Chapter 1

Celestial Mechanics
and the Solar System

Viewed from the Earth, the Sun slides eastward on its annual journey relative to the stars of the Zodiac while the wandering planets perform periodic gyrations within the Zodiac. (The Zodiac contains the traditional twelve constellations through which the sun travels on its annual journey with respect to the stars as seen from the earth; see Figure 1–1A.) The planets fall into two groups, based on their observed motions. The first, Mercury and Venus, can move eastward of the Sun, each appearing higher and higher in the western sky at sunset as an evening star. At greatest eastern elongation, each planet attains maximum angular distance east of the Sun; it then moves closer to the western horizon at sunset as it sweeps toward the Sun. Moving westward of the Sun, it rises as a predawn morning star in the eastern sky.

The second group visible to the naked eye is Mars, Jupiter, and Saturn. They progress steadily eastward across the sky relative to the stars until they approach opposition—180° away from the Sun in the sky. Then they slow to a halt at a stationary point, trace a retrograde loop toward the west, slow to another stationary point, and finally resume their eastward march (Figure 1–1B).

These planetary phenomena fascinated and frustrated the ancient astronomers. This chapter briefly discusses the historical efforts to explain these planetary motions—efforts that gave birth to physics and astrophysics and culminated in the creation of celestial mechanics by Isaac Newton in the seventeenth century.

1–1 \( \Box \)
THE HISTORICAL BASIS OF SOLAR SYSTEM MODELS

(A) THE HELIOCENTRIC MODEL OF COPERNICUS

In the sixteenth century, the Polish astronomer Nicolaus Copernicus grew dissatisfied with the traditional geocentric (Earth-centered) model of the Solar System and introduced a new heliocentric (Sun-centered) one. This model forms the foundation of the one we use today. Copernicus placed the Sun at the center of the Solar System
FIGURE 1–1 (A) The Sun on the ecliptic with Zodiacal constellations in the background. Note that Mercury, Saturn, and Venus lie close to the ecliptic (but would be invisible, as this view shows daytime). The field of view is 57° by 83° and displays approximately three months worth of the ecliptic. (Image generated by Voyager, the Interactive Desktop Planetarium™)

(B) Retrograde motion of Mars in 1994–95. The planet’s path is indicated relative to the stars of the constellation Leo. This view, which spans 28° by 39°, shows Mars about in the middle of its retrograde loop in February, 1995, when it shines the brightest. The retrograde loop begins in December, 1994 at the first stationary point, and ends in March, 1995, at the second stationary point. (Image generated by Voyager, the Interactive Desktop Planetarium™)
and had the planets (including the Earth) orbit it along circles. This heliocentric model was slow to be accepted because its predictions were not better than those of the geocentric model, but eventually it caught on, primarily for aesthetic reasons of simplicity and harmony.

To understand the explanatory purposes of the Copernican model, you need a little background on naked-eye observations of the Sun, Moon, and planets, which appear geocentric. Daily, the sky turns eastward with respect to the horizon. The Sun appears to move eastward with respect to the stars and circles the sky in a year. The imaginary path traced out by the Sun is called the ecliptic; behind it lies the special set of 12 constellations of the Zodiac. The planets (and Sun and Moon) move with respect to the stars within the band of the Zodiac.

In distance from the Sun in the heliocentric model, the planets range from Mercury (the closest) through Venus, Earth, Mars, Jupiter, Saturn, Uranus, and Neptune to Pluto (the most distant). The planets closer to the Sun than the Earth are termed inferior planets; these are Mercury and Venus. The planets orbiting farther from the Sun than the Earth are called superior planets; these are Mars through Pluto. The motions of the inferior planets as seen in our night sky differ markedly from the motions of the superior planets.

We define elongation (Figure 1–2) as the angle seen at the Earth between the direction to the Sun’s center and the direction to a planet. We speak of eastern or western elongation according to whether the planet lies east or west of the Sun, as seen from the Earth. Elongations of particular geocentric significance are given special names: an elongation of 0° is termed conjunction (inferior conjunction when the planet lies between the Earth and the Sun and superior conjunction when the planet lies on the opposite side of the Sun from the Earth). An elongation of 180° is called opposition, and one of 90° is quadrature. When an inferior planet attains its maximum elongation, we refer to greatest elongation. Only inferior planets may be at inferior conjunction or at greatest elongation (28° for Mercury and 48° for Venus), but they may never be at either quadrature or opposition. Inferior conjunction can never occur for superior planets, and their greatest elongation is 180° (when they are in opposition). Note that our Moon passes through inferior conjunction, quadrature, and opposition (its greatest elongation) because it is a satellite of the Earth.

Copernicus correctly stated that the farther a planet lies from the Sun, the slower it moves around the Sun. When the Earth and another planet pass each other on the same side of the Sun, the apparent retrograde loop occurs (Figure 1–3) from the relative motions of the other planet and the Earth. As we view the planet from the moving Earth, our line of sight reverses its angular motion twice, and the three-dimensional aspect of the loop comes about because the orbits of the two planets are not coplanar. This passing situation is the same for inferior or superior planets.

Copernicus derived a relationship between the synodic and sidereal periods of a planet in the heliocentric model. The synodic period S is the time it takes the planet to return to the same position in the sky relative to the Sun, as seen from the Earth.

**FIGURE 1–2** Heliocentric planetary configurations. Arrows indicate the direction of orbital motion as well as the rotational direction of the Earth. The phases of the planets’ illumination are also shown.
For the inferior planets, Mercury and Venus, this time is the interval between successive inferior conjunctions; for the superior planets, it is the interval between successive oppositions. The **sidereal period** $P$ is the time it takes the planet to complete one orbit of the Sun with respect to the stars (Figure 1–4). The Earth's sidereal period $E$ is 365.26 days. The Earth moves at the rate of $360^\circ/E$ degrees per day in its orbit, while a planet's rate of angular motion is $360^\circ/P$ as viewed from the Sun. For a superior planet, the Earth completes one orbit and must then traverse the additional angle $S \times (360^\circ/P)$ in the time $S - E$ to catch up to the superior planet at opposition again. Hence,

\[(S - E)(360^\circ/E) = S(360^\circ/P)\]

or

\[1/S = 1/E - 1/P\]

For an inferior planet, the Earth is a superior planet, and so we interchange $E$ and $P$ to arrive at Copernicus' result.

\[1/S = 1/P - 1/E \text{ (inferior)}\]
\[1/S = 1/E - 1/P \text{ (superior)}\]
As an example, consider Venus, an inferior planet with an observed synodic period of \( S = 583.92 \) days. The appropriate relationship is

\[
\frac{1}{583.92} = \frac{1}{P} - \frac{1}{365.26} \\
\frac{1}{P} = 0.00171 + 0.00274 = 0.00445 \\
P = 224.7 \text{ days}
\]

The telescopic observations reported by Galileo Galilei in his *Siderius Nuncius* in 1610 strongly supported the heliocentric model of the Solar System. His drawings of the wrinkled lunar surface and the moving sunspots on the Sun weakened the ancient belief in perfect and immutable heavens. Galileo also discovered the four largest moons of Jupiter and showed that they orbited Jupiter, not the Earth. This crack in the wall of geocentricism became an irreparable breach with his discovery of the phases of Venus.

Here’s why. The planets and the Moon shine from the sunlight they reflect. Half of a planet is always sunlit while the other half is dark. The fraction of the sunlit hemisphere seen from the Earth, however, varies with the configuration. So the phase termed **new** occurs when we see only the dark hemisphere (at inferior conjunction for the Moon, Mercury, and Venus), and **full** phase takes place at opposition when the entire sunlit hemisphere faces us. The superior planets can never be in **crescent** phase (when less than half the observable hemisphere is sunlit) and are almost always in **gibbous** phase (when more than half the planet appears sunlit). Galileo observed that Venus shows all phases—hence it must orbit the Sun. This observation confirmed Copernicus’ model.

**(B) THE METHODS OF KEPLER**

Using the heliocentric model of Copernicus and the positional observations of Mars that Tycho Brahe had painstakingly accumulated over 20 years, Johannes Kepler discovered the need for elliptical planetary orbits (Section 1–2). In 1609 and 1619, he published his three empirical laws of planetary motion. These set the stage for Newton’s great scheme of gravitation.

Let’s examine Kepler’s methods for determining distance to the planets. We use the mean Sun–Earth distance—called the **astronomical unit** (AU)—as the unit of distance.

The Sun–planet distance \( r \) (in AUs) may be found when an inferior planet reaches greatest elongation (Figure 1–5A). The angle \( SEP \) is observed (call it \( \alpha \)) and the angle \( EPS \) is 90°; hence, trigonometry yields \( r = \sin \alpha \). (This method was first worked out by Copernicus.)

Kepler’s method for finding the distance to a superior planet is more complicated than the preceding procedure (Figure 1–5B). The planet is at \( P \) at the beginning and end of one sidereal period, and the Earth is at \( E \) and \( E' \) at these two times. Note that point \( P \) is on the planet’s orbit but is otherwise arbitrary. Because we know the planet’s sidereal period, we also know the angle \( ESE' \); we must observe the angles \( PES \) and \( PE'S \). We can solve the triangle \( ESE' \) using the law of cosines and trigonometry (see Appendix 9, “Mathematical Operations”) to obtain \( EE' \) and the angles \( SEE' \) and \( SE'E \). By subtraction, the angles \( PE'E \) and \( PE'S \) are known, and we may solve the triangle \( EPE' \). Enough information is now available to solve for \( r \), using either triangle \( SEP \) or triangle

---

**FIGURE 1–5** Distance determinations in a heliocentric model.

(A) When an inferior planet reaches greatest elongation (\( P \)), we know angle \( SEP \) and can find \( r \) because angle \( SPE \) is a right angle. (B) A superior planet is at \( P \) at the start and end of one sidereal period; at these times, the Earth is at \( E \) and \( E' \). Angles \( PES \) and \( PE'S \) are observed; they are the elongations of the planet from the Sun.
SEP. This process was the method by which Kepler first traced out the orbit of Mars to find that it is elliptical—a significant break with the astronomical tradition of circular orbits.

1–2 PLANETARY ORBITS

(A) KEPLER’S THREE EMPIRICAL LAWS

Kepler tested many models of orbital shapes—even ovals—but discarded them all. He finally showed that the orbital planes of the planets pass through the Sun and discovered that the orbital shape was an ellipse [Section 1–2(b)]. These findings were announced in 1609 as Kepler’s first law—the law of ellipses: the orbit of each planet is an ellipse with the Sun at one focus (Figure 1–6A).

Kepler also investigated the speeds of the planets and found that the closer in its orbit a planet is to the Sun, the faster it moves. Drawing a straight line connecting the Sun and the planet (the radius vector), he discovered that he could express this fact in Kepler’s second law—the law of areas: the radius vector to a planet sweeps out equal areas in equal intervals of time (Figure 1–6B).

Ever seeking a greater harmony in the motions of the planets, Kepler toiled for another decade and in 1619 put forth his third law—the harmonic law: the squares of the sidereal periods of the planets are

FIGURE 1–6 Kepler’s laws of planetary motion. (A) Each planet (P) traces an elliptical orbit around the Sun (S), which is at one focus (F) of the ellipse. (B) Consider two equal time intervals, that from 1 to 2 and that from 3 to 4. The radius vector to the planet (SP) sweeps out the same area (A) during these times. (C) For all major planets, this log-log plot of semimajor axes (a) versus sidereal periods (P) falls very close to a straight line of slope 2/3, confirming Kepler’s third law.
proportional to the cubes of the semimajor axes (mean radii) of their orbits (Figure 1–6C).

The third law may be written algebraically as

$$p^2 = ka^3$$

where $P$ is a planet’s sidereal period and $a$ is its average distance from the Sun (the semimajor axis of an elliptical orbit; see below); the constant $k$ has the same value for every body orbiting the Sun. By 1621, Kepler had shown that the four moons of Jupiter discovered by Galileo obeyed the third law (with a different value of $k$), confirming its wide applicability.

(B) GEOMETRIC PROPERTIES OF ELLIPTICAL ORBITS

An ellipse is defined mathematically as the locus of all points such that the sum of the distances from two foci to any point on the ellipse is a constant (Figure 1–7); hence

$$r + r' = 2a = \text{constant} \quad (1-1)$$

The line joining the two foci $F$ and $F'$ intersects the ellipse at the two vertices $A$ and $A'$. Note that $a$ is half the distance between the vertices; $a$ defines the semimajor axis of the ellipse. The shape of the ellipse is determined by its eccentricity $e$, such that the distance from each focus to the center of the ellipse is $ae$. When $e = 0$, we have a circle; for $e = 1$, a straight line. One-half of the perpendicular bisector of the major axis is the semiminor axis $b$. Using the dashed lines ($r = r' = a$) in Figure 1–7 and the Pythagorean theorem, we find

$$b^2 = a^2 - a^2 e^2 = a^2(1 - e^2) \quad (1-2)$$

Kepler’s first law places the Sun at one focus, $F$. Then vertex $A$ is termed the perihelion of the orbit (point nearest the Sun), and vertex $A'$ is called the aphelion (point farthest from the Sun). The perihelion distance $AF$ is $a(1 - e)$, and the aphelion distance $A'F$ is $a(1 + e)$. The mean (average) distance from the Sun to a planet in elliptical orbit is just the semimajor axis $a$. We prove this fact by noting that for each point $P$ on the ellipse at a distance $r$ from focus $F$, there is a symmetrical point $P'$ a distance $r'$ from $F$; the average of these distances is $(r + r')/2 = a$. This result holds for any arbitrary but symmetrical pair of points.

We need to know the distance from one focus to a point on the ellipse (such as the Sun–planet or planet–satellite distance) as a function of the position of that point. Center a polar coordinate system $(r, \theta)$ at $F$ and let the line $FA$ correspond to $\theta = 0$. Now $r$ measures the distance $FP$, and then $\theta$—the true anomaly—measures the counterclockwise angle $AFP$. Using

$$\cos (\pi - \theta) = -\cos \theta$$

and the law of cosines, we have

$$r'^2 = r^2 + (2ae)^2 + 2r(2ae) \cos \theta$$

From Equation 1–1, however, $r' = 2a - r$, and so

$$r = a(1 - e^2)/(1 + e \cos \theta) \quad (1-3)$$

Equation 1–3 is the equation for an ellipse in polar coordinates for $0 \leq e < 1$.

To derive the area of an ellipse, we find the analog of Equation 1–3 in Cartesian coordinates $(x, y)$ positioned at the center of the ellipse. Then Figure 1–7 and the Pythagorean theorem give

$$r'^2 = (x + ae)^2 + y^2$$

$$r^2 = (x - ae)^2 + y^2$$

Subtracting these two equations and using Equation 1–1, we find $r' = a + ex$. Substituting back into the first of the above two equations and employing Equation 1–2, we obtain

$$(x/a)^2 + (y/b)^2 = 1 \quad (1-4)$$
which is the equation for an ellipse in Cartesian coordinates. The area of the ellipse is given by the double integral

\[ A = 4 \int_0^b dy \int_0^x dx \]

where, from Equation 1–4,

\[ x = a[1 - (y/b)^2]^{1/2} \]

The integration is easy if we use the substitution \( y = b \sin z \) (so that \( dy = b \cos z \, dz \)) and the relationship \( \sin^2 z + \cos^2 z = 1 \); the final answer is

\[ A = \pi ab \]  \hspace{1cm} (1–5) \]

The ellipse is one example of a class of curves called conic sections. This family of curves, all of which result from slicing a cone at different angles with a plane, includes the circle, ellipse, parabola, and hyperbola (Figure 1–8). From Equation 1–3, note that the ellipse degenerates to a circle of radius \( r = a \) when \( e = 0 \). If we increase \( e \), the foci move apart. When \( e = 1 \), one of the foci is at infinity and we have the parabola specified by

\[ r = 2p/(1 + \cos \theta) \]  \hspace{1cm} (1–6) \]

where \( p \) is the distance of closest approach (at \( \theta = 0 \)) to the remaining focus. When the eccentricity is greater than unity, the open hyperbola results:

\[ r = a(e^2 - 1)/(1 + e \cos \theta) \]  \hspace{1cm} (1–7) \]

Its distance of nearest approach to the sole focus is \( a(e - 1) \).

When one body moves under the gravitational influence of another, the relative orbit of the moving body must be a conic section. (The relative orbit is that seen from an observer on the more massive body.) Planets, satellites, and asteroids have elliptical orbits; many comets have eccentricities so close to unity that they follow essentially parabolic orbits. A few comets have hyperbolic orbits; after one perihelion passage, such comets leave the Solar System forever. Space probes have been launched into hyperbolic orbits with respect to the Earth, but they are nearly always captured into elliptical orbits about the Sun. Pioneer 10 was

![Conic sections diagram](image)

**FIGURE 1–8** Conic sections. (A) The family of conic-section curves includes the circle \((e = 0)\), the ellipse \((0 < e < 1)\), the parabola \((e = 1)\), and the hyperbola \((e > 1)\). (B) Conic sections are formed when a cone is cut with a plane. When the plane is perpendicular to the cone's axis, the result is a circle; when it is parallel to one side, the result is a parabola; intermediate angles result in ellipses. A hyperbola results when the angle the plane makes with the cone's side is greater than the opening angle of the cone.
the first spacecraft that, when perturbed by Jupiter, escaped from the Solar System.

1–3 \( \Box \)

**NEWTON’S MECHANICS**

Using Kepler’s empirical deductions about planetary orbits, Sir Isaac Newton created his unified scheme of dynamics and gravitation, which he published in his *Principia* in 1687. Newton’s brilliant insight and elegant formulation laid the foundations for the Newtonian physics that we know today. This section presents the first half of his unified structure—mechanics.

Newton assumed that the arena within which motions take place is three-dimensional, Euclidean space. These motions occur in time, which passes steadily and is unaffected by any phenomenon in the Universe. The basic entity in the scheme is the point particle, which has mass but no extent (Figure 1–9A). The position of the particle at time \( t \), relative to some origin, is indicated by a vector \( \mathbf{x}(t) \); the length of this vector is measured in units such as meters. At a slightly later time \( t + \Delta t \), the particle has moved to \( \mathbf{x} + \Delta \mathbf{x} \) at approximately the velocity

\[
v = \frac{[x(t + \Delta t) - x]}{t}\text{ or }\frac{[x(t + \Delta t) - x]}{t} = \Delta v/\Delta t
\]

As we let \( \Delta t \to 0 \), the velocity vector becomes tangent to the trajectory at point \( x \), where it is defined by the derivative

\[
v = \frac{d\mathbf{x}}{dt}
\]  

(1–8)

The magnitude of the velocity vector is called the *speed*, and its units are distance/time, such as meters per second (m/s) or kilometers per second (km/s). Noting that the velocity of the particle at \( t \) is \( v \), while at \( t + \Delta t \) it is \( v + \Delta v \), we may express the change in velocity by the acceleration vector (Figure 1–9B):

\[
a = \frac{[v + \Delta v] - v}{[t + \Delta t] - t} = \Delta v/\Delta t
\]

\[
a = \frac{dv}{dt} \quad \text{ (as } t \to 0 \text{)}
\]

(1–9)

The units of acceleration are speed/time, which is distance/time\(^2\), such as meters/second\(^2\) or kilometers/second\(^2\). Newton also considered the linear **momentum** vector (units of mass times speed—for example, kg \( \cdot \) m/s) of the particle, defined by the product

\[
p = m \mathbf{v}
\]

(1–10)

where \( m \) is the particle’s mass and \( \mathbf{v} \) is its instantaneous velocity. With these kinematic fundamentals in mind, let’s turn to Newton’s laws of motion.

**(A) THE LAW OF INERTIA**

In his *Physics*, Aristotle attempted to show that the natural state of a body is one of rest. Our daily experience seems to verify his observation that all moving objects eventually slow to a halt. Indeed, to explain the flight of an arrow, Aristotle said that the diverted air rushing in upon the tail of the arrow “pushed” it along.

Galileo came to a quite different conclusion. He released balls so that they rolled down smooth inclined planes and observed that they rolled up adjacent inclined planes to approximately the same height as that from which he had released them. As he made the planes smoother and inclined the second plane less to the horizontal, he found that the balls rolled farther. Galileo attributed any slowing down of the balls to friction and conjectured

---

**FIGURE 1–9** Motion of a particle. (A) A mass \( m \) at time \( t \) is at position \( x \) in its trajectory. At time \( t + \Delta t \), it is at position \( x + \Delta x \). The instantaneous velocity at \( x + \Delta x \) is \( v + \Delta v \). (B) The change in velocity between \( t \) and \( \Delta t \) is \( \Delta v \), from which instantaneous acceleration \( a \) is defined.
that a smooth ball on a horizontal plane would roll forever at a constant speed.

René Descartes later formulated this principle in the form Newton adopted as his first law of motion, the law of inertia: the velocity of a body remains constant (in both magnitude and direction) unless a force acts upon the body. For a freely moving body, the first law may be written \( v = \text{constant} \); when the constant is zero, a body initially at rest will remain at rest unless acted upon by a force.

The modern form of Newton’s first law is known as the law of conservation of linear momentum. For a body of mass \( m \), we may write \( \mathbf{p} = mv = \text{constant} \), which is equivalent to

\[
\frac{dp}{dt} = 0 \quad \text{(force-free)} \quad (1-11)
\]

This relation is true even when a body’s mass changes, as in the case of a rocket ship. In the case of a constant mass, then,

\[
m \frac{dv}{dt} = 0
\]

and

\[
m a = 0 \quad \text{(force-free, constant mass)}
\]

(B) THE DEFINITION OF FORCE

Newton’s second law is already implied by the proviso “unless a force acts upon the body” in the first law, for a force causes a change in velocity. Such a change in velocity (in speed or direction or both) is indicated by the acceleration vector (Equation 1–9). An important special case of accelerated motion is circular motion, in which the speed remains constant while the direction of motion changes.

The concept of force was defined in Newton’s second law of motion (the force law): the acceleration imparted to a body is proportional to and in the direction of the force applied and inversely proportional to the mass of the body. So we may write \( \mathbf{a} = F/m \) or, more commonly,

\[
F = ma \quad (1-12)
\]

Note that force is a vector, with the units mass times acceleration, such as kg \( \cdot \) m/s\(^2\). If several forces act upon a single body, the resultant acceleration is determined by Equation 1–12, using the force that is the vector sum of the individual forces—this is the principle of superposition. Two forces \( F_1 \) and \( F_2 \) add to give the resultant force \( \mathbf{F} \) (Figure 1–10); we may reverse this procedure to decompose the force \( F \) into any two or more component forces; two are indicated here by \( F_x \) and \( F_y \).

In Equation 1–12, the body’s mass \( m \) must remain constant. This restriction vanishes in the modern statement of the second law, which is formulated using the body’s linear momentum:

\[
F = \frac{dp}{dt} \quad (1-13)
\]

We may recover Newton’s form of the second law by using Equations 1–9, 1–10, and 1–13 when \( m \) is constant. Note that the first law of motion (Equation 1–11) now results from the second.

What is the meaning here of mass? In dynamics, mass represents the inertia of a body, that is, that body’s resistance to any change in its state of motion. If we apply the same force to two bodies, the more massive body will change velocity at a lower rate than the less massive body. In everyday terms, think of mass as the amount of material comprising a body. Mass is a scalar quantity characterizing a body, and it does not depend on the body’s location or state of motion [Section 1–4(b)].

(C) ACTION AND REACTION

To complete his dynamic theory, Newton developed his third law of motion, the law of action–reaction: for every force acting on a body (in a closed system), there is an equal and opposite force exerted by

![Figure 1-10](https://example.com/figure110.png)

**FIGURE 1-10** Superposition of forces. Two forces \( F_1 \) and \( F_2 \) act on a body. The resulting motion depends on \( F \), the vector sum of \( F_1 \) and \( F_2 \). The orthogonal components \( F_x \) and \( F_y \) yield the same motion because their sum results in \( F \).
that body. A simple example: The weight (a force) of a book lying on a table must be exactly balanced by the force that the table exerts on the book; otherwise, according to the second law, the book would accelerate off or through the table. The third law describes the static situation of balanced forces.

The modern version of the third law is the law of conservation of total linear momentum. Though we may treat any number of bodies, consider only two. The total linear momentum of the system is given by \( P = p_1 + p_2 = \) constant, when no external force acts upon the system (the first two laws). Considering two instants of time (the latter instant indicated by primes), we have

\[
p_1 + p_2 = p_1' + p_2'
\]

If the time interval is \( \Delta t \) and if we call \( \Delta p_1 = p_1' - p_1 \) and \( \Delta p_2 = p_2' - p_2 \), we can arrange the above equation and divide by \( \Delta t \) to get

\[
\frac{\Delta p_1}{\Delta t} = -\frac{\Delta p_2}{\Delta t}
\]

For arbitrarily small \( \Delta t \), the deltas become differentials \( (\Delta t \to dt) \) and Equation 1–13 yields the third law:

\[
F_1 = -F_2
\]

(D) SUMMARY: NEWTON’S LAWS OF MOTION

We present here the modern version of Newton’s three laws of mechanics. Recall that the Newtonian context is absolute space and time, wherein a point particle of mass \( m \) describes a trajectory \( x(t) \) with instantaneous velocity \( v(t) \), linear momentum \( p = mv \), and acceleration \( a(t) \).

First law (inertia):
\( v \) and \( p = \) constant

Second law (force):
\( F = dp/dt \)

Third law (action–reaction)
\[
F_{exerted} = -F_{acting}
\]

\[
P = \) constant
\]

In a closed system, the force exerted by a body is equal and opposite to the force acting upon the body, or, the total linear momentum of a closed system of bodies is constant in time.

1–4  0

NEWTON’S LAW OF UNIVERSAL GRAVITATION

Before Newton, Kepler suspected that some force acted to keep the planets in orbit about the Sun; he attributed the elliptical orbits to the force of magnetic attraction. Following a train of thought similar to the simplified derivation that follows, Newton discovered his law of universal gravitation, tested it on the motion of the Moon, and then explained the motions of the planets in detail.

(A) CENTRIPETAL FORCE AND GRAVITATION

Consider a body moving in a circular orbit of radius \( r \) about a center of force (Figure 1–11A). From symmetry, the speed \( v \) of the body must be constant, but the direction of the velocity vector is constantly changing. Such a changing velocity represents an acceleration—the centripetal acceleration that maintains the circular orbit; from the geometry of the figure, we deduce the acceleration. The time is \( t \) at point A, where the body’s velocity is \( v \). At an infinitesimal time interval \( \Delta t \) later, the body has traversed the angle \( \Delta \theta \) to B, where the velocity is \( v' \). In Figure 1–11B, we show the change in velocity \( \Delta v = v' - v \) by joining the tails of the two velocity vectors; the angle between \( v \) and \( v' \) is \( \Delta \theta \). Recalling that the magnitude of both \( v \) and \( v' \) is the speed \( v \), we use trigonometry to deduce, for small values of \( \Delta \theta \) (Figure 1–11A),

\[
\Delta \theta = \frac{\Delta \theta}{r}
\]

and (Figure 1–11B)

\[
\Delta \theta = \Delta \nu / v
\]

where the arc length \( s \) approximates the chord joining points A and B. So the centripetal acceleration has the magnitude
which is Newton’s law of universal gravitation; the direction of the gravitational force is along the line joining the two bodies (from the third law of motion). The gravitation constant $G$ has the measured value $6.67 \times 10^{-11} \text{ m}^3/\text{kg} \cdot \text{s}^2$ in SI units.

Equation 1–15 expresses the attractive gravitational force between two point masses. To find the gravitational attraction from an extended body, we must sum the vectorial contributions from each small piece of the body. In general, this process is rather difficult, but for a spherically symmetric body acting on a point mass, symmetry arguments alone tell us that the gravitational force must act along the line joining the centers of the bodies. By integrating the effects of all parts of the spherical body, we find that it behaves gravitationally as though its entire mass were concentrated at its center. This important result will be used many times in this book.

### (B) WEIGHT AND GRAVITATIONAL ACCELERATION

Near the Earth’s surface, bodies have a constant downward gravitational acceleration of magnitude (from Equation 1–15)

$$g = GM_\oplus/R_\oplus^2 \approx 9.8 \text{ m/s}^2$$

where $M_\oplus$ is the mass of the Earth and $R_\oplus$ is its radius. The Earth’s oblate surface and rapid rotation cause variations in the measured value of $g$ from 9.781 m/s² at the equator to 9.832 m/s² at the poles.

The **weight** of a body is the force on it in a gravitational field. At the Earth’s surface, we weigh a body on a scale, and if the mass of the body is $m$, we find

$$\text{Weight} = mg$$

In contrast to its mass, the weight of a body depends upon its **location**. An astronaut on the Moon’s surface will weigh approximately one-sixth her normal Earth weight; in a satellite orbit, her weight will be zero since she is falling freely in the gravitational field. In both cases, her mass is the same. Weight and force have the same units,
and the conventional SI unit is the newton (Appendix 6), where

\[ 1 \text{ newton} = 1 \text{ kg} \cdot \text{m/s}^2 \]

Keep in mind that weight is a force.

**(C) DETERMINATION OF G AND M_\oplus**

Let's describe how Henry Cavendish measured the gravitational constant \( G \) in 1798 and how Philip von Jolly established the mass of the Earth \( M_\oplus \) in 1881. These two methods represent early procedures to determine these important constants.

To find \( G \), Cavendish used an apparatus consisting of two small balls of equal mass \( m \) hung from a torsion beam and two larger balls of equal mass \( M \) attached to an independently suspended but coaxially aligned beam (Figure 1–12A). Each adjacent \( M - m \) pair is initially placed a distance \( D \) apart, but the gravitational force between the balls twists the torsion bars to a static equilibrium distance \( d \) between each \( M - m \) pair. From symmetry, the gravitational force causing the deflection is

\[ F_{\text{tot}} = 2GMm/d^2; \]

by directly measuring \( F_{\text{tot}}, M, m, \) and \( d \), Cavendish found \( G \).

Von Jolly's apparatus (Figure 1–12B) consisted of a balance bearing two small masses \( m \), with the rotation axis of the balance beam aligned horizontally. When he placed a large mass \( M \) below one of these small masses, the balance tipped; to restore the original balance, he placed a small mass \( n \) on the pan alongside the other mass \( m \). If the equilibrium \( M - m \) distance is \( d \), the forces acting on each side of the torsion beam are

\[ F_1 = GMm/d^2 + GM_\oplus m/R_\oplus^2 \]
\[ F_2 = GM_\oplus n/R_\oplus^2 + GM_\oplus m/R_\oplus^2 \]

The beam is horizontal again when \( F_1 = F_2 \), and so we find

\[ M_\oplus = (Mm/n)(R_\oplus/d)^2 = 5.976 \times 10^{24} \text{ kg} \]

**1–5 PHYSICAL INTERPRETATIONS OF KEPLER'S LAWS**

Newton combined his laws of motion and gravitation to derive all three of Kepler's empirical laws.

---

**FIGURE 1–12** Measuring Newton's constant of gravitation, \( G \). (A) The Cavendish experiment uses two torsion bars \( T \) and \( T' \), which are free to rotate about a vertical axis. After the gravitational force between adjacent \( M - m \) spheres swings the bars to their equilibrium positions (to position \( f \) from position \( i \)), the separations between \( M \) and \( m \) are \( d \). (B) Von Jolly's experiment has a horizontal balance that is in an equilibrium horizontal position when masses \( M \) and \( n \) are absent. When mass \( M \) is placed below the left-hand mass \( m \), mass \( n \) must be added to the right-hand mass \( m \) to restore equilibrium.
We could derive the elliptical orbits of Equation 1–3, using Equations 1–13 and 1–15; this requires a knowledge of vector differential equations, so we won’t do so. Newton’s form of Kepler’s laws conceptually views orbits as conic sections (Figure 1–8) with the center of mass of the system at one focus. We will accept Kepler’s first law and follow Newton’s footsteps in deducing Kepler’s second and third laws.

(A) THE LAW OF AREAS AND ANGULAR MOMENTUM

Let’s illustrate Kepler’s law of areas for an elliptical orbit (Figure 1–13). A body orbits the focus $F$ at the position $r$ and velocity $v$. During an infinitesimal time interval $\Delta t$, the body moves from $P$ to $Q$ and the radius vector sweeps through the angle $\Delta \theta$. This small angle is $\Delta \theta = v_t \Delta t/r$, where $v_t$ is the component of $v$ perpendicular to $r$. During this time, the radius vector has swept out the triangle $FPQ$, the area of which is $\Delta A = rv_t \Delta t/2$. Therefore, as $t \to 0$,

$$\frac{dA}{dt} = rv_t/2 = r^2(d\theta/dt)/2 = \frac{H}{2} \quad (1-16)$$

where the constant $H$ (the angular momentum per unit mass) appears because Kepler’s second law states that the rate of change of area with time is a constant.

Note that $A/P = H/2$, where $A = \pi ab$ is the total area of the ellipse and $P$ is the orbital period; this comes by integrating Equation 1–16. By combining this result with Equation 1–16 and noting that $v_t$ is the total speed at perihelion or aphelion, we deduce the perihelion and aphelion speeds of a

planet orbiting the Sun. For example, at perihelion we have

$$v = H/r = 2A/Pr = 2\pi a b P a (1 - e)$$

where Equation 1–3 with $\theta = 0$ is used in the last equality. Carrying through the derivation at aphelion and using Equation 1–2, we get

$$v_{per} = (2\pi a/\rho)[(1 + e)/(1 - e)]^{1/2} \quad (1-17a)$$

$$v_{ap} = (2\pi a/\rho)[(1 - e)/(1 + e)]^{1/2} \quad (1-17b)$$

For the Earth, $a$ is 1 AU (1.496 × 108 km), $P$ is one year (3.156 × 107 s), and the orbital eccentricity is $e = 0.0167$; hence, the orbital speed varies from 30.3 km/s at perihelion to 29.3 km/s at aphelion.

A modern Newtonian derivation of Kepler’s second law requires the concept of the orbiting body’s angular momentum:

$$L = r \times p = m (r \times v) \quad (1-18)$$

where $m$ is the body’s mass, $r$ its position vector, and $p$ its linear momentum (Equation 1–10). The vector cross product (denoted by $\times$) in Equation 1–18 is an operation that yields the product of the perpendicular components of two vectors (see Appendix 9, ‘Mathematical Operations’); hence, if $r$ and $p$ are parallel, then $r \times p = 0$. Angular momentum is a vector quantity $L$ with the units kg·m²/s. Differentiating Equation 1–18, we have

$$\frac{dL}{dt} = v \times p + r \times (dp/dt) = r \times F \quad (1-19)$$

since $v$ is parallel to $p$ and $dp/dt$ defines force. We call $dL/dt$ the torque (with units kg·m²/s²) and see that when $F$ is collinear with $r$—a central force, such as gravitation—the torque vanishes. Hence, $L$ is constant in time so that angular momentum is conserved for all central forces. Applying Equation 1–18 to the situation in Figure 1–13, we find

$$L/m = r \times v_t = H = \text{constant}$$

which is Kepler’s second law.

(B) NEWTON’S FORM OF KEPLER’S THIRD LAW

The external forces that act upon the Solar System are essentially negligible; hence, the total linear momentum of the Solar System is constant. So the Sun must move about the center of mass of the Solar System. We apply this idea to an isolated sys-
tem of two bodies moving in circular orbits from their mutual gravitational attraction; our final result, Newton’s form of Kepler’s third law, is also applicable to elliptical orbits.

Consider two bodies of masses $m_1$ and $m_2$, orbiting their stationary center of mass at distances $r_1$ and $r_2$ (Figure 1–14). Because the gravitational force acts only along the line joining the centers of the bodies, both bodies must complete one orbit in the same period $P$ (though they move at different speeds $v_1$ and $v_2$). The centripetal forces of the orbits are therefore

$$F_1 = m_1v_1^2/r_1 = 4\pi^2m_1r_1/P^2 \quad (1–20a)$$
$$F_2 = m_2v_2^2/r_2 = 4\pi^2m_2r_2/P^2 \quad (1–20b)$$

Newton’s third law requires $F_1 = F_2$, and so we obtain

$$r_1/r_2 = m_2/m_1 \quad (1–21)$$

The more massive body orbits closer to the center of mass than does the less massive one; Equation 1–21 defines the position of the center of mass.

The total separation of the two bodies $a = r_1 + r_2$ is also the radius of their relative orbits. Now Equation 1–21 may be expressed in the form

$$r_1 = m_2a/(m_1 + m_2) \quad (1–22)$$

However, the mutual gravitational force $F_{grav} = F_1 = F_2$ is

$$F_{grav} = Gm_1m_2a^2 \quad (1–23)$$

so that combining Equations 1–20a, 1–22, and 1–23 gives Newton’s form of Kepler’s third law:

$$p^2 = 4\pi^2a^3/G(m_1 + m_2) \quad (1–24)$$

If body 1 is the Sun and body 2 any planet, then $m_1 \gg m_2$; therefore $k = 4\pi^2/Gm_1$ is the proportionality “constant” (to a good approximation) in Kepler’s third law. Note that the third law presents a way to describe gravitational effects without specifically discussing forces.

(C) ORBITAL VELOCITY

To better understand elliptical orbits, consider the orbital velocity $v$. We may decompose this velocity into two perpendicular components (Figure 1–15): $v_r$, the radial speed, and $v_\theta$, the angular speed. Now, from Equation 1–16 and the discussion following it, we have

$$d\theta/dt = (2\pi a/P)(a/r)^2(1 - e^2)^{1/2} \quad (1–25)$$

Using Equation 1–3, which is the polar equation of an ellipse, and Equation 1–25, we compute the time derivatives below to find:

$$v_r = dr/dt = (2\pi a/P)(e \sin \theta)(1 - e^2)^{-1/2} \quad (1–26a)$$
$$v_\theta = r(d\theta/dt) = (2\pi a/P)(1 + e \cos \theta)(1 - e^2)^{-1/2} \quad (1–26b)$$

Note that Equation 1–26b reduces to Equation 1–17 at perihelion and aphelion. The total orbital speed now follows from Equations 1–26 as

$$v^2 = v_r^2 + v_\theta^2 = (2\pi a/P)^2(1 + 2e \cos \theta + e^2)/(1 - e^2) \quad (1–27)$$

Rearranging the polar equation for an ellipse, we get

$$e \cos \theta = \sqrt{a(1 - e^2)} - r/r$$

and substituting this result into Equation 1–27, we finally obtain (with the help of Equation 1–24)

$$v^2 = G(m_1 + m_2)((2/r) - (1/a)) \quad (1–28)$$

Therefore, for given masses, the total orbital speed depends only upon the separation and the orbit’s semimajor axis. This useful result is commonly called the vis viva equation.

(D) THE CONSERVATION OF TOTAL ENERGY

The concept of energy provides an alternative approach to that of forces and Newtonian mechanics.
The vector dot product operation in Equation 1–29 yields the product of the parallel components of $\mathbf{F}$ and $d\mathbf{x}$ (Appendix 9, "Mathematical Operations"); when $\mathbf{F}$ and $d\mathbf{x}$ are mutually perpendicular, their dot product vanishes. To evaluate Equation 1–29, we note:

$$\mathbf{F} \cdot d\mathbf{x} = m(d\mathbf{v}/dt) \cdot \mathbf{v} \cdot dt = m(\mathbf{v} \cdot d\mathbf{v}) = d(mv^2/2)$$

where we have used Newton's second law and the definitions of velocity and speed. We integrate Equation 1–29 directly to give

$$W = (mv^2/2)_B - (mv^2/2)_A = KE_B - KE_A \quad (1–30)$$

where the kinetic energy is specified by $KE = mv^2/2$. Therefore, the work done by the force on the body changes the body's kinetic energy. In a system of two bodies of masses $m_1$ and $m_2$, the kinetic energy is $KE = (m_1v_1^2/2) + (m_2v_2^2/2)$. If the force is the gravitational force between the two bodies, we have

$$\mathbf{F} \cdot d\mathbf{x} = -Gm_1m_2/r^2 \, dr = d(Gm_1m_2/r)$$

where $r$ is the radial separation of the bodies. Equation 1–29 is again integrated:

$$W = (Gm_1m_2/r)_B - (Gm_1m_2/r)_A = PE_A - PE_B \quad (1–31)$$

where we have inserted the definition of the mutual gravitational potential energy $PE = -Gm_1m_2/r$. Thus the potential energy is the work done by the gravitational force of the body moves from $r = \infty$ to $r = r$ with $m$. Note that the mutual gravitational potential energy changes when the two bodies are interated.

Both kinetic and potential energy has units (joules) as work, so we can combine 1–30 and 1–31 to obtain

$$(KE + PE)_B = (KE + PE),$$

or, in terms of the total energy of the system,

$$TE_B = TE_A = \text{constant} \quad (1–32)$$

Therefore, the total energy of our gravitational system is conserved:
for problems relating to orbits. Energy often proves a more powerful way to gain insight into orbital problems.

Energy is a quantity assigned to one body that indicates that body’s ability to change the state of another body. Heat is a form of energy, for a hot body will warm a cold body if the two are brought into contact. Electrical energy causes the filament of a light bulb to glow (become hot), showing that one type of energy may be converted to another. Kinetic energy (KE) is a body’s energy of motion, but if we decide to move along with the body, it has no relative motion; it has no kinetic energy in this reference frame. Potential energy (PE) is due to the position of the body; if the body is free to move, this energy may be converted to kinetic energy. In celestial mechanics, we sum the kinetic and potential energies to obtain the **total energy**: 
\[ TE = KE + PE. \]

Assume that a force \( F \) acts upon a body of mass \( m \) that is moving in the trajectory \( x(t) \) about a central force. In the infinitesimal time \( dt \), the body moves through the vector distance \( dx \). As the body moves from position A to position B, we define the work \( W \) (SI units = \( \text{kg} \cdot \text{m}^2/\text{s}^2 = \text{joules} \); see Appendix 6) done on the body by the force as

\[ W = \int_A^B F \cdot dx \quad (1-29) \]

**The vector dot product** operation in Equation 1–29 yields the product of the **parallel components** of \( F \) and \( dx \) (Appendix 9, “Mathematical Operations”); when \( F \) and \( dx \) are mutually perpendicular, their dot product vanishes. To evaluate Equation 1–29, we note:

\[ F \cdot dx = m(dv/dt) \cdot v \ dt = m(v \cdot dv) = d(mv^2/2) \]

where we have used Newton’s second law and the definitions of velocity and speed. We integrate Equation 1–29 directly to give

\[ W = (mv^2/2)_B - (mv^2/2)_A = KE_B - KE_A \quad (1-30) \]

where the kinetic energy is specified by \( KE = mv^2/2 \). Therefore, the work done by the force on the body changes the body’s kinetic energy. In a system of two bodies of masses \( m_1 \) and \( m_2 \), the kinetic energy is

\[ KE = (m_1v_1^2/2) + (m_2v_2^2/2). \]

If the force is the gravitational force between the two bodies, we have

\[ F \cdot dx = -GM_1m_2/r^2 \ dr = d(Gm_1m_2/r) \]

where \( r \) is the radial separation of the bodies. Equation 1–29 is again integrated:

\[ W = (Gm_1m_2/r)_B - (Gm_1m_2/r)_A = PE_A - PE_B \quad (1-31) \]

where we have inserted the definition of the mutual gravitational potential energy \( PE = -Gm_1m_2/r \). Thus the potential energy is the negative of the work done by the gravitational force as \( m_1 \) moves from \( r = \infty \) to \( r = r \) with \( m_2 \) held fixed. Note that the mutual gravitational potential energy vanishes when the two bodies are infinitely separated.

Both kinetic and potential energy have the same units (joules) as work, so we can combine Equations 1–30 and 1–31 to obtain

\[(KE + PE)_B = (KE + PE)_A\]

or, in terms of the total energy of the two-body system,

\[ TE_B = TE_A = \text{constant} \]

Therefore, the total energy of our gravitating system is conserved:

\[ TE = (m_1v_1^2/2) + (m_2v_2^2/2) - (Gm_1m_2/r) = \text{constant} \quad (1-32) \]
As the bodies move, kinetic and potential energy may be interchanged, but the total energy of the system remains constant. Now to evaluate the constant TE in Equation 1–32 for elliptical orbits. Refer to Section 1–3(c) and recall that the total linear momentum of our isolated system is constant; we choose this constant to be zero so that \( m_1v_1 = -m_2v_2 \) and, in terms of the speeds of the bodies, \( m_1v_1 = m_2v_2 \). Since \( v = v_1 + v_2 \) is the relative speed of either body with respect to the other, we find
\[
\begin{align*}
v_1 &= m_2v/(m_1 + m_2) \\
v_2 &= m_1v/(m_1 + m_2)
\end{align*}
\]
Substituting this result into Equation 1–32 yields
\[
TE = m_1m_2[v^2/2(m_1 + m_2) - (G/r)] \quad (1–33)
\]
Using Equation 1–24, let’s evaluate Equation 1–33 at the perihelion of the orbit, where \( r = a(1 - e) \) and \( v \) is the periheleon speed given by Equation 1–17. The result, \( TE = -Gm_1m_2/2a \), shows that the total energy is negative—the orbit is bound. Now Equation 1–33 takes its final form:
\[
v^2 = G(m_1 + m_2)[(2/r) - (1/a)] \quad (1–34)
\]
exactly the expression found in Equation 1–28. This classic result, the vis viva equation, is a statement of total energy conservation.

PROBLEMS

1. Assume that the orbital plane of a superior planet is inclined 10° to the ecliptic and that the planet crosses the ecliptic moving northward at opposition. Make a diagram similar to Figure 1–1B, showing the retrograde path of this superior planet.

2. Imagine you are observing the Earth from Jupiter. What would you observe the Earth’s synodic orbital period to be? What would it be from Venus? (Hint: See Appendix 3.)

3. (a) Explicitly carry out the derivation of Equation 1–4, showing all the appropriate steps. (b) On graph paper, plot the following polar equations: Equation 1–3 for an ellipse, Equation 1–6 for a parabola, and Equation 1–7 for a hyperbola.

4. In terms of the gravitational acceleration \( g \) at the surface of Earth, find the surface gravitational acceleration of (a) the Moon (\( M_m = 0.0123M_{\oplus}, R_m = 1738 \text{ km} \)), (b) the Sun (\( M_S = 2 \times 10^{30} \text{ kg}, R_S = 7 \times 10^8 \text{ m} \)), and (c) Jupiter (\( M_J = 318M_{\oplus}, R_J = 11.2 \text{ R}_\oplus \)).

5. What are the perihelion and aphelion speeds of Mercury? What are the perihelion and aphelion distances of this planet? Compute the product \( vr \) (speed times distance) at each of these two points and interpret your result physically.

6. Find the relative position of the center of mass for (a) the Sun–Jupiter system and (b) the Earth–Moon system.

7. A television satellite is in circular orbit about the Earth, with a sidereal period of exactly 24 h. What is the distance from the Earth’s surface for such a satellite? (Hint: Use Kepler’s laws.) If the satellite appears stationary to an earthbound observer, what is the orientation of its orbital plane?

8. Using orbital data for Titan (Appendix 3), find the mass of Saturn.

9. A stone is released from rest at the Moon’s orbit and falls toward Earth. What is the stone’s speed when it is 192,000 km from the center of Earth?

10. An object is observed from the Earth to have a synodic period of 1.5 years. What are the two possible values for the semimajor axis of the object’s orbit?

11. Using orbital data for the Earth found in Appendix 3, estimate the mass of the Sun. Does the mass of the Earth matter significantly in this calculation?

12. Compare the mutual gravitational force between you and the following two objects: (a) another person with mass 100 kg located 1 m from you; (b) Mars at opposition. Comment on your result.

13. (a) Venus has a maximum elongation of 47°. What is its distance from the Sun in astronomical units? (b) Mars has a synodic period of 779.9 days and a sidereal period of 686.98 days. On February 11, 1990, Mars had an elongation of 43° West. The elongation of Mars 687 days later, on December 30, 1991, was 15° West. What is the distance of Mars from the Sun in astronomical units?