.)

must actually be adjacent to each other on the spherical surface of the Earth. (This last, elephant-based reason sounds absurd to modern ears, but it's actually an illustration of how you can arrive at the right answer for the wrong reasons.)

The astronomer Aristarchus (ca. 310–230 BC) was notorious in his day for his unprecedented belief that the Earth orbits the Sun, rather than vice versa. The only surviving book of Aristarchus, *On the Sizes and Distances of the Sun and Moon*, doesn't explicitly mention his **heliocentric** (Sun-centered) model for the universe; instead, it puts forward geometric methods for determining the relative distances to the Sun and Moon, and their relative sizes. Aristarchus realized that when we, on the Earth, see half the Moon's disk illuminated, then the Earth–Moon–Sun angle must be exactly 90°, as seen in Figure 2.1. When the Earth–Moon–Sun angle is 90°, then the ratio of the Earth–Moon distance A to the Earth–Sun distance C is

$$\frac{A}{C} = \cos\theta, \qquad (2.1)$$

where θ is the measurable angle between the Sun and Moon as seen from the Earth. Unfortunately, the angle θ is difficult to measure with sufficient accuracy, since the difference between θ and 90° is tiny. Aristarchus thought the angle was $\theta = 87°$, which would give

$$C = A/\cos 87° = 19A. \qquad (2.2)$$

However, the actual value of the angle is $\theta = 89.853°$, much closer to a right angle, which gives

$$C = A/\cos 89.853° = 390A. \qquad (2.3)$$

Because of the difficulty of measuring θ with sufficiently high accuracy, Aristarchus underestimated the distance to the Sun, relative to that of the Moon, by a factor of 20.

Nevertheless, Aristarchus did correctly deduce that the Sun is much farther away than the Moon is. Since the Sun and the Moon are the same angular size as seen from Earth, we know from similar triangles that the ratio of their diameters is the same as the ratio of their distances from Earth. That is, Aristarchus thought that the Sun was 19 times bigger than the Moon in diameter (whereas, the Sun is actually 390 times bigger than the Moon). Aristarchus knew that the Moon was smaller than the Earth, since it fits inside the Earth's shadow during a total lunar eclipse. Moreover, he calculated, by further geometric arguments, that the diameters of Moon, Earth, and Sun had the approximate relative values 1:3:19. Again, although the exact numbers are wrong (they are actually closer to 1:4:390), Aristarchus correctly deduced that the Sun is much larger than the Earth, thus lending support to, or perhaps even inspiring, his heliocentric model for the universe. It seemed sensible to Aristarchus that the small Earth should go around the large Sun rather than the reverse.

Aristarchus deduced the relative sizes of the Moon, Earth, and Sun; absolute values for their sizes, in physical units, were provided by the work of Eratosthenes (276–195 BC), who served as the head librarian of the famous Library of Alexandria. Although the original works of Eratosthenes have been lost, a later textbook by the astronomer Cleomenes records the method by which Eratosthenes determined the circumference of the Earth.[4] Eratosthenes was aware that exactly at noon at the time of the summer solstice, the Sun was at the zenith as seen from the town of Syene (the modern city of Aswan, in upper Egypt).

On the same date, however, the Sun doesn't pass through the zenith as seen from Alexandria; instead, as shown in Figure 2.2, it is an angle α south of the zenith at noon. Eratosthenes measured the angle α and found it to be 1/50 of a full circle, or $\alpha = 7°12'$. At this point, Eratosthenes assumed that the Earth is spherical (he had read his Aristotle) and that the Sun is far enough away that the Alexandria–Sun line is effectively parallel to the Syene–Sun line. In that case, angle β in Figure 2.2 must be equal to angle α.[5] Since β, the angular distance between Alexandria and Syene, is equal to 1/50 of a full circle, the physical distance D between Alexandria and Syene must be 1/50 of the circumference of the Earth. That is,

$$C = 50D, \tag{2.4}$$

where C is the circumference of the Earth. The distance between Alexandria and Syene was known to be 5000 stades; the *stade* was a Greek unit of length, based on the length of the stadium in which foot races were held. This meant that the Earth's circumference was

$$C = 50 \times 5000 \text{ stades} = 250{,}000 \text{ stades}. \tag{2.5}$$

The length of the stade was not uniform throughout the ancient world, and historians of science have had a grand time debating the exact length of the stade used by Eratosthenes. Perhaps the most widely used stade at the time of Eratosthenes was the Attic, or Athenian, stade, equal in length to 185 meters. If we assume that Eratosthenes used the Attic stade, then the circumference that he computed was

$$C = 250{,}000 \text{ stades} \left(\frac{185 \text{ m}}{1 \text{ stade}}\right) = 4.6 \times 10^7 \text{ m} = 46{,}000 \text{ km}. \tag{2.6}$$

[4] Like most textbook writers, Cleomenes labored in humble obscurity; in fact, so obscure was Cleomenes that estimates of when he wrote his text range from 100 BC to AD 470.

[5] This equality is proved as Proposition 29 in Book I of Euclid's *Elements*, written ca. 300 BC.

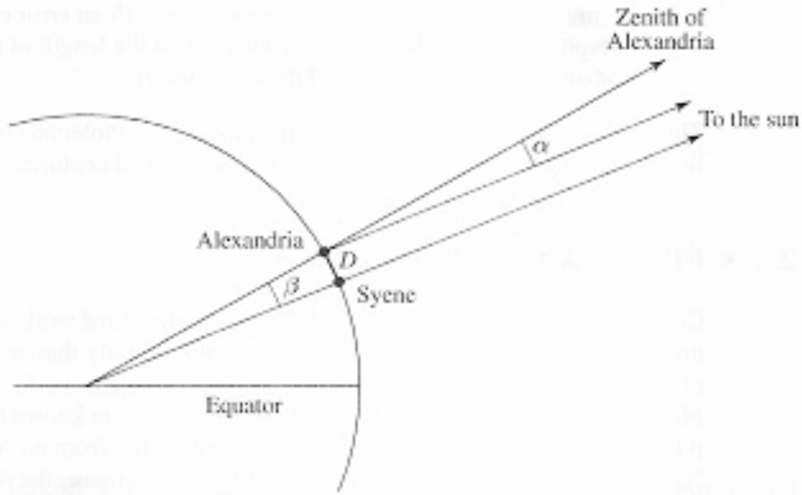


FIGURE 2.2 The geometrical method of Eratosthenes for determining the circumference of the Earth.

This is only 15% bigger than the correct value of 40,000 km. Thus, by the time of Eratosthenes, Greek astronomers not only knew the Earth is spherical but also had a reasonably correct idea of its size.

Hipparchus (ca. 190–120 BC) was perhaps the greatest astronomical observer during antiquity. Hipparchus is credited with a number of accomplishments:

- He produced an accurate catalog of hundreds of star positions. It was his careful observations that led Hipparchus to the discovery of the precession of the equinoxes, mentioned on page 14. He noted that the bright star Spica, which lies close to the ecliptic, was 6° west of the autumnal equinox. However, a star catalog made 150 years earlier had described Spica as being 8° west of the autumnal equinox. Hence, Hipparchus deduced that the equinoxes were slipping westward relative to Spica and the other stars at a rate of 2° per 150 years, equivalent to $48'' \text{ yr}^{-1}$; this is close to the accurate modern value of $50.3'' \text{ yr}^{-1}$.
- He established the **magnitude** system for describing the brightness of stars. He called the brightest stars "first magnitude," and then worked downward through second, third, fourth, and fifth magnitudes, all the way down to the faintest stars he could see, which were labeled "sixth magnitude." The more quantitative magnitude system that is used by astronomers today (described in more detail in Section 13.2) is based on that of Hipparchus.
- He computed a more accurate distance to the Moon. Although the original work of Hipparchus is lost, like so many works of Greek astronomy, a later commentator stated that Hipparchus found the average Earth–Moon distance to be roughly 70 times the Earth's radius. The actual average separation is 60.5 Earth radii.

- He measured the length of the tropical year with an error of less than 7 minutes. (Despite having such an accurate measure of the length of the year available, the Roman pontifices *still* botched their calendar!)

The observations of Hipparchus were the basis of the Ptolemaic model for the universe, which dominated Western astronomy for more than 14 centuries.

2.2 ▪ PTOLEMAIC ASTRONOMY

Claudius Ptolemaeus (called "Ptolemy" for short) lived and worked in Alexandria, Egypt, during the mid-second century AD. The scanty details that we know about his life come from his surviving astronomical books. His main work, which Ptolemy called *Mathematike Syntaxis* ("Mathematical Treatise") is better known by the name applied to it in the middle ages: the *Almagest*, a name that comes from an Arabic phrase meaning "the best." As you might guess from its flattering nickname, the *Almagest* was the most highly regarded astronomical work in the Western world from the time it was written until the sixteenth century.[6] The main portion of the *Almagest* is devoted to a geometrical model that describes the motions of the stars, Sun, Moon, and planets as seen from Earth. Before going into detail about Ptolemy's model, let's briefly review the motions of celestial bodies that he was attempting to explain.

- Stars move in diurnal circles about the celestial poles, with one complete circuit requiring one sidereal day. The stars are fixed in position relative to each other (this is only approximately true, but the relative motions of the stars are too gradual for the Greeks to have discovered).
- The Sun moves eastward relative to the stars along the ecliptic, which is tilted at 23.5° relative to the celestial equator. The average rate of motion is roughly 1° per day, but this varies over the course of a year.
- The Moon moves eastward relative to the stars along a path close to, but not identical with, the ecliptic. The average rate of motion is roughly 13° per day, but this varies over the course of a month.
- The planets Mercury, Venus, Mars, Jupiter, and Saturn usually move eastward relative to the stars, along a path close to the ecliptic; sometimes, however, they reverse course and move westward. An example of the **prograde** (eastward) and **retrograde** (westward) motion of Mars is shown in Figure 1.7.

Ptolemy's job was made unnecessarily complicated by the erroneous assumptions that he made. First, he assumed that the Earth was stationary at the center of the universe. In other words, the Ptolemaic model was **geocentric** (Earth-centered) rather than heliocentric (Sun-centered). Second, he assumed that celestial bodies moved in perfect circles at constant speed. This doctrine of **uniform circular motion** can be traced to early Greek

[6] From now on, all dates in this textbook will be AD, unless otherwise indicated.

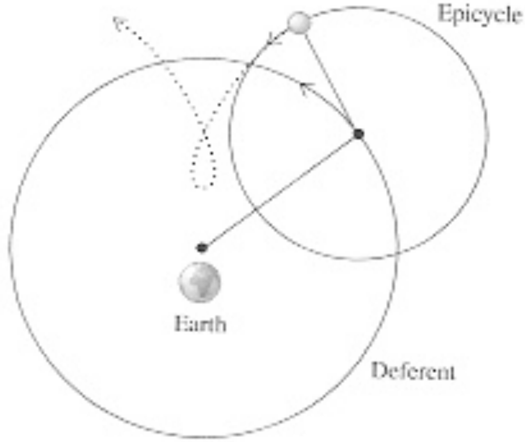


FIGURE 2.3 A planet moves at constant speed around the center of its epicycle, while the center of the epicycle moves at a constant speed along the deferent. The combination causes a model planet to move in a "loop-the-loop" motion.

philosophers such as Pythagoras and Plato. They believed that the heavens were perfect, in contrast to the obviously imperfect Earth, and that heavenly bodies must therefore move in circles (which were regarded as a perfect shape) at a perfectly constant speed.

Given these assumptions, explaining the apparent motions of the "fixed stars" was easy; Ptolemy assumed they were affixed to a rigid spherical shell, which rotated from east to west about the celestial poles, completing one rotation every sidereal day. Explaining the apparent motion of the Sun was more difficult. How could the nonuniform motion of the Sun along the ecliptic be reconciled with the dogma of uniform circular motion? Ptolemy followed the example of his predecessors by using a concept known as the *eccentric*. The Sun, Ptolemy assumed, moved along a circular orbit at a constant speed; however, the Earth was offset from the orbital center by a short distance. This small offset was referred to as the orbit's eccentric.[7] As the Sun moves along the orbit at a constant physical speed, its angular speed as seen from Earth is greatest when it's closest to the Earth, and smallest when it's farthest from the Earth. Ptolemy found that when he displaced the Earth from the orbital center by roughly 4% of the orbital radius, he could reproduce the observed motions of the Sun with fair accuracy.

Although the eccentric can describe an angular speed that varies with time, it cannot describe retrograde motion, in which the angular speed of a planet actually changes sign, rather than simply slowing down and speeding up. Ptolemy explained retrograde motion of a planet by using an **epicycle**, illustrated in Figure 2.3. In the epicyclic model, a planet travels at a constant speed around a circular path called an epicycle. At the

[7] The word "eccentric" literally means "away from the center"; thus, when you call a friend's behavior eccentric, that's another way of saying that he's a few standard deviations away from the mean.

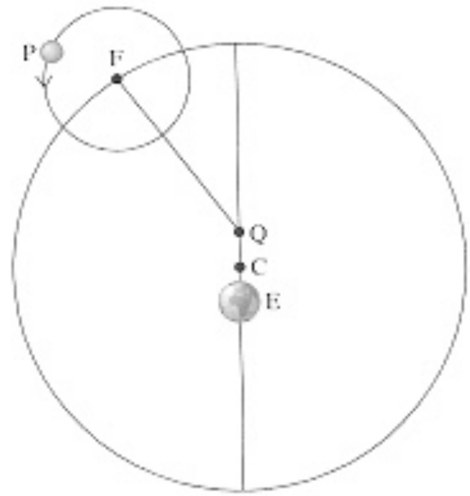


FIGURE 2.4 The complete Ptolemaic model for a planet's motion, including the equant.

same time, the center of the epicycle moves at a constant speed around the center of a larger circle called the **deferent**. The combination of an epicycle and a deferent can produce retrograde motion. Suppose the planet moves counterclockwise at a speed v around its epicycle, while the center of the epicycle moves counterclockwise at a speed w around its deferent, as shown in Figure 2.3. When the planet is at the outside of its epicycle, its speed relative to the center of the deferent is $v + w$; when it's at the *inside* of its epicycle, its speed is $v - w$. Thus, if $w > v$, the planet is actually moving backward (or in retrograde) when it is closest to the center of the deferent. A typical path traced out by a planet on an epicycle is shown in Figure 2.3. By fiddling with the sizes of eccentrics and epicycles, and by playing with the relative orbital speeds of epicycles and deferents, Ptolemy could get a fairly good fit to the observed motions of planets on the celestial sphere, but not quite good enough. His models were unable to match the observations exactly. Eccentrics, deferents, and epicycles were ideas that Ptolemy had inherited from previous Greek astronomers. However, in order to match the observations with the necessary accuracy, Ptolemy introduced a new device called the **equant**, illustrated in Figure 2.4.

In Ptolemy's new construction, the Earth (labeled E in the figure) is offset from the center of the planet's deferent (labeled C) by a small distance. Ptolemy dictated, however, that the center of the epicycle (labeled F) moved along the deferent at a changing physical speed, such that its angular speed would be constant as seen from the **equant point** (labeled Q). The equant (Q), orbital center (C), and Earth (E) lie along a straight line and are spaced so that the distance Q–C is equal to the distance C–E. The concept of the equant stretched the doctrine of uniform circular motion to the absolute limit; according to Ptolemy's critics, it stretched it beyond the limit. Many medieval astronomers were dissatisfied by the rather contrived notion of the equant.

Nevertheless, Ptolemy's complete model for a planetary orbit, including a deferent, epicycle, and equant, had enough adjustable parameters to enable Ptolemy to make quite accurate predictions of the motions of planets as seen from Earth. It is not clear that Ptolemy intended his complicated geocentric model to be an actual physical model of the cosmos. It worked adequately as a mathematical model, which accounted for its popularity during medieval times; people wanted reasonably accurate predictions of the locations of the Sun, Moon, and planets, which the Ptolemaic model provided. The fact that Ptolemy's model was geocentric also made it conceptually acceptable. There were a number of plausible arguments, during Ptolemy's time and later, why a geocentric model seemed correct:

- We cannot feel the motion of the Earth. A circumference of 250,000 stades implies a rotation speed at the Earth's equator of roughly 3 stades per second, or about 50 times the speed of the fastest sprinter. It seemed inconceivable that such a rapid speed should be imperceptible.
- The Earth's centrality and importance was somehow gratifying. (The Earth must be important; after all, *we* live on it.)
- **Stellar parallax** is not observed. This is the most serious scientific objection to a heliocentric model and deserves a fuller discussion, which is given below.

In general, the term **parallax** refers to the shift in apparent position of an object when seen from two different locations. For instance, if you hold up your thumb at arm's length and view it first through your right eye and then through your left, you will see your thumb's image jump from left to right by roughly 5° relative to objects in the background. In astronomy, the term **geocentric parallax** refers to the shift in apparent position of a relatively nearby object, such as the Moon or a planet, when seen from two different points on the Earth's surface. Geocentric parallax, illustrated in Figure 2.5a, is also referred to as **diurnal parallax**. If you want to observe geocentric parallax, you don't have to go on an expedition; during the course of 12 hours, the daily (or diurnal) rotation of the Earth will carry you through a distance $d = D \cos \ell$, where $D \approx 12{,}700$ km is the Earth's diameter and ℓ is your latitude. The closer an object is to the Earth, the larger its geocentric parallax will be. The Moon shifts in apparent position by as much as 2° when viewed from antipodal points on the Earth; however, the Sun's corresponding shift in apparent position is smaller by a factor of 390, since the Sun is 390 times farther away than the Moon is. Thus, the geocentric parallax of the Moon was easily measured by ancient astronomers (it's how Hipparchus measured the distance to the Moon, in fact), but the diurnal parallax of the Sun, and of the yet more distant stars, is too small to be measured by the naked eye.

The daily rotation of the Earth causes a change in position of an observer on the Earth; so does the annual revolution of the Earth around the Sun. **Heliocentric parallax** is the shift in apparent position of a relatively nearby star when seen from two different points on the Earth's orbit. Heliocentric parallax, illustrated in Figure 2.5b, is also referred to as **annual parallax**. If you want to observe heliocentric parallax, you don't have to launch

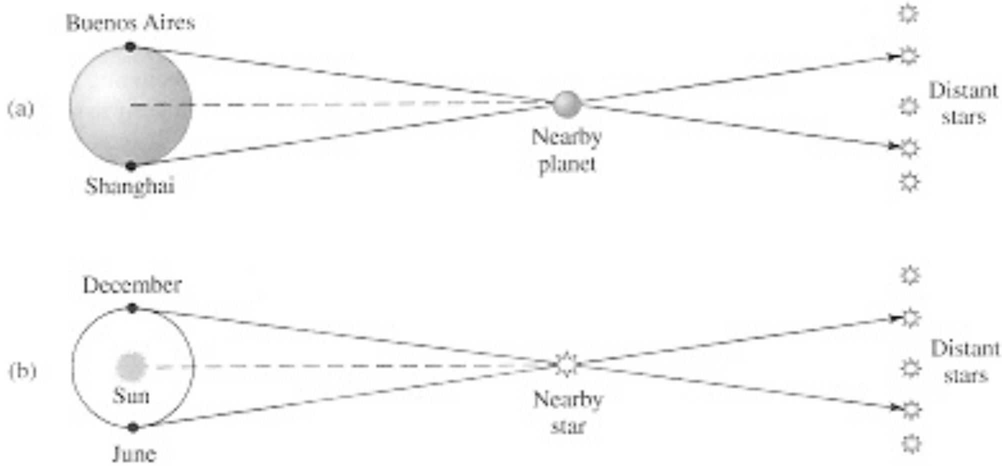


FIGURE 2.5 (a) Geocentric, or diurnal, parallax due to a change in position relative to the Earth's center. (b) Heliocentric, or annual, parallax due to a change in position relative to the Sun.

a spacecraft; during the course of half a year, the annual revolution of the Earth will carry you through a distance equal to the diameter of the Earth's orbit.

Before the invention of the telescope, astronomers attempted to measure the annual parallax of nearby stars but were unsuccessful. They recognized two possible explanations for the lack of detectable annual parallax: either the Earth was stationary or the stars were so far away that the annual parallax, like the diurnal parallax, was too small to be measured. Given the accuracy with which stellar positions could be measured in antiquity, Ptolemy and others deduced that if the solar system were heliocentric, then the nearest stars would have to be at a distance of *at least* a few thousand times the Earth–Sun distance. Such a large amount of empty space made astronomers uneasy. They preferred the more compact geocentric model. As we discuss further in Chapter 13, stellar parallax was not measured until long after the invention of the telescope. Even the Sun's nearest neighbors among the stars are at a distance of 270,000 times the Earth–Sun distance. The small, tidy Ptolemaic universe may have been psychologically comforting, but the universe is under no obligation to make us comfortable.

2.3 ▪ COPERNICAN ASTRONOMY

The Polish astronomer Nicolaus Copernicus (1473–1543) was the first scientist since antiquity to advance a heliocentric model for the universe. Copernicus was a Renaissance man metaphorically as well as chronologically; in addition to studying astronomy and mathematics, he also traveled to Italy in order to study medicine and law. After taking minor orders in the Church, he served in a variety of administrative positions. His work for the Church left Copernicus with enough time to make astronomical observations and work out his heliocentric model in detail. By the year 1514, Copernicus was

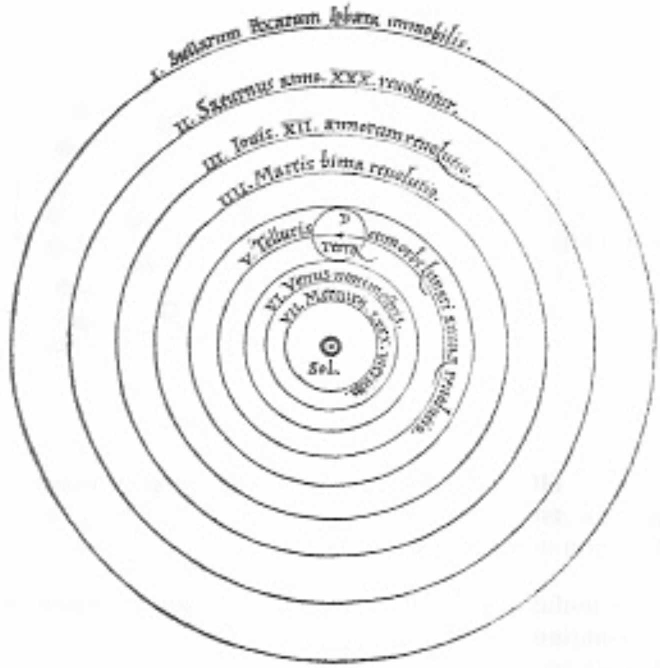


FIGURE 2.6 A schematic diagram of the heliocentric model, drawn by Copernicus (note that *Sol*, the Sun, is at center).

circulating a brief manuscript about his ideas among his friends; the grand summary of his work, the book *De Revolutionibus Orbium Coelestium* ("On the Revolutions of the Heavenly Spheres"), was not published until Copernicus was on his deathbed, in the year 1543.

The most radical aspect of the Copernican model was its insistence that the Sun, not the Earth, was at the center of the solar system (Figure 2.6), and that the Earth was both rotating about its axis and revolving about the Sun. The Copernican model, however, also had conservative aspects. For instance, Copernicus wholeheartedly embraced the dogma of uniform circular motion. One of his proudest claims for his heliocentric model was that it eliminated the need for equants (however, to match the observations, it still needed eccentrics and epicycles).

The Copernican model, although it retained eccentrics and epicycles, was conceptually simpler than the Ptolemaic model in many respects. In the Copernican model, retrograde motion of the planets is accounted for by the fact that inner planets move faster along their orbits than the outer planets do. Thus, as an inner planet, such as the Earth, overtakes an outer planet, such as Mars, the outer planet undergoes retrograde motion as seen from the inner planet. This is demonstrated graphically in Figure 2.7. In a heliocentric model, with the Earth being one of many planets orbiting the Sun, it

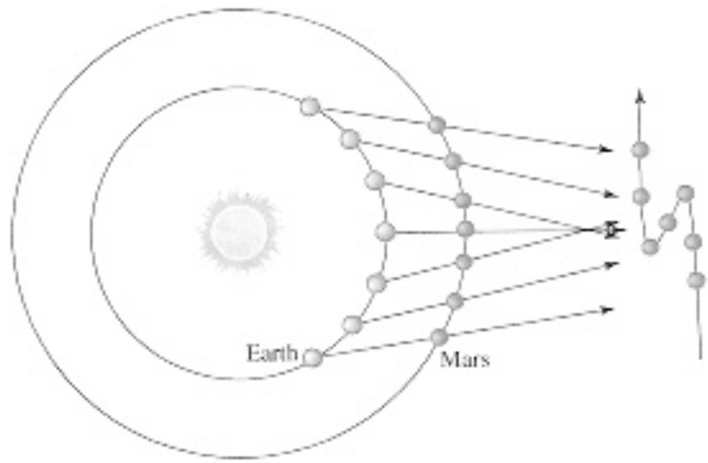


FIGURE 2.7 Explanation of the retrograde motion of Mars in a heliocentric system.

is useful to divide the planets into two groups, based on their distance from the Sun compared to that of the Earth:

- **Inferior planets** are those with orbits smaller than the Earth's orbit, that is, Mercury and Venus.
- **Superior planets** are those with orbits larger than the Earth's orbit. Mars, Jupiter, and Saturn were the superior planets known at the time of Copernicus; the planets Uranus and Neptune and the dwarf planets Ceres, Pluto, Haumea, Makemake, and Eris were not discovered until after the invention of the telescope.

In the Copernican model, the Earth is in motion around the Sun. Thus, for an Earthly observer, the positions of planets are measured from a reference frame that is co-rotating with the Earth–Sun line. It is particularly useful, as we shall see, to measure the position of planets on the celestial sphere relative to the Sun.

Some special positions of the superior planets relative to the Sun are shown in Figure 2.8. Names have been given to these special positions:

- **Opposition** occurs when the Earth lies between the Sun and the superior planet. That is, the Sun and planet are 180° apart on the celestial sphere as seen from the Earth.
- **Conjunction** occurs when the Sun lies between the Earth and the superior planet. That is, the Sun and planet are 0° apart as seen from the Earth.
- **Quadrature** occurs when the Sun and the superior planet are 90° apart as seen from the Earth. The quadrature can be either eastern, when the planet appears 90° east of the Sun on the sky, or western, when the planet appears 90° west of the Sun.

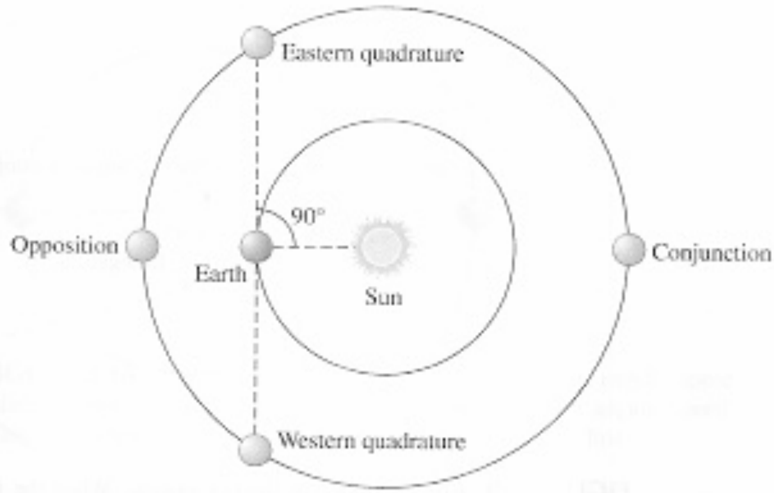


FIGURE 2.8 Configurations of superior planets. In this and following diagrams, we adopt a convention of looking down on the solar system from above the Earth's north pole.

Although inferior planets cannot be seen in opposition or in quadrature, they do have two different conjunctions, as shown in Figure 2.9:

- **Inferior conjunction** occurs when the inferior planet lies between the Earth and the Sun.
- **Superior conjunction** occurs when the Sun lies between the Earth and the inferior planet.

When a planet is not in conjunction, it is separated from the Sun on the celestial sphere by an angle θ referred to as the planet's **elongation**. Note from Figure 2.9 that an inferior planet can have the same elongation θ at two different distances from the Earth.

One of the happy results of the Copernican model is that it enabled Copernicus to compute the orbital periods of the planets, relative to the Earth's orbital period, and compute the size of planetary orbits, relative to the size of the Earth's orbit. Let's first see how Copernicus computed orbital periods, and then how he computed orbital sizes.

As seen from the Earth, planets undergo motion that can be described as periodic; that is, there is a fixed time interval between consecutive appearances of a particular planetary configuration. This time interval, known as the **synodic period** of the planet, can be found by measuring the time elapsed between successive conjunctions (for a superior planet) or the time elapsed between successive inferior conjunctions (for an inferior planet).[8] The synodic period is different from the **sidereal period** of the planet, which is the time

[8] The term "synodic" comes from the Greek word *synodos*, meaning a "coming together"—in this case, a coming together of the Sun and the planet when the planet is at conjunction.

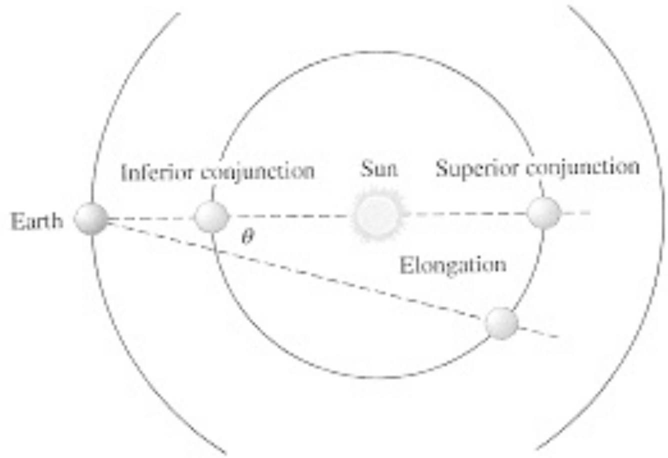


FIGURE 2.9 Configurations of inferior planets. When the planet is not in conjunction, the angle θ between the Sun and the planet, as seen from Earth, is the planet's elongation.

it takes the planet to complete one full circuit of the sky relative to the fixed stars. The synodic period of a planet is longer than its sidereal period for much the same reason that the solar day is longer than the sidereal day (as discussed in Section 1.5). As a reminder, the sidereal day is the Earth's rotation period in the nonrotating frame of reference of the distant stars (the sidereal frame); the solar day is the Earth's rotation period in a frame of reference co-rotating with the Earth–Sun line. Similarly, the sidereal period of a planet is the planet's orbital period in the nonrotating sidereal frame; the synodic period is its orbital period in a frame of reference co-rotating with the Earth–Sun line.

As in equation (1.1), let $\vec{\omega}_E$ be the angular velocity of the Earth's orbital motion in the sidereal frame; let $\vec{\omega}_P$ be the angular velocity of the planet's orbital motion in the same frame. Figure 2.10 shows the orbital motions of the Earth and an inferior planet; for an inferior planet, $\omega_P > \omega_E$. The difference between these two angular velocities is $\vec{\omega}_{syn}$, the angular velocity of the planet's orbital motion in the frame co-rotating with the Earth–Sun line. Specifically, we see that

$$\vec{\omega}_P = \vec{\omega}_E + \vec{\omega}_{syn}. \tag{2.7}$$

If $\vec{\omega}_P$ and $\vec{\omega}_E$ are parallel (that is, if the orbits of the Earth and the planet are coplanar and they orbit in the same direction about the Sun), we may write, for an inferior planet,

$$\begin{aligned}\omega_P &= \omega_E + \omega_{syn}\\ \frac{2\pi}{P_P} &= \frac{2\pi}{P_E} + \frac{2\pi}{P_{syn}}\\ \frac{1}{P_P} &= \frac{1}{P_E} + \frac{1}{P_{syn}}.\end{aligned} \tag{2.8}$$

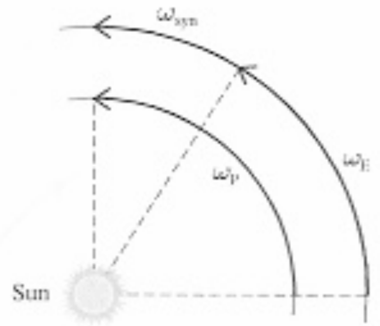


FIGURE 2.10 The angular speed of the Earth is ω_E and the angular speed of an inferior planet is ω_P. The difference between them is ω_{syn}, the angular speed of the planet in a reference frame that co-rotates with the Earth–Sun line.

In equation (2.8), P_E is the sidereal orbital period of the Earth, P_P is the sidereal orbital period of the inferior planet, and P_{syn} is the synodic orbital period of the inferior planet, as seen from Earth. As an example of an inferior planet, consider Venus. The synodic period of Venus is measured to be $P_{syn} = 583.92$ days. The Earth's sidereal orbital period is $P_E = 365.256$ days.[9] We can then compute the sidereal period of Venus:

$$P_{\rm Venus} = \left[\frac{1}{365.256\ \rm days} + \frac{1}{583.92\ \rm days}\right]^{-1} = 224.70\ \rm days. \tag{2.9}$$

In the case of a superior planet, $\omega_P < \omega_E$. If we refer to Figure 2.11, we see that $\vec{\omega}_{syn}$ is in the opposite sense to $\vec{\omega}_E$ and $\vec{\omega}_P$. Equation (2.8) then becomes, for a superior planet,

$$\omega_P = \omega_E - \omega_{syn}$$

$$\frac{1}{P_P} = \frac{1}{P_E} - \frac{1}{P_{syn}}. \tag{2.10}$$

As an example of a superior planet, consider Mars. The synodic period of Mars is measured to be $P_{syn} = 779.95$ days. Given the length of the sidereal period of Earth, $P_E = 365.256$ days, we compute the sidereal period of Mars to be

$$P_{\rm Mars} = \left[\frac{1}{365.256\ \rm days} - \frac{1}{779.95\ \rm days}\right]^{-1} = 686.98\ \rm days. \tag{2.11}$$

In addition to permitting a determination of a planet's sidereal orbital period, the Copernican model also permits us to compute the distance of each planet from the Sun. For an inferior planet, this computation is straightforward. We need only measure the

[9] Remember that due to the precession of the equinoxes, the sidereal year is slightly longer than the tropical year of $P = 365.24219$ days.

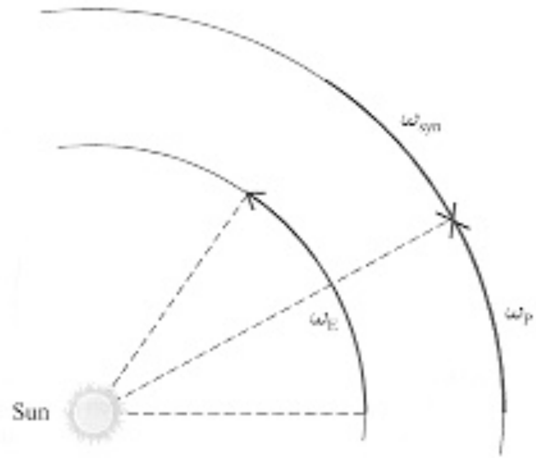


FIGURE 2.11 The angular speed of the Earth is ω_E and the angular speed of a superior planet is ω_P. The difference between them is ω_{syn}. To an observer on Earth, the angular velocity ω_{syn} of a superior planet is negative.

inferior planet's **greatest elongation**, that is, the maximum angular separation between the planet and Sun as seen from the Earth. As shown in Figure 2.12, if we approximate the orbit of the inferior planet as a perfect circle, then greatest elongation occurs when the line of sight from the Earth to the planet is exactly tangent to the planet's orbit. When that happens, the angle Earth–planet–Sun is a right angle, as the figure shows. The distance B from the planet to the Sun is then given by the relation

$$B/C = \sin\theta, \tag{2.12}$$

where θ is the angle of greatest elongation and C is the Earth–Sun distance. This method, therefore, only gives the radius of the planet's orbit in units of the Earth–Sun distance. The average distance from the Earth to the Sun is of such importance to astronomers that it is given the name **astronomical unit**, or **AU** for short. Copernicus, like the Greek astronomers before him, did not have an accurate knowledge of the absolute length of the astronomical unit.[10] However, he did know the *relative* sizes of the planets' orbits. For instance, the greatest elongation of Venus is $\theta = 46°$, so its orbital radius is

$$B = (\sin 46°)(1\ \text{AU}) = 0.72\ \text{AU}. \tag{2.13}$$

The size of the orbits of superior planets can be determined by a similar but slightly more complicated method. First, we must measure the time interval τ between opposition and eastern quadrature of the superior planet. As shown in Figure 2.13, the angle swept out by the Earth during the time interval τ is $\omega_E \tau$, where ω_E is the angular speed of the Earth's orbital motion. Over the same time interval, the superior planet (assumed to

[10] We now know that 1 AU = 149,597,870.7 km.

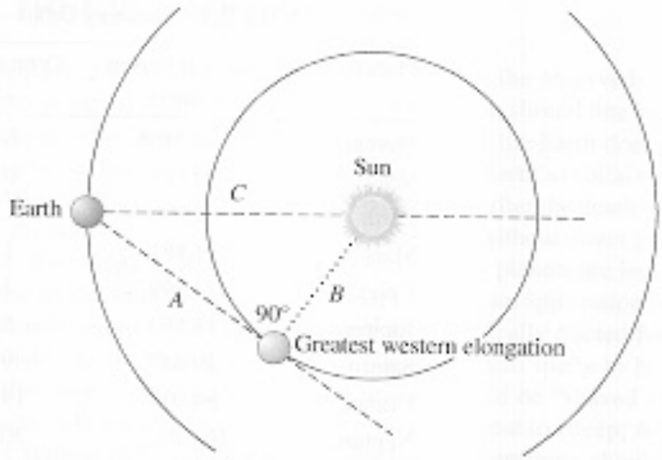


FIGURE 2.12 Measurement of the greatest elongation θ of an inferior planet allows determination of its distance B from the Sun.

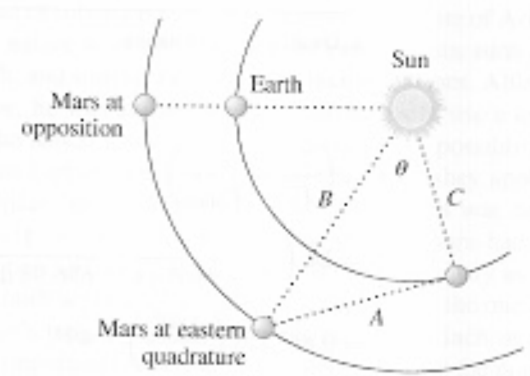


FIGURE 2.13 In the time τ between opposition and eastern quadrature, Mars sweeps out an angle $\omega_{\text{Mars}}\tau$ and the Earth sweeps out an angle $\omega_{\text{E}}\tau$. The difference between these angles is θ, with $\cos\theta = C/B$.

be Mars in the figure) sweeps out an angle $\omega_{\text{P}}\tau$, where ω_{P} is the angular speed of the planet's orbital motion. The difference between these angles is the angle $\theta = (\omega_{\text{E}} - \omega_{\text{P}})\tau$ shown in the figure. When Mars is at quadrature, the angle Mars–Earth–Sun is a right angle, so we have the relation

$$C/B = \cos\theta, \tag{2.14}$$

where C is the Earth–Sun distance and B is the Mars–Sun distance. In the case of Mars, the time from opposition to eastern quadrature is $\tau = 107$ days. Thus, the angle θ is

TABLE 2.1 Planetary Orbits

Planet[a]	Sidereal Period (years)	Orbital Radius (AU)
Mercury	0.2408	0.3871
Venus	0.6152	0.7233
Earth	1.000	1.000
Mars	1.881	1.524
Ceres	4.599	2.766
Jupiter	11.863	5.203
Saturn	29.447	9.537
Uranus	84.017	19.189
Neptune	164.79	30.070
Pluto	247.92	39.482
Haumea	283.28	43.133
Makemake	306.17	45.426
Eris	559.55	67.903

a. Dwarf planets in italics.

$$\begin{aligned}\theta &= \left(\frac{2\pi}{P_{\rm E}} - \frac{2\pi}{P_{\rm Mars}}\right)\tau \\ &= 2\pi\left(\frac{1}{365.256\ \text{days}} - \frac{1}{686.98\ \text{days}}\right)(107\ \text{days}) \\ &= 0.862\ \text{rad}\left(\frac{180^\circ}{\pi\ \text{rad}}\right) = 49^\circ. \end{aligned} \tag{2.15}$$

The distance from Mars to the Sun is then

$$B = \frac{C}{\cos\theta} = \frac{1\ \text{AU}}{\cos 49^\circ} = 1.52\ \text{AU}. \tag{2.16}$$

Table 2.1 shows the sidereal orbital period and the orbital radius for each of the planets and dwarf planets in the solar system (including those that were unknown at the time of Copernicus).[11]

[11] Truth in advertising: the simple calculations we have done in this section assume that planetary orbits are perfectly circular. Although this is a good first approximation, the orbits are actually ellipses, and what we call the "orbital radius" in Table 2.1 is actually the semimajor axis of the ellipse.

2.4 ▪ GALILEO: THE FIRST MODERN SCIENTIST

Both the Ptolemaic and Copernican models could explain the observed motions of the Sun, Moon, and planets on the celestial sphere. Why, then, should one believe that the Earth is in motion rather than the Sun? We know now that the Earth does orbit the Sun rather than vice versa, but direct experimental proof of the Earth's orbital motion was not provided until the eighteenth century, nearly two centuries after the death of Copernicus. Nevertheless, the heliocentric model came to be accepted without direct proof. This was partly because of its elegant simplicity; the motions of the planets are less complicated in a heliocentric model than in a geocentric model. This is an application of the general principle often referred to as **Occam's Razor**.[12] In its typically quoted form, Occam's Razor states that "the simplest description of Nature is most likely to be most nearly correct." In other words, unnecessary complications should be "shaved away" from a theory. Of course, when using a razor, it is important not to cut too deep; Albert Einstein is said to have rephrased Occam's Razor in the form "Everything should be made as simple as possible . . . but not simpler."

In addition to the aesthetic appeal of the heliocentric model's relative simplicity, compelling indirect evidence for heliocentrism was provided by the telescopic observations of Galileo Galilei (1564–1642). Galileo is sometimes called the first modern experimental physicist. Instead of relying purely on the pronouncements of Aristotle, Galileo tried to understand how nature works by carrying out experiments, such as swinging pendulums back and forth, and sliding weights down inclined planes. Although Galileo didn't invent the telescope, he was among the first individuals to use a telescope as a scientific instrument. The actual inventor of the telescope may possibly have been a Dutch optician called Hans Lippershey. In October 1608, Lippershey applied for a patent on what he called a *kijker*, or "looker" in English. The patent was denied by the Dutch government, however, on the grounds that "many other persons had a knowledge of the invention." Indeed, news of the telescope reached Galileo in Italy as early as May 1609; soon thereafter, he built several telescopes, each superior to the one before.

Although Galileo's telescopes had apertures of only an inch or two, they provided Galileo with many important observations. Galileo, knowing the potentially revolutionary impact of his discoveries, rushed into print in March 1610 with a pamphlet entitled *Sidereus Nuncius* ("Starry Messenger"). Many of Galileo's observations were startling to his contemporaries:

- The Moon is not smooth and perfect. Instead, as Galileo wrote, it is "uneven, rough, and crowded with depressions and bulges. And it is like the face of the Earth itself, which is marked here and there with chains of mountains and depths of valleys." In other words, there is not a vast difference between the Earth's surface and that of a celestial object, namely the Moon.

[12] Occam's Razor is named after William of Occam, a fourteenth-century friar and logician.

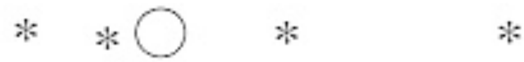


FIGURE 2.14 Galileo's illustration of the four bright satellites of Jupiter (the four asterisks), shown relative to Jupiter itself (the central disk).

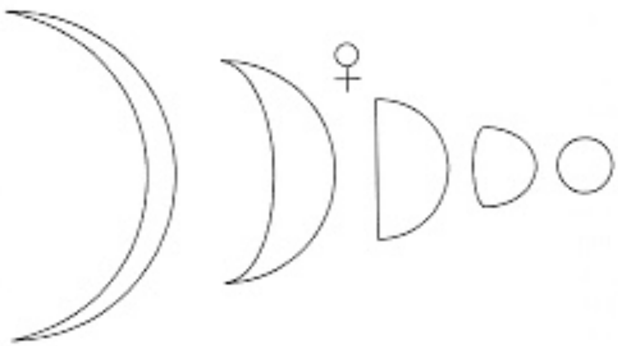
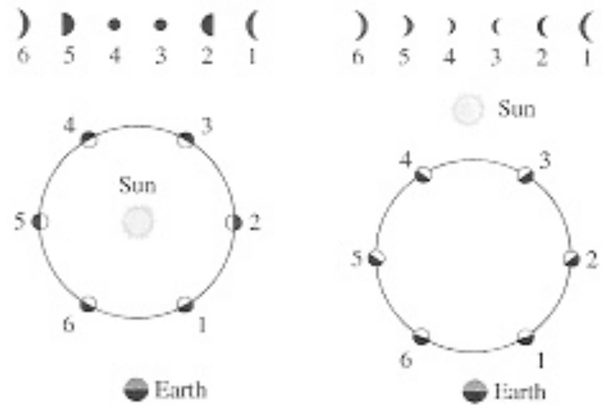


FIGURE 2.15 The phases and relative angular size of Venus, from crescent to full.

- The Milky Way, the nebulous band of light that extends around the sky, actually consists of numerous faint stars. "To whatever region of it you direct your spyglass," Galileo wrote, "an immense number of stars immediately offer themselves to view."
- Through a telescope, stars remain unresolved points, but planets show as disks. As Galileo put it, "the planets present entirely smooth and exactly circular globes that appear as little moons." (Unfortunately for astronomers, even the nearest stars are too distant to be resolved with conventional telescopes, even telescopes much larger than Galileo's.)
- The planet Jupiter has four large, bright satellites. Although Galileo called these satellites the "Medicean Stars," in honor of Cosimo de Medici, Grand Duke of Tuscany, later astronomers named them the **Galilean satellites**. The individual names of the four Galilean satellites are Io, Europa, Ganymede, and Callisto.

The Galilean satellites of Jupiter, shown in Figure 2.14 were an indirect piece of support for the Copernican system. One objection to a heliocentric model was that it required multiple centers of motion; the Earth went around the Sun while the Moon went around the Earth. This was regarded as more complex than a geocentric model in which everything goes around the Earth. However, Galileo provided clear evidence that there *had* to be multiple centers of motion; obviously, the Galilean satellites were going around Jupiter, regardless of whether Jupiter was going around the Sun or around the Earth.



(a) Copernican system (b) Ptolemaic system

FIGURE 2.16 (a) The phases of Venus in the Copernican model. (b) The phases of Venus in the Ptolemaic model.

By the end of the year 1610, Galileo made another telescopic discovery that further undermined the Ptolemaic model. He found that Venus went through all the phases that the Moon did, from full to new. Moreover, he found that the angular size of Venus was smallest when it was full and largest when it was a thin crescent. The phases of Venus, illustrated by Galileo in his later work *Il Saggiatore*, are shown in Figure 2.15. Ptolemy, in his geocentric system, had the task of explaining why Venus should always lie within 46° of the Sun if the two bodies were on independent orbits around the Earth. Ptolemy managed it by saying that the center of Venus's epicycle always lies directly between the Earth and the Sun (as shown in Figure 2.16b) and that the epicycle is big enough to subtend an angle of 92° as seen from the Earth. The geometry of this situation requires that we see primarily the nighttime side of Venus, that is, the side away from the Sun. In the Ptolemaic system, then, we would always see a new or crescent phase for Venus, as illustrated in Figure 2.16b, top.

Galileo demonstrated, however, that we see gibbous and full Venuses, as well as crescent and new Venuses. This is easily explained in the Copernican system, as shown in Figure 2.16a. In the Copernican model, the sunlit side of Venus is turned toward us when Venus is at superior conjunction; this is when Venus is at its greatest distance from Earth, and hence has its smallest angular size. Conversely, the nighttime side of Venus is turned toward us when it is at inferior conjunction, when it is closest to Earth and has its largest angular size.[13] This is in accord with the observations of Galileo.

[13]Tantalizingly, when Venus is in its crescent phase, it is just under an arcminute across and thus can *almost* be resolved by the human eye. If our eyes were a bit better, or Venus were a bit larger, the phases of Venus would have been seen before the invention of the telescope, thus altering the course of astronomical history.

2.5 ▪ KEPLER'S LAWS OF PLANETARY MOTION

As increasingly accurate observations of planetary motions were made, the flaws of both the Ptolemaic and Copernican models became more evident. Tycho Brahe (1546–1601) was probably the greatest astronomical observer prior to the invention of the telescope; it was his observations of planetary motions that both revealed the inadequacy of the Copernican system and provided the necessary data for calculating the true nature of planetary orbits around the Sun. Tycho was a Danish aristocrat and received large sums of money from the King of Denmark to set up an elaborate observatory on the island of Hven, near Copenhagen. For more than 20 years, Tycho observed the positions of planets and stars with an accuracy of 1 arcminute. Interestingly, Tycho did not believe that the heliocentric model was correct. He noted, as did the Greeks before him, that the stars do not show parallax. The absence of parallaxes larger than 1 arcminute implies that the nearest stars must be farther away than a few thousand AU, given a heliocentric solar system. Tycho thought this distance was implausibly large and thus devised a compound system in which all the planets other than the Earth went around the Sun, while the Sun orbited the Earth, carrying its entourage of planets along with it.

In the year 1599, after a major falling-out with the Danish king, Tycho accepted a post as Imperial Mathematician to the Holy Roman Emperor in Prague. There he hired a new assistant named Johannes Kepler (1571–1630). Initially, Kepler was frustrated by Tycho's reluctance to share his data. However, Kepler soon had complete access to Tycho's observations; in October 1601, less than two years after Kepler arrived in Prague, Tycho died, and Kepler was appointed his successor as Imperial Mathematician. By using Tycho's observations of the planet Mars, and by doing several years' worth of calculations, Kepler was able to formulate a mathematical description of its orbit, and by extension, the orbits of other planets. His basic findings are encapsulated in **Kepler's laws of planetary motion**.

1. **Kepler's first law:** *Planets travel on elliptical orbits with the Sun at one focus.* The properties of the closed curve known as an **ellipse** are best described by explaining how to draw one (Figure 2.17). Take a piece of string and tie each end to a pin. Stick the pins into a piece of paper, separated by a distance less than the string's length. Use a pencil to stretch the string taut and draw a complete, closed curve; this is an ellipse. The two pins are located at the **foci** of the ellipse.[14] Expressed mathematically, the ellipse is the locus of points for which the sum of the distances to the foci is a constant (equal to the length of the string, in our graphic example). If the pins are moved closer together, for a given length of string, the ellipse becomes more nearly circular; if they are moved farther apart, the ellipse becomes more flattened.

   The longest distance across the ellipse (which passes through both foci) is called the **major axis**. The shortest distance across the ellipse, passing through the ellipse's center, is called the **minor axis**. The **semimajor axis** is half the major axis, and the **semiminor axis** is half the minor axis. The **eccentricity** of the ellipse

[14] "Foci" is the plural of the word "focus."

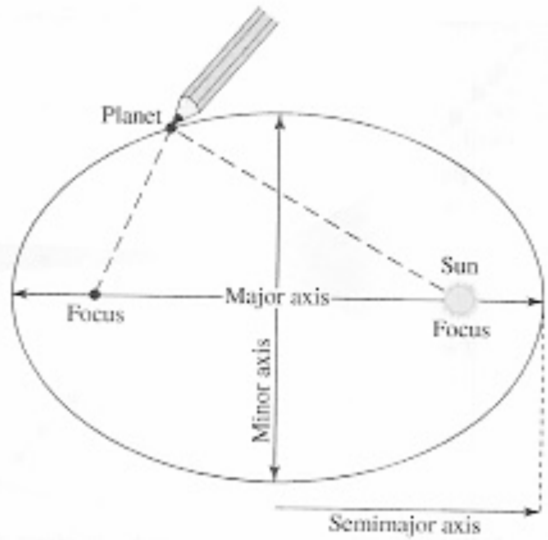


FIGURE 2.17 The properties of an ellipse.

is the distance between the foci divided by the length of the major axis. If the foci coincide, then $e = 0$, and the ellipse is a circle. The other limiting case, $e = 1$, represents the case in which the foci are separated by the full length of the string. It was quite a feat for Kepler to discover the elliptical shape of planetary orbits, since most planets have orbits with small eccentricity. Of the planets known to Kepler, Mercury had the largest eccentricity, $e = 0.21$; all the others had $e < 0.1$.

2. **Kepler's second law:** *A line drawn from the Sun to a planet sweeps out equal areas in equal time intervals.* This law provides a quantitative description of how the orbital speed of planets changes with their distance from the Sun; not only is motion not circular, Kepler discovered, it doesn't have uniform speed, either. The second law is graphically demonstrated in Figure 2.18. A mythical planet has its motion plotted during two time intervals, each 10 days long, separated by half the planet's orbital period. The two wedge-shaped areas swept out by the planet–Sun line are of equal area, even though they are of different shape. Kepler's second law implies that planets move most rapidly at **perihelion**, the point on their orbit closest to the Sun, and least rapidly at **aphelion**, the point farthest from the Sun.[15] As we show in Section 3.1, Kepler's second law is a simple consequence of the conservation of angular momentum.

3. **Kepler's third law:** *The squares of the sidereal orbital periods of the planets are proportional to the cubes of the semimajor axis of their orbits.* Kepler's third law

[15] Sometimes you hear "aphelion" pronounced as "ap-helion," sometimes as "af-felion." Both pronunciations can be found in reputable dictionaries.

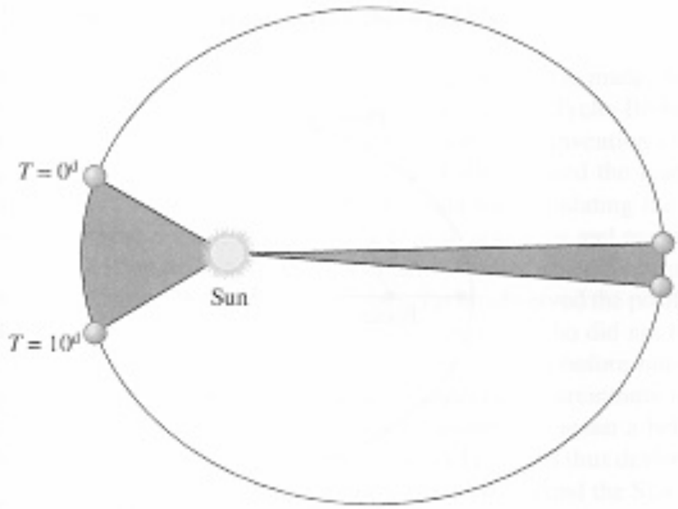


FIGURE 2.18 The area swept out by the planet–Sun line in each 10-day interval is identical.

can be expressed more compactly in mathematical notation:

$$P^2 = Ka^3, \tag{2.17}$$

where P is a planet's sidereal orbital period, a is the length of the semimajor axis of its orbit, and K is a constant. For objects orbiting the Sun,

$$K = 1\ \text{yr}^2\ \text{AU}^{-3}. \tag{2.18}$$

A plot of orbital period versus semimajor axis (like that of Figure 2.19) shows that all planets in the solar system, even those unknown to Kepler, follow his third law. In addition, Figure 2.19 shows that the Galilean satellites of Jupiter also obey equation (2.17), but with $K \approx 1050\ \text{yr}^2\ \text{AU}^{-3}$ rather than $K \approx 1\ \text{yr}^2\ \text{AU}^{-3}$.

2.6 ▪ PROOF OF THE EARTH'S MOTION

Although Galileo's discoveries convinced many individuals that the heliocentric model was correct, definitive proof that the Earth revolves around the Sun and rotates on its axis wasn't provided until much later. The rotation of the Earth about its axis was proved by detecting the Coriolis effect; this was done most famously by Jean Foucault, using what is now called a Foucault pendulum. The revolution of the Earth about the Sun was proved by detecting the effect known as aberration of starlight; later confirmation came from measuring the annual parallax of nearby stars.

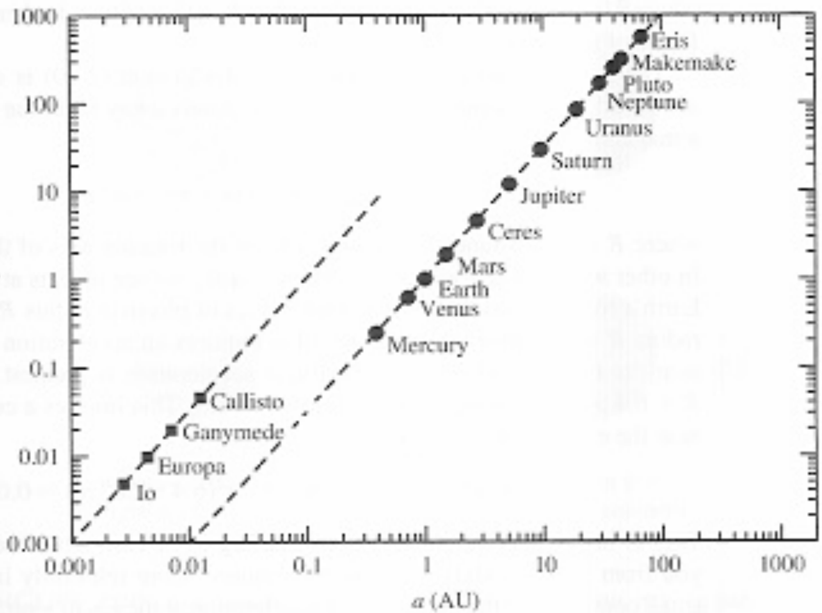


FIGURE 2.19 Orbital period P versus semimajor axis a for planets and dwarf planets orbiting the Sun (circular dots) and for the Galilean satellites orbiting Jupiter (square dots).

2.6.1 Rotation of the Earth

When we measure the trajectory of a projectile (such as a bullet or a thrown ball), we are measuring the trajectory relative to the Earth's surface. However, because of the Earth's rotation, any set of coordinates fixed to the Earth's surface is rotating with an angular velocity $\vec{\omega}$. The magnitude of $\vec{\omega}$ is $\omega \approx 2\pi \text{ day}^{-1} \approx 7.3 \times 10^{-5}\text{ s}^{-1}$, and the direction of $\vec{\omega}$ is pointing from south to north, parallel to the Earth's rotation axis. By watching the motion of the projectile, we can detect the Earth's rotation; its trajectory in the Earth's rotating frame of reference is subtly different from what it would be in a nonrotating frame of reference.

To quantify the difference in trajectories, let's start by writing down the relevant equations of motion. In a nonrotating frame, the motion of an object is famously given by Newton's second law of motion:

$$\vec{\mathbf{a}} = \vec{\mathbf{F}}/m, \tag{2.19}$$

where $\vec{\mathbf{a}}$ is the measured acceleration of the object, $\vec{\mathbf{F}}$ is the net force applied, and m is the object's mass. However, the equation of motion is different when the acceleration $\vec{\mathbf{a}}$ is measured in a frame of reference rotating with angular velocity $\vec{\omega}$:

$$\vec{\mathbf{a}} = \vec{\mathbf{F}}/m + 2(\vec{\mathbf{v}} \times \vec{\omega}) - \vec{\omega} \times (\vec{\omega} \times \vec{\mathbf{r}}), \tag{2.20}$$

where $\vec{\mathbf{v}}$ is the object's velocity and $\vec{\mathbf{r}}$ is the object's position, both measured in the rotating frame of reference.

The last term on the right-hand side of equation (2.20) is called the **centrifugal** acceleration. The centrifugal acceleration points away from the rotation axis, and has a magnitude

$$a_{\text{cent}} = |\vec{\omega} \times (\vec{\omega} \times \vec{\mathbf{r}})| = \omega^2 R, \tag{2.21}$$

where R is the distance of the object from the rotation axis of the frame of reference. In other words, when we rotate with the Earth, we see objects at a distance R from the Earth's rotation axis move in diurnal circles of physical radius R; motion in a circle of radius R with uniform angular speed ω requires an acceleration $a = \omega^2 R$. For objects near the Earth's surface, the centrifugal acceleration is greatest at the equator, where $R \approx 6.4 \times 10^6$ m is equal to the Earth's radius. This implies a centrifugal acceleration near the equator of

$$a_{\text{cent}} = \omega^2 R \approx (7.3 \times 10^{-5}\,\text{s}^{-1})^2 (6.4 \times 10^6\,\text{m}) \approx 0.034\,\text{m s}^{-2}. \tag{2.22}$$

This is not a large acceleration. In the jargon of auto advertisements, it would take you from "zero to sixty mph" in 13 minutes. More relevantly in this context, a_{cent} is small compared to the gravitational acceleration at the Earth's surface, $g = 9.8\,\text{m s}^{-2}$. In principle, traveling from the poles to the equator should reduce your acceleration toward the Earth's center, and thus reduce your weight. However, the fractional weight loss will be only $a_{\text{cent}}/g \approx 0.003$.

The middle term on the right-hand side of equation (2.20) is called the Coriolis acceleration, or the **Coriolis effect**, after a French scientist named Gustave Coriolis, who published the equations of motion for a rotating frame in the year 1835. It is sometimes computationally convenient to think of the Coriolis acceleration,

$$\vec{\mathbf{a}}_{\text{cor}} = 2(\vec{\mathbf{v}} \times \vec{\omega}), \tag{2.23}$$

as being due to a fictitious "Coriolis force" equal to $2m(\vec{\mathbf{v}} \times \vec{\omega})$. In truth, however, no physical force is being applied to the particle; the Coriolis acceleration results from the fact that the particle is being observed from a rotating, and hence accelerated, reference frame. The cross-product in equation (2.23) tells us that the Coriolis acceleration is always perpendicular to the direction of motion of the particle. When the cross-product is worked out in detail, it is seen that a moving particle is deflected to its right in the northern hemisphere and to its left in the southern hemisphere as Figure 2.20 demonstrates.

The magnitude of the Coriolis acceleration is

$$a_{\text{cor}} = 2v\omega \sin \Theta, \tag{2.24}$$

where Θ is the angle between $\vec{\mathbf{v}}$ and $\vec{\omega}$. Thus, the Coriolis effect is maximized when the particle's motion is perpendicular to the Earth's rotation axis; it vanishes when the particle's motion is parallel to the rotation axis. For other directions of motion, we may make the rough approximation

$$a_{\text{cor}} \sim v\omega. \tag{2.25}$$

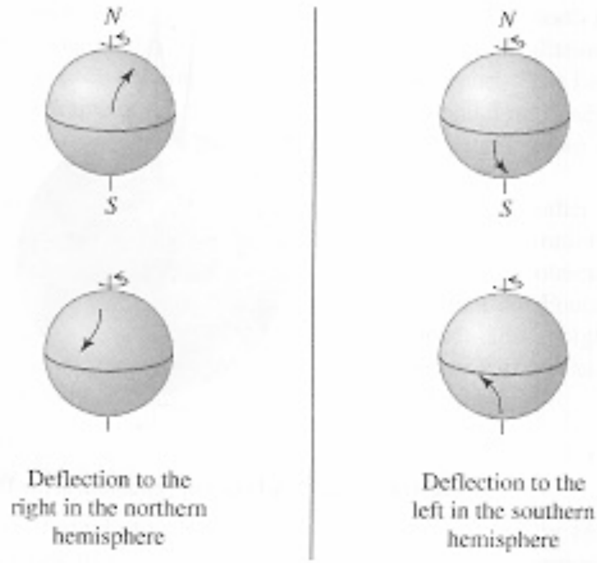


FIGURE 2.20 In a reference frame co-rotating with the Earth, moving particles are deflected to the right in the northern hemisphere, and to the left in the southern hemisphere.

If a particle is in flight for a time Δt, its velocity will be altered by a fractional amount

$$\frac{\Delta v}{v} \sim \frac{a_{\rm cor}\Delta t}{v} \sim \omega \Delta t. \tag{2.26}$$

Thus, the change in the particle's direction of motion will be small as long as its time of flight is much shorter than

$$\omega^{-1} \sim \frac{1}{2\pi}\text{ days} \sim 4\text{ hr} \sim 14{,}000\text{ s}. \tag{2.27}$$

Usually, when a ball is thrown or a bullet is fired, it reaches its target within a few seconds, so the Coriolis effect is negligible. However, the Coriolis acceleration can significantly affect the ballistic trajectory of projectiles when the time of flight is sufficiently long. During the projectile's flight, it will be deflected by a distance

$$\Delta d \sim \frac{1}{2}a_{\rm cor}(\Delta t)^2 \sim \frac{1}{2}v\omega(\Delta t)^2, \tag{2.28}$$

to the right of its initial trajectory in the northern hemisphere and to the left in the southern hemisphere. During World War I, for instance, the German army used an immense artillery gun to bombard Paris from a distance of ~ 120 km. The Paris Gun had a muzzle velocity $v \sim 1.6\text{ km s}^{-1}$; shells were sent on a parabolic trajectory with a maximum altitude of ~ 40 km and a time of flight $\Delta t \sim 170$ s. This led to a deflection

FIGURE 2.21 A Foucault pendulum at the Earth's north pole.

$$\Delta d \sim \frac{1}{2} v\omega (\Delta t)^2 \sim 2\ \text{km}, \tag{2.29}$$

to the right of where the gun was aimed.

The Coriolis acceleration also affects wind patterns. As air moves inward toward an area of low pressure, the Coriolis acceleration causes it to swerve to the right (in the northern hemisphere of Earth), and sets up a counterclockwise circulation. As a consequence, hurricanes in the northern hemisphere rotate counterclockwise; conversely, circular storms in the southern hemisphere rotate clockwise. Urban legend to the contrary, water draining from a sink doesn't invariably spiral counterclockwise in the northern hemisphere and clockwise in the southern hemisphere. Draining a sink takes much less time than forming a hurricane; during the time it takes a sink to empty, the Δv caused by the Coriolis effect remains small compared to the speed of the eddies that form as you fill the sink and wash your hands.[16]

A celebrated demonstration of the Coriolis effect is the **Foucault pendulum**, first demonstrated in the year 1851 by a French scientist named Jean Foucault. A Foucault pendulum is nothing more than a long pendulum suspended from a ball-and-socket joint overhead, so it is free to swing in any direction. Although Foucault set up his own pendulum in Paris, it is easier to visualize the principle behind the Foucault pendulum if we imagine one installed at the Earth's north pole (Figure 2.21). If we set the pendulum oscillating, it will continue to oscillate back and forth in the same plane, as viewed by a nonrotating observer. Thus, a sidereal nonrotating observer would report, "The

[16] In a classic experiment, A. H. Shapiro of MIT (latitude 42° N) managed to detect the Coriolis effect by filling a 6-foot diameter tank with water, letting it sit covered with a plastic sheet for 24 hours at constant temperature, then carefully pulling out the small, centrally located drain plug. Under such controlled conditions, the water did indeed spiral counterclockwise down the drain. (*Nature*, 1962, vol. 196, p. 1080).

Earth rotates counterclockwise (viewed from above the Earth's north pole), completing one rotation in a sidereal day; the plane of the pendulum's oscillation is not rotating." However, an observer co-rotating with the Earth would report, "The Earth is not rotating with respect to my frame of reference; the plane of the pendulum's oscillation is rotating clockwise (viewed from above the Earth's north pole), completing one rotation in a sidereal day."

Analyzing the rotation of a Foucault pendulum at locations other than the north or south pole requires a more detailed analysis of the Coriolis acceleration of the pendulum bob; the result found is that the pendulum's plane of oscillation rotates at a rate $2\pi \sin \ell$ radians per sidereal day, where ℓ is the latitude at which the Foucault pendulum is located. (This accounts for the popularity of Foucault pendulums at high-latitude science museums; near the equator, the excruciatingly slow rotation of a Foucault pendulum is a less visually exciting demonstration of the Earth's rotation.)

2.6.2 Revolution of the Earth

The **aberration of starlight** was first detected by Jean Picard in 1680, but it wasn't explained until 1729, by the astronomer James Bradley. The aberration of starlight is an effect that causes the apparent positions of stars on the celestial sphere to be deflected in the direction of the observer's motion. The common analogy to explain the aberration of starlight involves running through a rainshower with an umbrella; even if the rain is falling straight down, you have to tilt your umbrella in the direction of motion in order to keep your head dry. Similarly, in order to catch photons from a distant star, you have to tilt your telescope in the direction of motion (Figure 2.22). Photons travel at a large but finite speed, $c = 3.0 \times 10^5$ km s^{-1}. The orbital speed of the Earth averages

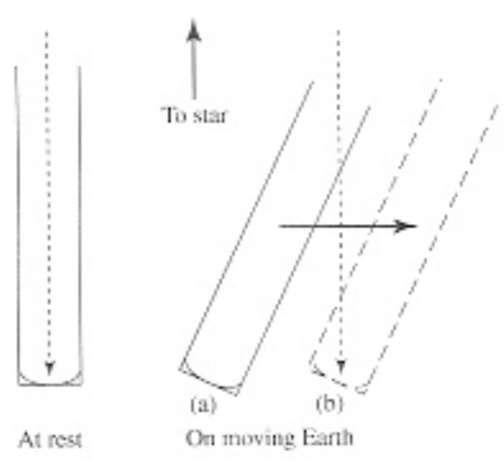


FIGURE 2.22 Telescopes must be tilted in the direction of the Earth's motion by an angle $\theta \approx v/c$ to assure that photons arrive at point P at the same time as the bottom of the telescope.

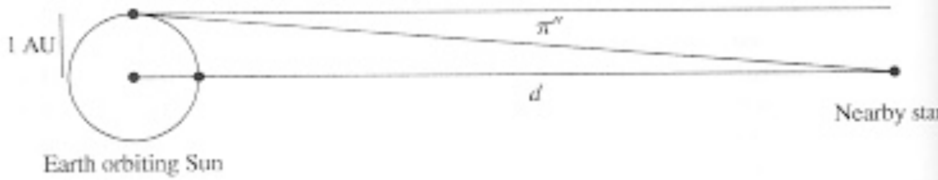


FIGURE 2.23 Definition of the parallax π'' of a star.

$v = 29.8\ \mathrm{km\ s^{-1}} \approx 10^{-4}\, c$. If your telescope is 1 m long, then during the time it takes light to pass through the telescope, the Earth's motion will have translated the telescope through a distance of 0.1 mm. Figure 2.22 shows that the angle through which the telescope must be tilted is given by the relation

$$\tan\theta = v/c. \tag{2.30}$$

Since the Earth's speed is so much smaller than the speed of light, we may use the small-angle approximation:

$$\theta \approx \frac{v}{c} \approx \frac{29.8\ \mathrm{km\ s^{-1}}}{3.0\times 10^5\ \mathrm{km\ s^{-1}}}\left(\frac{180^\circ}{\pi\ \mathrm{rad}}\right)\left(\frac{3600''}{1^\circ}\right) \approx 20.5''. \tag{2.31}$$

Aberration of starlight causes the positions of stars in the sky to follow an annual path that is the projection of the Earth's motion onto the sky: an ellipse of semimajor axis $20.5''$ and a semiminor axis $20.5''\beta$, where β is the angular distance of the star from the ecliptic.

Stellar parallax was introduced earlier, on page 37, when we emphasized the fact that observers couldn't detect it prior to the invention of the telescope. In fact, even after the invention of the telescope, it took a long time before stellar parallax was first measured. It wasn't until 1838, more than two centuries after the first telescopes, that the astronomer Friedrich Wilhelm Bessel announced that he had finally measured the annual parallax of a star. Formally, astronomers define parallax π'' (Figure 2.23) as the apparent displacement of a star, in arcseconds, due to a change in the position of the observer by 1 AU perpendicular to the line of sight to the star.[17] Although parallaxes are defined in terms of a 1 AU displacement, the actual baseline used for parallax measurements can be as large as 2 AU, by using observations six months apart at the appropriate times of year. From Figure 2.23, we see that the distance d from the Sun to another star is simply related to the star's parallax:

$$d = \frac{a}{\tan\pi''}. \tag{2.32}$$

[17] We avoid confusion with the irrational number $\pi = 3.14159265\ldots$ by using the double prime, gently reminding us that parallaxes are generally measured in units of arcseconds.

Using the small-angle approximation, and converting the parallax from radians to arcseconds, we find that

$$d = \frac{a}{\pi''[\text{arcsec}]}\left(\frac{180^\circ}{\pi\ \text{rad}}\right)\left(\frac{3600''}{1^\circ}\right) = \frac{206{,}265\ \text{AU}}{\pi''}. \tag{2.33}$$

The distance at which a star has a parallax of exactly 1″ is known as the **parsec**, short for "**par**allax of one arc**sec**." The number of AU in one parsec is equal to the number of arcseconds in a radian: 206,265. The nearest star to the Sun, Proxima Centauri, has a parallax $\pi'' = 0.76''$, and hence is at a distance $d = 270{,}000$ AU $= 1.3$ parsecs. Stellar parallax causes the positions of stars on the celestial sphere to follow a path that is the projection of the Earth's orbit onto the sky: an ellipse with semimajor axis π'' and semiminor axis $\pi''\beta$, where β is the angular distance of the star from the ecliptic.[18] It took a while for stellar parallax to be measured, but when it was, it confirmed two initially controversial assertions made by Copernicus. First, the Earth goes around the Sun, rather than vice versa. Second, space is big (really big).

PROBLEMS

2.1 Over the course of the year, which gets more hours of daylight, the Earth's north pole or south pole? (Hint: the Earth is at perihelion in January.)

2.2 On 2003 August 27, Mars was in opposition as seen from the Earth. On 2005 July 14 (687 days later), Mars was in western quadrature as seen from the Earth. What was the distance of Mars from the Sun on these dates, measured in astronomical units (AU)? Is this greater than or less than the semimajor axis length of the Martian orbit? You may assume the Earth's orbit is a perfect circle. (Hint: the sidereal period of Mars is also 687 days.)

2.3 In the 1670s, the astronomer Ole Rømer observed eclipses of the Galilean satellite Io as it plunged through Jupiter's shadow once per orbit. He noticed that the time between observed eclipses became shorter as Jupiter came closer to the Earth and longer as Jupiter moved away. Rømer calculated that the eclipses were observed 17 minutes earlier when Jupiter was in opposition compared to when it was close to conjunction. This was attributed by Rømer to the finite speed of light. From Rømer's data, compute the speed of light, first in AU min^{-1}, then in m s^{-1}.

[18] Note that the aberrational shift of 20.5″, which is independent of the star's distance, is much greater than the parallactic shift for even the nearest stars, which have $\pi'' \leq 0.75''$.

3 Orbital Mechanics

Isaac Newton (1642/3–1727) was born in rural England; his birth date was 1642 December 25 according to the Julian calendar (still in use in England at the time), but 1643 January 4 according to the Gregorian calendar. When young Newton proved to be incompetent at managing his family's farm, he was sent to Cambridge University and started to thrive as a scholar. In 1665, the year in which Newton earned his bachelor's degree, an outbreak of the plague closed down the university, and Newton retreated to his family's farm and began to think—very hard. The period when the university was closed was Newton's *annus mirabilis*, during which he discovered calculus, formulated his three **laws of motion** and his **law of universal gravitation**, and performed ground-breaking experiments in optics. Much of the remainder of Newton's long life was dedicated to developing the ideas he had in this burst of youthful creativity.[1]

Newton didn't publish his laws of motion and law of universal gravitation until 1687, when his book *Philosophiae Naturalis Principia Mathematica* ("Mathematical Principles of Natural Philosophy") was published. The laws of motion can be summarized as follows:

1. An object's velocity remains constant unless a net outside force acts upon it.
2. If a net outside force acts on an object, its acceleration is directly proportional to the force and inversely proportional to the mass of the object. In short, $\vec{F} = m\vec{a}$, where $\vec{F}$ is the outside force, m is the mass, and $\vec{a}$ is the acceleration.
3. Forces come in pairs, equal in magnitude and opposite in direction. (As Newton put it: *Actioni contrariam semper et aequalem esse reactionem*, or "Every action has an equal and opposite reaction.")

Newton's law of universal gravitation can be concisely expressed in mathematical form. Suppose that two spherical objects, of mass M and m, are separated by a distance r.

[1] He also performed many alchemical experiments while trying to systematize chemistry in the way he did physics, not to mention writing reams of theological works, becoming Master of the Royal Mint, and serving as president of the Royal Society for nearly a quarter-century.

(The distance r is measured between the centers of the two objects.) Newton's law tells us that the gravitational attraction between the two objects is

$$F = -\frac{GMm}{r^2}, \tag{3.1}$$

where G, called the **gravitational constant**, is a universal constant whose value is $G = 6.67 \times 10^{-11}\ \mathrm{Nm^2\ kg^{-2}}$ (where N stands for newton).[2] The negative sign in equation (3.1) tells us that gravity is always an attractive force.

3.1 ▪ DERIVING KEPLER'S LAWS

Newton derived the form of equation (3.1) by requiring that the force of gravity result in planetary orbits that obey Kepler's laws of planetary motion. Newton was solving the problem in the difficult direction: he deduced the form of the law of gravitation starting from the observations. Since we aren't as smart as Newton, we will take the easier direction in the following section; starting with Newton's law of universal gravitation, we'll show that Kepler's laws follow as a consequence. Although it may seem numerically incongruous, the derivations will flow more smoothly if we begin by deriving Kepler's second law, then go on to the first and third laws.

3.1.1 Kepler's Second Law

Gravity is an example of a **central force**, defined as a force directed straight toward or away from some central point, with a magnitude that depends only on the distance r from that point. The gravitational force qualifies as a central force because the force $\vec{\mathbf{F}}$ acting on the mass m always points toward the mass M (the central point of the force), and the magnitude of the gravitational force is $\propto 1/r^2$, where r is the separation of the two masses.[3] While analyzing the motion of a particle responding to a central force, it is convenient to be able to switch from Cartesian coordinates to polar coordinates.

In a Cartesian coordinate system (Figure 3.1), the unit vectors along the x, y, and z axes are $\hat{\mathbf{i}}$, $\hat{\mathbf{j}}$, and $\hat{\mathbf{k}}$, respectively. Suppose we choose our Cartesian coordinate axes such that the larger mass M lies at the origin, and the position $\vec{\mathbf{r}}$ and velocity $\vec{\mathbf{v}}$ of the smaller mass m lie in the xy plane. (For the sake of concreteness, let's call mass M the Sun, and mass m a planet, although the situation applies in general to any system of two spherical masses: a planet and a moon, a planet and an artificial satellite, a supermassive black hole and a star—you name it.) The planet's position (x, y) can now be expressed in polar coordinates, where the polar coordinates (r, θ) are related to the Cartesian coordinates (x, y) by the relations $x = r\cos\theta$ and $y = r\sin\theta$. In polar coordinates, as illustrated in

[2] The newton (N)—the force required to accelerate 1 kilogram at one meter per second per second—is equivalent to 3.6 ounces, or about the weight of a small apple.

[3] The electrostatic repulsion or attraction between two charged particles is another example of a central force.

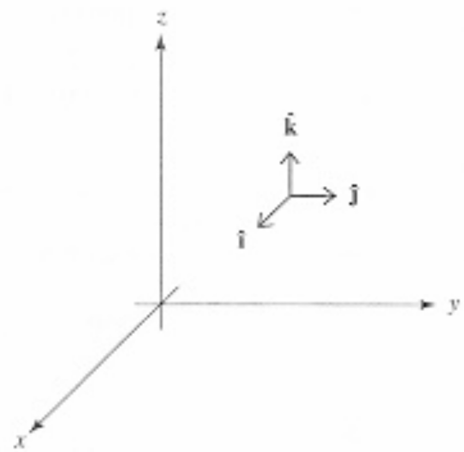


FIGURE 3.1 Axes and unit vectors in a Cartesian coordinate system.

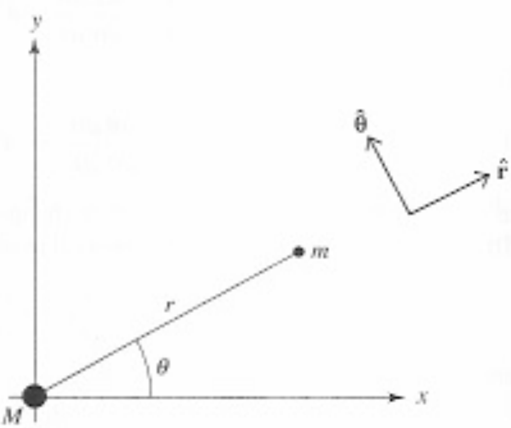


FIGURE 3.2 Axes and unit vectors in a polar coordinate system.

Figure 3.2, the unit vectors $\hat{\mathbf{r}}$ and $\hat{\boldsymbol{\theta}}$ are

$$\hat{\mathbf{r}} = \hat{\imath}\cos\theta + \hat{\jmath}\sin\theta \tag{3.2}$$

and

$$\hat{\boldsymbol{\theta}} = -\hat{\imath}\sin\theta + \hat{\jmath}\cos\theta. \tag{3.3}$$

The dot product (or scalar product) of these unit vectors is

$$\hat{\mathbf{r}} \cdot \hat{\boldsymbol{\theta}} = -\cos\theta\sin\theta + \sin\theta\cos\theta = 0. \tag{3.4}$$

and their cross product (or vector product) is

$$\hat{\mathbf{r}} \times \hat{\boldsymbol{\theta}} = \begin{vmatrix} \hat{\mathbf{i}} & \hat{\mathbf{j}} & \hat{\mathbf{k}} \\ \cos\theta & \sin\theta & 0 \\ -\sin\theta & \cos\theta & 0 \end{vmatrix} = \hat{\mathbf{k}}(\cos^2\theta + \sin^2\theta) = \hat{\mathbf{k}}, \tag{3.5}$$

thus demonstrating that $\hat{\mathbf{r}}$ and $\hat{\boldsymbol{\theta}}$ are mutually orthogonal as well as being orthogonal to $\hat{\mathbf{k}}$, the unit vector in the z direction.

From equations (3.2) and (3.3), we see that

$$\frac{d\hat{\mathbf{r}}}{d\theta} = \frac{d}{d\theta}(\hat{\mathbf{i}}\cos\theta + \hat{\mathbf{j}}\sin\theta) = -\hat{\mathbf{i}}\sin\theta + \hat{\mathbf{j}}\cos\theta = \hat{\boldsymbol{\theta}} \tag{3.6}$$

and

$$\frac{d\hat{\boldsymbol{\theta}}}{d\theta} = \frac{d}{d\theta}(-\hat{\mathbf{i}}\sin\theta + \hat{\mathbf{j}}\cos\theta) = -\hat{\mathbf{i}}\cos\theta - \hat{\mathbf{j}}\sin\theta = -\hat{\mathbf{r}}. \tag{3.7}$$

We can then apply the chain rule to find the rate of change of the unit vectors $\hat{\mathbf{r}}$ and $\hat{\boldsymbol{\theta}}$:

$$\frac{d\hat{\mathbf{r}}}{dt} = \frac{d\hat{\mathbf{r}}}{d\theta}\frac{d\theta}{dt} = \hat{\boldsymbol{\theta}}\frac{d\theta}{dt} \tag{3.8}$$

and

$$\frac{d\hat{\boldsymbol{\theta}}}{dt} = \frac{d\hat{\boldsymbol{\theta}}}{d\theta}\frac{d\theta}{dt} = -\hat{\mathbf{r}}\frac{d\theta}{dt}. \tag{3.9}$$

Note that since $\hat{\mathbf{r}}$ and $\hat{\boldsymbol{\theta}}$ are unit vectors, they change only in direction, not in magnitude.

The velocity of the planet can be expressed in polar coordinates as

$$\vec{\mathbf{v}} \equiv \frac{d\vec{\mathbf{r}}}{dt} = \frac{d(r\hat{\mathbf{r}})}{dt} = \frac{dr}{dt}\hat{\mathbf{r}} + r\frac{d\hat{\mathbf{r}}}{dt} = v_r\hat{\mathbf{r}} + v_t\hat{\boldsymbol{\theta}}, \tag{3.10}$$

where

$$v_r = \frac{dr}{dt} \tag{3.11}$$

is the **radial velocity** and

$$v_t = r\frac{d\theta}{dt} \tag{3.12}$$

is the **tangential velocity**.

The angular momentum of the planet is defined as

$$\vec{\mathbf{L}} = \vec{\mathbf{r}} \times \vec{\mathbf{p}}, \tag{3.13}$$

where $\vec{\mathbf{p}} = m\vec{\mathbf{v}}$ is the linear momentum. The rate of change of the angular momentum is then

$$\frac{d\vec{\mathbf{L}}}{dt} = \frac{d\vec{\mathbf{r}}}{dt} \times \vec{\mathbf{p}} + \vec{\mathbf{r}} \times \frac{d\vec{\mathbf{p}}}{dt} = \vec{\mathbf{v}} \times m\vec{\mathbf{v}} + \vec{\mathbf{r}} \times m\frac{d\vec{\mathbf{v}}}{dt}. \tag{3.14}$$

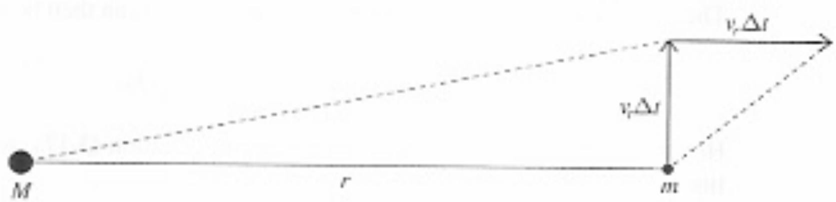


FIGURE 3.3 The motions of a planet during a short time interval Δt.

From Newton's second law of motion, we know that $m d\vec{\mathbf{v}}/dt = \vec{\mathbf{F}}$. Thus, equation (3.14) can be rewritten as

$$\frac{d\vec{\mathbf{L}}}{dt} = m(\vec{\mathbf{v}} \times \vec{\mathbf{v}}) + \vec{\mathbf{r}} \times \vec{\mathbf{F}}. \tag{3.15}$$

However, $\vec{\mathbf{v}} \times \vec{\mathbf{v}} = 0$ (that's just a vector identity), and for a central force, $\vec{\mathbf{F}}$ is parallel to $\vec{\mathbf{r}}$ and thus $\vec{\mathbf{F}} \times \vec{\mathbf{r}} \propto \vec{\mathbf{r}} \times \vec{\mathbf{r}} = 0$. We conclude that for gravity or any other central force, angular momentum is conserved:

$$\frac{d\vec{\mathbf{L}}}{dt} = 0. \tag{3.16}$$

Note that the direction as well as the magnitude of $\vec{\mathbf{L}}$ is constant; this tells us that the motion of an object moving under the influence of a central force is confined to a plane.

The conservation of angular momentum is equivalent to Kepler's second law; to demonstrate that this is true, we use equation (3.10) to write the angular momentum explicitly as

$$\vec{\mathbf{L}} = \vec{\mathbf{r}} \times m\vec{\mathbf{v}} = m r v_t \hat{\mathbf{k}} = L\hat{\mathbf{k}}, \tag{3.17}$$

where v_t is the tangential velocity. Referring to Figure 3.3, consider a planet of mass m; at a time t, it is at a distance r from the Sun, which has mass M. During a brief time interval Δt, the planet moves a distance $v_t \Delta t$ in the tangential direction and a distance $v_r \Delta t$ in the radial direction. The area ΔA swept out by the planet–Sun line during this brief interval can be approximated as the sum of two triangles:

$$\Delta A \approx \frac{1}{2} r (v_t \Delta t) + \frac{1}{2}(v_r \Delta t)(v_t \Delta t), \tag{3.18}$$

where the two terms represent the left-hand triangle and the right-hand triangle in Figure 3.3.[4] In the limit $v_r \Delta t \ll r$, the right-hand triangle is vanishingly small compared to the left-hand triangle, and the area swept out can be further simplified as

$$\Delta A \approx \frac{1}{2} r (v_t \Delta t). \tag{3.19}$$

[4] In Figure 3.3, we are looking at the specific case $v_r > 0$, but performing a time reversal will yield the case $v_r < 0$.

The rate at which the planet–Sun line sweeps out area can then be written

$$\lim_{\Delta t \to 0} \frac{\Delta A}{\Delta t} = \frac{dA}{dt} = \frac{1}{2} r v_t. \tag{3.20}$$

However, since we know that $L = mrv_t$, from equation (3.17), we can rewrite equation (3.20) in the form

$$\frac{dA}{dt} = \frac{1}{2}\frac{L}{m}. \tag{3.21}$$

Since L and m are constant, so is the rate dA/dt at which the planet–Sun line sweeps out area. In other words, we have demonstrated that Kepler's second law will be true for a body acting under any central force, not just the force of gravity.

3.1.2 Kepler's First Law

To demonstrate that Kepler's first law follows from Newton's law of universal gravitation, we will have to demonstrate that the trajectory $r(\theta)$ of the mass m (the planet) constitutes an ellipse with the larger mass M (the Sun) at one focus. Using equations (3.12) and (3.17), we can write the angular momentum per unit mass of the orbiting body as

$$\frac{L}{m} = r^2 \frac{d\theta}{dt}, \tag{3.22}$$

which is constant for any central force. If the force acting on the mass m is gravitational, then from Newton's law of universal gravitation and second law of motion,

$$\vec{\mathbf{F}} = -\frac{GMm}{r^2}\hat{\mathbf{r}} = m\frac{d\vec{\mathbf{v}}}{dt}. \tag{3.23}$$

The orbital acceleration under the influence of gravity is then

$$\frac{d\vec{\mathbf{v}}}{dt} = -\frac{GM}{r^2}\hat{\mathbf{r}}. \tag{3.24}$$

From equation (3.9), we know that

$$\hat{\mathbf{r}} = -\left(\frac{d\theta}{dt}\right)^{-1} \frac{d\hat{\boldsymbol{\theta}}}{dt}. \tag{3.25}$$

By combining equations (3.24) and (3.25), we find that the acceleration of the planet is

$$\frac{d\vec{\mathbf{v}}}{dt} = \frac{GM}{r^2}\left(\frac{d\theta}{dt}\right)^{-1} \frac{d\hat{\boldsymbol{\theta}}}{dt}. \tag{3.26}$$

Combining this equation with equation (3.22), we see

$$\frac{L}{GMm}\frac{d\vec{\mathbf{v}}}{dt} = \frac{d\hat{\boldsymbol{\theta}}}{dt}. \tag{3.27}$$

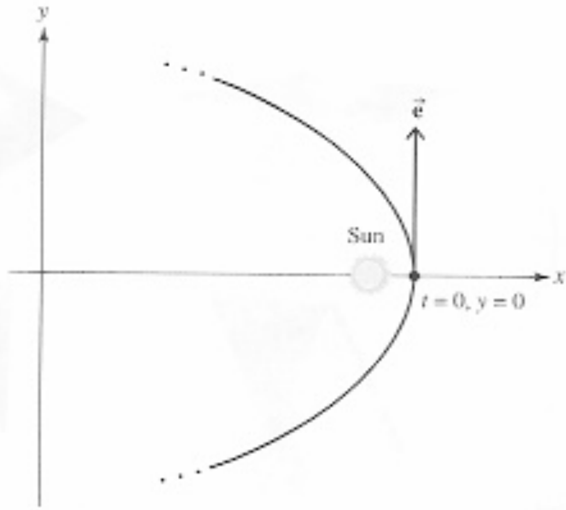


FIGURE 3.4 Time $t = 0$ corresponds to perihelion passage, with the planet crossing the x axis with its velocity in the positive y direction.

Integration of this simple differential equation yields

$$\frac{L}{GMm}\vec{v} = \hat{\theta} + \vec{e}, \tag{3.28}$$

where $\vec{e}$ is a constant of integration that depends on the initial conditions of the orbiting planet. We may choose the initial conditions for our own convenience. Let's choose the time $t = 0$ to correspond to a perihelion passage of the planet, and orient the axes so that perihelion passage occurs on the positive x axis (Figure 3.4). With this choice of coordinates, $\vec{v}$ and $\hat{\theta}$ both point in the y direction at $t = 0$; thus, we may write $\vec{e} = e\hat{j}$, where e is a constant. Equation (3.28) is then

$$\frac{L}{GMm}\vec{v} = \hat{\theta} + e\hat{j}. \tag{3.29}$$

We now take the dot product of this equation and the unit vector $\hat{\theta}$:

$$\frac{L}{GMm}\vec{v}\cdot\hat{\theta} = \hat{\theta}\cdot\hat{\theta} + e\hat{j}\cdot\hat{\theta}. \tag{3.30}$$

To simplify the right-hand side of equation (3.30), we use equation (3.3) to find that $\hat{j}\cdot\hat{\theta} = \cos\theta$. To simplify the left-hand side, we write

$$\vec{v}\cdot\hat{\theta} = \left[v_r\hat{r} + v_t\hat{\theta}\right]\cdot\hat{\theta} = v_t. \tag{3.31}$$

But, since equation (3.17) tells us that $mrv_t = L$, we may write

$$\vec{v}\cdot\hat{\theta} = v_t = \frac{L}{mr}. \tag{3.32}$$

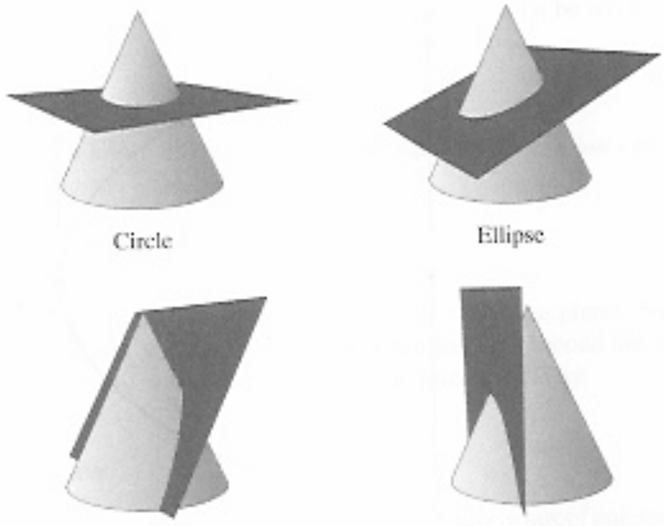


FIGURE 3.5 Conic sections demonstrated by slicing a cone.

Substituting equation (3.32) back into equation (3.30), we find a relationship between r and θ for fixed values of M, m, L, and e:

$$\frac{L^2}{GMm^2r} = 1 + e\cos\theta, \tag{3.33}$$

which can also be written in the form

$$r = \frac{L^2}{GMm^2(1 + e\cos\theta)}. \tag{3.34}$$

Equation (3.34) is the equation of a **conic section** in polar coordinates; as such, it provides a generalization of Kepler's first law.

Conic sections can be obtained by slicing a cone with a plane, as illustrated in Figure 3.5. If the plane is perpendicular to the cone's axis, then the conic section is a **circle**; from equation (3.34), we see that a circle corresponds to the special case $e = 0$, and hence $r = L^2/(GMm^2) =$ constant. If the slicing plane is tilted from the perpendicular by an angle less than the half-opening angle of the cone, the conic section obtained is an **ellipse**; this corresponds to the special case $0 < e < 1$.[5] When the slicing plane is tilted from the perpendicular by an angle exactly equal to the half-opening angle of the cone, the conic section resulting is a **parabola**; this is the special case $e = 1$. Finally, when the slicing plane is tilted by a larger angle, the conic section that results is a **hyperbola**,

[5] Yes, the parameter e in equation (3.34) is the same as the eccentricity e that we encountered while discussing elliptical orbits in Section 2.5, that is, the distance between foci divided by the major axis length.

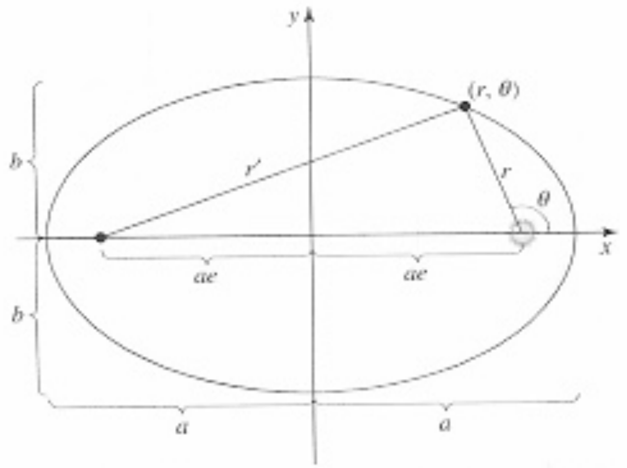


FIGURE 3.6 An ellipse of semimajor axis a and semiminor axis b.

which has $e > 1$. Kepler's first law is thus a special case that deals with **closed orbits**; that is, orbits with $e < 1$, which form closed curves (ellipses or circles). The basic physics of gravitation, however, permits **open orbits** as well, that is, parabolic or hyperbolic orbits with $e \geq 1$.

We have blithely asserted that the parameter e in equation (3.34), when it lies in the range $0 \leq e < 1$, is precisely the same as the eccentricity of an ellipse, defined as the distance between the foci divided by the length of the major axis. It is time to support that assertion by looking at the properties of ellipses in more depth. In Figure 3.6, an ellipse is shown along with a set of Cartesian coordinates; the origin of the coordinates is the center of the ellipse; the x axis lies along the major axis of the ellipse; and the y axis lies along the minor axis. We also define a system of polar coordinates centered on one of the foci. Let's call the focus at the origin the **principal focus** and require that it be the focus where the Sun is located, if the ellipse is regarded as a planetary orbit. The angular coordinate θ is measured counterclockwise from the x axis in the manner shown in Figure 3.6. The semimajor axis has length a and the semiminor axis has length b; each of the foci is displaced from the origin of the Cartesian coordinates by a distance ae. An arbitrary point on the ellipse is displaced by a distance r from the principal focus and a distance r' from the other focus; the basic property of an ellipse is that $r + r'$ is constant. By considering the two points of the ellipse lying on the x axis ($x = \pm a$, $y = 0$), we find that $r + r' = 2a$. It also follows that the perihelion distance, if the ellipse is regarded as a planetary orbit, is $q = a(1 - e)$ and the aphelion distance is $Q = a(1 + e)$.

Consider the point of the ellipse that lies on the positive y axis, where $r = r' = a$ as shown in Figure 3.7. From the Pythagorean theorem, as applied to the right triangle drawn in the figure, we find that $b^2 + (ae)^2 = r^2$, or since $r = a$,

$$b^2 = a^2(1 - e^2). \tag{3.35}$$

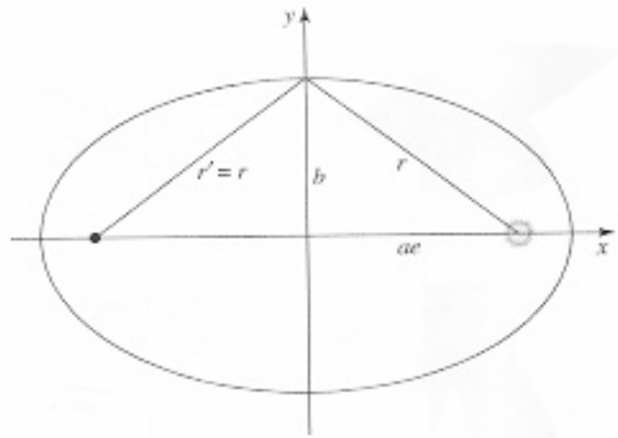


FIGURE 3.7 The relationship among the semimajor axis a, the semiminor axis b, and the eccentricity e.

This enables us to translate between the axis ratio of an ellipse, b/a, and its eccentricity,

$$e = (1 - b^2/a^2)^{1/2}. \tag{3.36}$$

It can also be shown that the average distance of all points on the ellipse from either focus is equal to the semimajor axis length a. To prove this, consider an arbitrary point on the ellipse, $P(x, y)$, and its reflection across the y axis, $P'(-x, y)$, as shown in Figure 3.8. The distance from point P to the focus on the positive x axis is r. By symmetry, the distance from the complementary point P' to the focus on the positive x axis is r', where r' is the distance from point P to the focus on the negative x axis. The average distance of the two points from the focus on the positive x axis is then

$$\langle r \rangle = \frac{r + r'}{2} = \frac{2a}{2} = a. \tag{3.37}$$

Since this relation holds for all (P, P') pairs, regardless of the choice of P, it is true that the average distance $\langle r \rangle$ from the focus over the entire ellipse is a.

Let us now describe the ellipse in terms of the polar coordinates (r, θ), where r is the distance from the principal focus and θ is the polar angle measured counterclockwise from the positive x axis, as shown in Figure 3.9. (When the ellipse represents an orbit, the angle θ is called the **true anomaly**.) Note in the figure that we can draw a triangle from the principal focus at $r = 0$, to an arbitrary point (r, θ) on the ellipse, to the other focus, then back to the principal focus. The internal angle of the vertex at the principal focus (as shown in Figure 2.17) is $\pi - \theta$. We can thus use the law of cosines to write

$$r'^2 = r^2 + (2ae)^2 - 2(2ae)r\cos(\pi - \theta). \tag{3.38}$$

Using the trigonometric identity $\cos(\pi - \theta) = -\cos\theta$, this becomes

$$r'^2 = r^2 + 4a^2e^2 + 4aer\cos\theta. \tag{3.39}$$

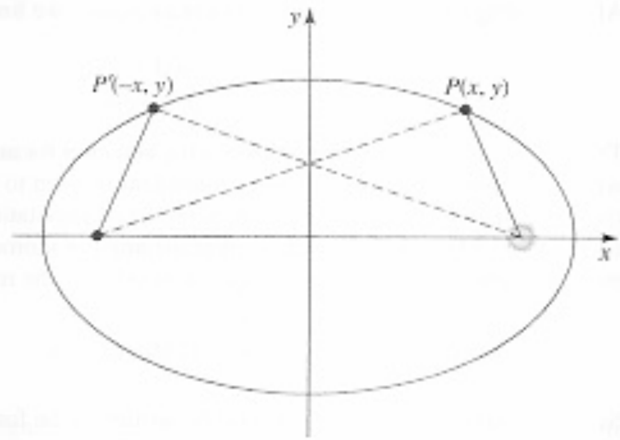


FIGURE 3.8 The point $P(x, y)$ is at a distance r from the focus on the positive x axis and a distance r' from the other focus. The complementary point $P'(-x, y)$ is at a distance r' from the focus on the positive x axis and a distance r from the other focus.

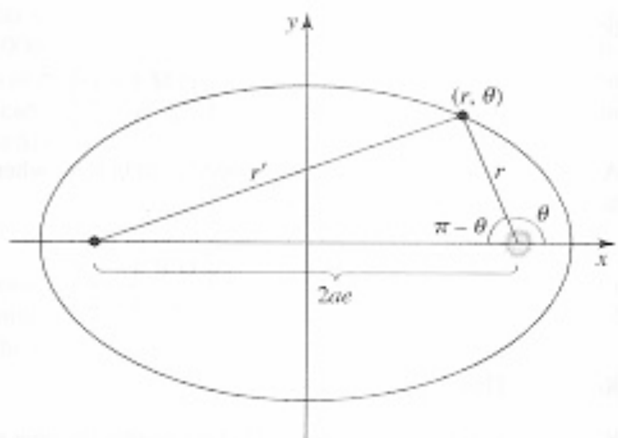


FIGURE 3.9 An ellipse in polar coordinates.

However, from the definition of the ellipse, we know that $r' = 2a - r$, which yields (squaring each side of the equation)

$$r'^2 = 4a^2 - 4ar + r^2. \tag{3.40}$$

Since the right-hand sides of equations (3.39) and (3.40) are equal, this tells us

$$4a^2e^2 + 4aer\cos\theta = 4a^2 - 4ar. \tag{3.41}$$

After dividing by $4a$ and doing a bit of rearranging, we find

$$r = \frac{a(1-e^2)}{1+e\cos\theta}. \tag{3.42}$$

This equation for r as a function of θ is the equation for an ellipse in polar coordinates, with the origin at one focus. This is equivalent in form to equation (3.34), which gives the shape of an orbit if Newton's law of universal gravitation holds true. Comparison of equations (3.34) and (3.42) tells us that the angular momentum L of a planet's orbital motion is related to the size and shape of its orbit by the relation

$$\frac{L^2}{m^2} = GMa(1-e^2). \tag{3.43}$$

Since $L = mrv_t$, this relation can also be written in the form

$$r^2 v_t^2 = GMa(1-e^2). \tag{3.44}$$

When a planet is at perihelion, its velocity is entirely tangential ($v_{\rm pe} = v_t$), and its distance from the Sun is $q = a(1-e)$. This implies that for a planet at perihelion,

$$v_{\rm pe}^2 a^2 (1-e)^2 = GMa(1-e^2), \tag{3.45}$$

or

$$v_{\rm pe} = \left[\frac{GM}{a}\frac{1+e}{1-e}\right]^{1/2}. \tag{3.46}$$

A similar analysis of the planet's speed at aphelion, where its velocity is also entirely tangential ($v_{\rm ap} = v_t$), tells us that

$$v_{\rm ap} = \left[\frac{GM}{a}\frac{1-e}{1+e}\right]^{1/2}. \tag{3.47}$$

3.1.3 Kepler's Third Law

Kepler's second law (equation 3.21) tells us that the area swept out per unit time by the planet–Sun line is a constant, $L/(2m)$. The area swept out in one orbital period, P, is the area of the ellipse, given by the standard formula $A = \pi ab$. For one complete orbital period, then, we may write

$$\frac{\pi ab}{P} = \frac{L}{2m}. \tag{3.48}$$

By squaring this equation and making the substitution $b^2 = a^2(1-e^2)$, we have

$$\frac{\pi^2 a^4 (1-e^2)}{P^2} = \frac{L^2}{4m^2}. \tag{3.49}$$

Since equation (3.43) gives us a relation among L, a, and e, namely,

$$\frac{L^2}{m^2} = GMa(1-e^2), \tag{3.50}$$

we can substitute back into equation (3.49) to find

$$\frac{\pi^2 a^4 (1-e^2)}{P^2} = \frac{GMa(1-e^2)}{4}, \tag{3.51}$$

or

$$P^2 = \frac{4\pi^2}{GM} a^3, \tag{3.52}$$

which we recognize as Kepler's third law, $P^2 = Ka^3$, with the proportionality constant $K \propto 1/M$. With somewhat more exertion, taking into account the acceleration of the Sun (mass M) as well as the lower-mass planet (mass m), it is possible to reach the more general form

$$P^2 = \frac{4\pi^2}{G(M+m)} a^3. \tag{3.53}$$

Within the solar system, however, even the most massive of the planets, Jupiter, has a mass only 1/1000 that of the Sun, so the approximation $M + m \approx M$ is adequate.

The masses of celestial bodies are measured by how they accelerate nearby masses. In particular, we can use the orbital periods and semimajor axes of the planets to determine the mass of the Sun:

$$M = \frac{4\pi^2 a^3}{GP^2}. \tag{3.54}$$

The orbital period of the Earth, for instance, is 365.256 days $\times$ 86,400 s day^{-1} = 3.16×10^7 s.[6] The semimajor axis of the Earth's orbit is $a = 1\,\mathrm{AU} = 1.496 \times 10^{11}$ m. Thus, we can compute the mass of the Sun as

$$\begin{aligned} M &= \frac{4\pi^2 (1.496 \times 10^{11}\,\mathrm{m})^3}{6.67 \times 10^{-11}\,\mathrm{m^3\,s^{-2}\,kg^{-1}} (3.16 \times 10^7\,\mathrm{s})^2} \\ &= 1.98 \times 10^{30}\,\mathrm{kg} = 1 M_\odot. \end{aligned} \tag{3.55}$$

Later in this book, we will find that the solar mass ($M_\odot$) is a useful unit for expressing the masses of stars (and larger objects).[7]

[6] A useful approximation is that the length of the year is $\pi \times 10^7$ s.

[7] The "dot in a circle" symbol $\odot$ is the standard astronomical symbol for the Sun. It is of great antiquity, being identical to the Egyptian hieroglyph for the Sun god Ra, seen here, for instance, as the first syllable in the name of the pharaoh Ramses the Great: 

3.2 ▪ ORBITAL ENERGETICS

Suppose you place a particle of mass m at a location $\vec{\mathbf{r}}$ relative to an object of mass M; you give it a kick so that it is initially moving at a velocity $\vec{\mathbf{v}}$. What determines whether its orbit is closed (a circle or ellipse, with $e < 1$) or open (a parabola or hyperbola, with $e \geq 1$)? In a sense, it's all about the energy. The particle will have an energy E that is the sum of its kinetic energy K and its gravitational potential energy U:

$$E = K + U = \frac{1}{2}mv^2 - \frac{GMm}{r}. \tag{3.56}$$

The square of the velocity can be determined by squaring equation (3.28):

$$\left(\frac{L}{GMm}\right)^2 \vec{\mathbf{v}} \cdot \vec{\mathbf{v}} = \hat{\boldsymbol{\theta}} \cdot \hat{\boldsymbol{\theta}} + 2e\hat{\boldsymbol{\theta}} \cdot \hat{\mathbf{j}} + e^2 \hat{\mathbf{j}} \cdot \hat{\mathbf{j}}$$

$$\left(\frac{L}{GMm}\right)^2 v^2 = 1 + 2e\hat{\boldsymbol{\theta}} \cdot \hat{\mathbf{j}} + e^2. \tag{3.57}$$

Since $\hat{\boldsymbol{\theta}} \cdot \hat{\mathbf{j}} = \cos\theta$, from equation (3.3), we may now write the kinetic energy as

$$K = \frac{1}{2}mv^2 = \frac{1}{2}m\left(\frac{GMm}{L}\right)^2 (1 + e^2 + 2e\cos\theta). \tag{3.58}$$

The kinetic energy is greatest at perihelion ($\theta = 0$), which is as it should be, since that's when the particle is moving fastest. Now using equation (3.34) for r as a function of θ, we can write the potential energy as

$$U = -\frac{GMm}{r} = -\frac{(GM)^2 m^3}{L^2}(1 + e\cos\theta). \tag{3.59}$$

The amplitude of the potential energy, $|U|$, is greatest at perihelion ($\theta = 0$), which is as it should be, since that's when the particle is closest to the mass M. By adding together the kinetic energy (equation 3.58) and the potential energy (equation 3.59), and doing a bit of rearranging, we find

$$E = \left(\frac{GMm}{L}\right)^2 \frac{m}{2}(e^2 - 1). \tag{3.60}$$

This is constant, which is as it should be, since energy is conserved for this isolated two-body system. We can also, if we so choose, write the orbital eccentricity as a function of energy E and angular momentum L:

$$e = \left(1 + \frac{2EL^2}{G^2M^2m^3}\right)^{1/2}. \tag{3.61}$$

We can readily identify three distinct cases:

1. **Hyperbolic orbits:** As we recall from our discussion of conic sections (page 68), the case $e > 1$ represents a hyperbola. Equation (3.60) shows that $e > 1$ corresponds

to a total energy $E > 0$; that is, $K > |U|$. This is an open orbit; the mass m is not gravitationally bound to the mass M. The mass m makes a single perihelion passage at $\theta = 0$ and does not return—its value of r, the distance from the mass M, continues to increase monotonically after perihelion passage.

2. **Parabolic orbits:** In the case where $e = 1$ exactly, the mass m is marginally unbound to M; that is, its velocity approaches zero asymptotically as r approaches infinity. In the case of a parabolic orbit, equation (3.60) shows that $e = 1$ corresponds to $E = 0$, or $K = |U|$. Equation (3.56) reveals that a particle will be on a parabolic orbit if its speed is equal to the **escape speed**:

$$v_{esc}(r) = \left(\frac{2GM}{r}\right)^{1/2}. \qquad (3.62)$$

If its velocity is greater than v_{esc}, it will be on a hyperbolic orbit.

3. **Elliptical orbits:** In the case where $e < 1$, the mass m is gravitationally bound; it goes around the mass M on an elliptical orbit. The total energy, when $e < 1$, is $E < 0$, corresponding to $K < |U|$. The special case $e = 0$ corresponds to a perfectly circular orbit. Equation (3.60) shows that a circular orbit is the orbit that minimizes the energy E for a given angular momentum L.

3.3 ▪ ORBITAL SPEED

It is not possible in general to obtain a simple equation that gives the time dependence of a planet's distance from the Sun, $r(t)$, or orbital speed, $v(t)$.[8] However, it is possible to find the orbital speed v as a simple function of r, which can be useful. We start with the equation for a conic section (equation 3.42), which we write in the form

$$e\cos\theta = \frac{a(1-e^2) - r}{r}. \qquad (3.63)$$

The orbital speed as a function of θ is given by equation (3.58):

$$v^2 = \frac{2K}{m} = \left(\frac{GMm}{L}\right)^2 (1 + e^2 + 2e\cos\theta). \qquad (3.64)$$

Thus, by combining equations (3.63) and (3.64), we find an equation that gives the orbital speed as a function of r:

$$v^2 = \frac{G^2M^2m^2}{L^2}\left(1 + e^2 + \frac{2}{r}[a(1-e^2) - r]\right). \qquad (3.65)$$

Using equation (3.43), which tells us $L^2/m^2 = GMa(1-e^2)$, we find

[8] This also implies that there is no simple equation for $\theta(t)$, since if we had one, we could use the conic section equation for $r(\theta)$ to find $r(t)$.

$$v^2 = \frac{G^2M^2}{GMa(1-e^2)}\left(\frac{r+e^2r+2a(1-e^2)-2r}{r}\right)$$

$$= \frac{GM}{a(1-e^2)}\left(\frac{2a(1-e^2)-r(1-e^2)}{r}\right)$$

$$= \frac{GM}{a}\left(\frac{2a}{r}-1\right) = GM\left(\frac{2}{r}-\frac{1}{a}\right). \quad (3.66)$$

The resulting equation

$$v^2 = GM\left(\frac{2}{r}-\frac{1}{a}\right) \quad (3.67)$$

is called the **vis viva** equation. The Latin term *vis viva*, which translates literally to "living force," is an archaic bit of scientific terminology that was first employed by Gottfried Leibniz (best known as the other discoverer of calculus). Leibniz used the term *vis viva* to refer to the quantity mv^2, what we would now call $2K$, or twice the kinetic energy. The *vis viva* equation is a statement of how the kinetic energy of an orbiting object changes as a function of r. By using Kepler's third law (equation 3.52), we can also write the *vis viva* equation in the form

$$v(r) = \frac{2\pi a}{P}\left(2\frac{a}{r}-1\right)^{1/2}. \quad (3.68)$$

This implies that the orbital angular speed $\omega = v/r$ of a planet is

$$\omega(r) = \frac{2\pi}{P}\frac{a}{r}\left(2\frac{a}{r}-1\right)^{1/2}. \quad (3.69)$$

At perihelion, where $r = q = a(1-e)$, the angular speed of the planet is

$$\omega_{pe} = \frac{2\pi}{P}\frac{(1+e)^{1/2}}{(1-e)^{3/2}}, \quad (3.70)$$

and at aphelion, where $r = Q = a(1+e)$, the angular speed is

$$\omega_{ap} = \frac{2\pi}{P}\frac{(1-e)^{1/2}}{(1+e)^{3/2}}. \quad (3.71)$$

Here on Earth, for instance, the observed average angular speed of the Sun along the ecliptic is equal to 2π radians per sidereal year, or $\omega = 0.986°/\text{day}$. However, since the Earth's orbit has an eccentricity $e = 0.017$, the observed angular speed is greatest at the time of perihelion (early January), when $\omega_{pe} = 1.020°/\text{day}$, and smallest at the time of aphelion (early July), when $\omega_{ap} = 0.953°/\text{day}$.

An interesting application of the *vis viva* equation (eq. 3.68) addresses the problem of the **transfer orbit**. In traveling from the Earth to another planet, the transfer orbit is the route you would take from the Earth to the other planet's orbit. The **Hohmann transfer orbit**, illustrated in Figure 3.10, is an ellipse whose perihelion is at the orbit of

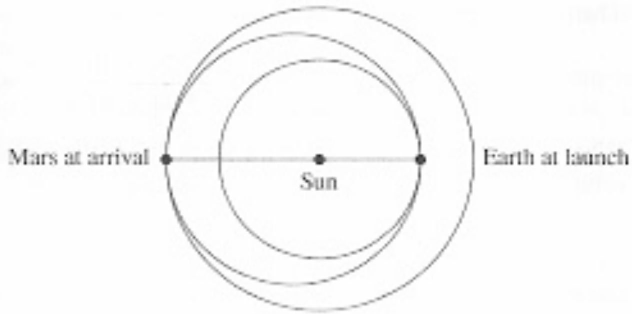


FIGURE 3.10 A Hohmann transfer orbit for interplanetary travel (here from Earth to Mars). The transfer orbit is an ellipse with its perihelion at Earth and its aphelion at the orbit of Mars.

the inner planet and whose aphelion is at the orbit of the outer planet. As the German engineer Walter Hohmann pointed out in the 1920s, the Hohmann transfer orbit has two desirable properties. First, it requires only two engine burns when done properly: one when leaving Earth and one when the destination planet is reached. The rest of the time, the spacecraft is "coasting" on a Newtonian orbit. Second, it is economical in its fuel use; launching your spacecraft on a hyperbolic orbit will cause it to reach its destination faster but requires more energy.

As a concrete example, suppose you want to send a spacecraft to Mars. As a first approximation, we can assume that the orbit of the Earth is a circle of radius $a_\oplus = 1\text{ AU} = 1.50 \times 10^8$ km, with orbital period $P_\oplus = 1\text{ yr} = 3.16 \times 10^7$ s.[9] We further assume that the orbit of Mars is a larger circle, of radius $a_{\rm Mars} = 1.52 a_\oplus = 2.27 \times 10^8$ km, with orbital period $P_{\rm Mars} = 1.88\text{ yr} = 5.94 \times 10^7$ s. The semimajor axis of the Hohmann transfer orbit from Earth to Mars is

$$a_{\rm to} = \frac{a_\oplus + a_{\rm Mars}}{2} = \frac{1\text{ AU} + 1.52\text{ AU}}{2} = 1.26\text{ AU}. \tag{3.72}$$

The orbital period for the transfer orbit is then

$$P_{\rm to}[\text{ yr}] = (a[\text{ AU}])^{3/2} = (1.26)^{3/2} = 1.41. \tag{3.73}$$

Traveling from Earth to Mars requires half an orbit, or a time $t = P_{\rm to}/2 = 0.71\text{ yr} \approx$ 260 days.

The average speed of the Earth on its orbit is

$$v_\oplus = \frac{2\pi a_\oplus}{P_\oplus} = \frac{2\pi(1.50 \times 10^8\text{ km})}{3.16 \times 10^7\text{ s}} = 29.8\text{ km s}^{-1}. \tag{3.74}$$

[9] The "cross in a circle" symbol ⊕ is the standard astronomical symbol for the Earth.

The average speed of Mars is slower:

$$v_{\rm Mars} = \frac{2\pi a_{\rm Mars}}{P_{\rm Mars}} = \frac{2\pi(2.27\times10^8\ {\rm km})}{5.94\times10^7\ {\rm s}} = 24.0\ {\rm km\ s^{-1}}. \tag{3.75}$$

When the spacecraft has just left the Earth, it is at the perihelion of the Hohmann transfer orbit. Its speed, from the *vis viva* equation (eq. 3.68), is

$$\begin{aligned} v_{\rm pe} &= \frac{2\pi a_{\rm to}}{P_{\rm to}}\left(\frac{2a_{\rm to}}{a_\oplus}-1\right)^{1/2} \\ &= \frac{2\pi(1.26\ {\rm AU})(1.50\times10^8\ {\rm km\ AU^{-1}})}{(1.41\ {\rm yr})(3.16\times10^7\ {\rm s\ yr^{-1}})}\left[\frac{2(1.26\ {\rm AU})}{1.00\ {\rm AU}}-1\right]^{1/2} \\ &= 26.7\ {\rm km\ s^{-1}}(1.52)^{1/2} = 32.9\ {\rm km\ s^{-1}}. \end{aligned} \tag{3.76}$$

Thus, at the perihelion of the Hohmann transfer orbit, the spacecraft must be going *faster* than the Earth by an amount $\Delta v = v_{\rm pe} - v_\oplus = 3.1\ {\rm km\ s^{-1}}$. When the spacecraft is just reaching Mars, it is at the aphelion of the Hohmann transfer orbit. Its speed, from equation (3.68), is then

$$v_{\rm ap} = \frac{2\pi a_{\rm to}}{P_{\rm to}}\left(\frac{2a_{\rm to}}{a_{\rm Mars}}-1\right)^{1/2} = 26.7\ {\rm km\ s^{-1}}\left[\frac{2(1.26\ {\rm AU})}{1.52\ {\rm AU}}-1\right]^{1/2} = 21.7\ {\rm km\ s^{-1}}. \tag{3.77}$$

Thus, in order to match its velocity to that of Mars, the spacecraft must increase its speed by $\Delta v = v_{\rm Mars} - v_{\rm ap} = 2.3\ {\rm km\ s^{-1}}$. (If you want your spacecraft to go into orbit around Mars, like the *Mars Reconnaissance Orbiter*, the time, direction, and duration of your engine burn depend on the orbital parameters you want to attain.)

Use of a Hohmann transfer orbit requires careful timing. If you are sending a spacecraft to Mars, for instance, the craft must reach the aphelion of its orbit just as Mars reaches that point. This restricts launches to certain times, known as **launch windows**. If you fail to launch during one launch window, you could wait for one synodic period of the target planet before launching again. For a mission to Mars, whose synodic period is 2.1 years, this could be a frustrating wait.

3.4 ▪ THE VIRIAL THEOREM

If a system contains only two spherical bodies, such as a star and planet, there is a simple analytic solution (first seen in Section 2.5) for the planet's trajectory, $r(\theta)$. Similarly, Section 3.2 yields simple formulas for the planet's kinetic energy $K(\theta)$ and potential energy $U(\theta)$, while Section 3.3 gives the *vis viva* equation for v as a function of r. In a system containing more than two bodies, however, there are no longer any simple analytic solutions for the bodies' properties. Thus, when astronomers study large stellar systems such as star clusters and galaxies, they generally use numerical techniques to compute the stellar orbits using a computer. However, despite the complexity of many-body systems such as star clusters, it is possible to find useful statistical results that describe the average global properties of the system. One such result is the **virial theorem**, which relates the total kinetic energy of a system to its total potential energy.

To derive the virial theorem, let's suppose we have a system containing N stars (or planets, or other compact massive bodies). The mass of the ith star is m_i, and its location is $\vec{\mathbf{r}}_i$. We can define a function

$$A \equiv \sum_{i=1}^{N} m_i \frac{d\vec{\mathbf{r}}_i}{dt} \cdot \vec{\mathbf{r}}_i. \tag{3.78}$$

The reason for defining this function starts to become a bit more obvious when we take the time derivative of A:

$$\frac{dA}{dt} = \sum_{i=1}^{N} \left(m_i \frac{d\vec{\mathbf{r}}_i}{dt} \cdot \frac{d\vec{\mathbf{r}}_i}{dt} + m_i \frac{d^2\vec{\mathbf{r}}_i}{dt^2} \cdot \vec{\mathbf{r}}_i \right). \tag{3.79}$$

The first term on the right-hand side of equation (3.79) is twice the kinetic energy, and the second term can be transformed using Newton's second law,

$$m_i \frac{d^2\vec{\mathbf{r}}_i}{dt^2} = \vec{\mathbf{F}}_i, \tag{3.80}$$

where $\vec{\mathbf{F}}_i$ is the net force acting on the ith star. Thus, we may write

$$\frac{dA}{dt} = 2K + \sum_{i=1}^{N} \vec{\mathbf{F}}_i \cdot \vec{\mathbf{r}}_i, \tag{3.81}$$

where K is the sum of the kinetic energies of all the stars in the system. The term $\sum \vec{\mathbf{F}}_i \cdot \vec{\mathbf{r}}_i$ was named the **virial** by the physicist Rudolf Clausius.[10]

Equation (3.81) is the most general form of the virial theorem. It applies to any system of bodies that follow Newton's second law, regardless of the forces $\vec{\mathbf{F}}_i$ acting on them. A more useful form of the virial theorem can be found by taking the time average of equation (3.81). If we average over the time interval $t = 0 \rightarrow t = \tau$, we find

$$\begin{aligned} 2\langle K \rangle + \langle \sum_{i=1}^{N} \vec{\mathbf{F}}_i \cdot \vec{\mathbf{r}}_i \rangle &= \langle \frac{dA}{dt} \rangle \\ &= \frac{1}{\tau} \int_0^{\tau} \frac{dA}{dt} dt \\ &= \frac{A(\tau) - A(0)}{\tau}. \end{aligned} \tag{3.82}$$

If the system is bound, then the velocity of each particle, as well as its displacement from the origin, remains finite. In that case, $A(t)$, as given by equation (3.78), is finite at all

[10] Clausius also coined the term "entropy," probably his most memorable contribution to the scientific vocabulary.

times, and the right-hand side of equation (3.82) goes to zero in the limit $\tau \to \infty$. Thus, for any bound system of particles, the time-averaged virial theorem has the form

$$2\langle K\rangle + \langle\sum_{i=1}^{N} \vec{\mathbf{F}}_i \cdot \vec{\mathbf{r}}_i\rangle = 0. \tag{3.83}$$

The virial theorem as expressed in equation (3.83) can be applied to any bound system, for instance, to a gas of molecules enclosed within a box. However, as astronomers, we are interested in the specific case of an isolated bound stellar system, in which the force acting on the ith star is the sum of the gravitational forces exerted by the other $N-1$ stars in the system:

$$\vec{\mathbf{F}}_i = \sum_{j\neq i} \frac{Gm_i m_j(\vec{\mathbf{r}}_j - \vec{\mathbf{r}}_i)}{|\vec{\mathbf{r}}_j - \vec{\mathbf{r}}_i|^3}. \tag{3.84}$$

For such a system, what is the value of the virial, $\sum \vec{\mathbf{F}}_i \cdot \vec{\mathbf{r}}_i$? Let's start with a simple system containing only two stars. For this system, the virial will be

$$\begin{aligned}
\vec{\mathbf{F}}_1 \cdot \vec{\mathbf{r}}_1 + \vec{\mathbf{F}}_2 \cdot \vec{\mathbf{r}}_2 &= \frac{Gm_1m_2(\vec{\mathbf{r}}_2 - \vec{\mathbf{r}}_1)\cdot\vec{\mathbf{r}}_1}{|\vec{\mathbf{r}}_2 - \vec{\mathbf{r}}_1|^3} + \frac{Gm_2m_1(\vec{\mathbf{r}}_1 - \vec{\mathbf{r}}_2)\cdot\vec{\mathbf{r}}_2}{|\vec{\mathbf{r}}_1 - \vec{\mathbf{r}}_2|^3} \\
&= -\frac{Gm_1m_2|\vec{\mathbf{r}}_2 - \vec{\mathbf{r}}_1|^2}{|\vec{\mathbf{r}}_2 - \vec{\mathbf{r}}_1|^3} \\
&= -\frac{Gm_1m_2}{|\vec{\mathbf{r}}_2 - \vec{\mathbf{r}}_1|}.
\end{aligned} \tag{3.85}$$

The right-hand side of equation (3.85) is simply the potential energy U of the two-star system. By extension, for a three-star system, the virial will be equal to the sum of the potential energies of all three pairs: (1,2), (2,3), and (3,1). For a system containing N stars, the virial will be equal to the sum of the potential energies of all $N_{\rm pair} = N(N-1)/2$ pairs of stars that can be drawn from the system. We can thus write

$$\sum_{i=1}^{N} \vec{\mathbf{F}}_i \cdot \vec{\mathbf{r}}_i = U = \sum_{i=1}^{N}\sum_{j>i} -\frac{Gm_im_j}{|\vec{\mathbf{r}}_i - \vec{\mathbf{r}}_j|}, \tag{3.86}$$

and the virial equation (eq. 3.83) becomes

$$2\langle K\rangle + \langle U\rangle = 0. \tag{3.87}$$

The virial theorem is useful to astronomers, as we find in Section 20.2, when it enables us to estimate the mass of distant galaxies.