Chapter

3 Orbital Dynamics

In 1608, Johannes Kepler (1571–1630) published two laws of planetary motion that he deduced from an analysis of the accurate observational data he inherited from his employer, the Danish astronomer Tycho Brahe (1546–1601). Ten years later, Kepler published a third law called the Harmonic Law. These three laws, which fifty years later led Isaac Newton (1642–1727) to the discovery of the law of gravitation, can be expressed as follows:

1. *The planets move in ellipses, with the Sun at one focus;*
2. *The radius vector from the Sun to a planet sweeps out equal areas in equal times;*
3. *The period of revolution squared is proportional to the semi-major axis cubed.*

The second law is the most general because it is true for any two-body central force problem. The first law is a consequence of the fact that the gravitational force of one body on another depends on the inverse square of their separation, and the third law reflects the fact that the gravitational force is proportional to the masses of the bodies. We will derive Kepler's Laws from Newton's Law of Gravitation, and then will discuss what happens when there are more than two bodies to consider.

3.1 Two-Body Central Force

As a prelude to understanding the motions of solar system objects, we treat the problem of two bodies that move under the influence of a mutual central force. Consider two bodies with masses m_1 and m_2 whose position vectors relative to the center of mass are $\mathbf{r}_1(t)$ and $\mathbf{r}_2$(t), respectively. A central force is one that depends only on the difference vector, $\mathbf{r} \equiv \mathbf{r_2} - \mathbf{r_1}$ (where $|\mathbf{r}| = |r_x^2 + r_y^2|^{1/2}$) and its time derivatives

$$\mathbf{F} = \mathbf{F}(\mathbf{r},\, \dot{\mathbf{r}},\, \ldots)\,. \tag{3.1}$$

A **conservative central force**, like gravity, is one that can be expressed as the gradient of a potential that depends only on the magnitude of the difference vector, $|\mathbf{r}|$. We use polar coordinates in which $(x, y) = (rcos\phi, rsin\phi)$ and so $|\mathbf{r}| = [r^2cos^2\phi + r^2sin^2\phi]^{1/2} = r$. In this system

$$\mathbf{F}(r) = -\nabla\Phi(r) = -\left[\hat{\mathbf{e}}_r\frac{\partial\Phi}{\partial r} + \hat{\mathbf{e}}_\phi\frac{1}{r}\frac{\partial\Phi}{\partial\phi}\right]. \tag{3.2}$$

Along with the difference vector there is a second natural position vector, the **center of mass** of the system, $\mathbf{R}(t)$:

$$\mathbf{R}(t) \equiv \frac{\sum_i m_i\,\mathbf{r}_i(t)}{\sum_i m_i} = \frac{m_1\,\mathbf{r}_1 + m_2\,\mathbf{r}_2}{m_1 + m_2}. \tag{3.3}$$

where i represents the summation over the particles in the system. By expressing $\mathbf{r}_1$ and $\mathbf{r}_2$ in terms of $\mathbf{R}$ and $\mathbf{r}$:

$$\mathbf{r}_1 = \mathbf{R} - \frac{m_2}{m_1 + m_2}\mathbf{r}, \tag{3.4}$$

$$\mathbf{r}_2 = \mathbf{R} + \frac{m_1}{m_1 + m_2}\mathbf{r}, \tag{3.5}$$

the kinetic energy K may be written:

$$\begin{aligned}
K &= \sum_i \frac{1}{2}m_i\,|\dot{\mathbf{r}}_i|^2 = \frac{1}{2}m_1|\dot{\mathbf{r}}_1|^2 + \frac{1}{2}m_2|\dot{\mathbf{r}}_2|^2\,, \\
&= \frac{1}{2}m_1\left|\dot{\mathbf{R}} - \left(\frac{m_2}{m_1+m_2}\right)\dot{\mathbf{r}}\right|^2 + \frac{1}{2}m_2\left|\dot{\mathbf{R}} + \left(\frac{m_1}{m_1+m_2}\right)\dot{\mathbf{r}}\right|^2, \\
&= \frac{1}{2}(m_1+m_2)|\dot{\mathbf{R}}|^2 + \left[\frac{1}{2}m_1\left(\frac{m_2}{m_1+m_2}\right)^2 + \frac{1}{2}m_2\left(\frac{m_1}{m_1+m_2}\right)^2\right]|\dot{\mathbf{r}}|^2\,, \\
&= \frac{1}{2}(m_1+m_2)|\dot{\mathbf{R}}|^2 + \frac{1}{2}\left(\frac{m_1m_2}{m_1+m_2}\right)|\dot{\mathbf{r}}|^2\,.
\end{aligned} \tag{3.6}$$

In Equation (3.6), the first term on the RHS corresponds to the motion *of* the center of mass and the second term represents the motion *about* the center of mass. The quantity

$$\mu \equiv \left(\frac{m_1m_2}{m_1+m_2}\right) \tag{3.7}$$

is called the **reduced mass** of the system, which can be rewritten

$$\frac{1}{\mu} = \frac{1}{m_1} + \frac{1}{m_2}.$$

Because there are no external forces in this problem, no equations of the motion for $\mathbf{r}$ contain terms with $\mathbf{R}$. The center of mass is either at rest or is moving uniformly, meaning that this term is constant and can be dropped. Thus, for conservative central forces, it is possible to express a **two-body problem** as an **equivalent one-body problem**, with a center of force at the origin and with a single body of mass μ located at a distance $\mathbf{r}(t)$ from the origin.

3.1.1 Constants of Motion

For a body under the influence of a central conservative force, the potential is only a function of the radial distance from the center of mass. The problem thus has spherical symmetry in which any rotation about a fixed axis will have no effect on the solution. This yields important simplifications. In a spherically symmetric system of the form

$$\mathbf{F}(r) = F(r)\,\frac{\mathbf{r}}{|\mathbf{r}|}\,,$$

angular momentum, $\mathbf{L}$, is conserved. In such a system $\mathbf{r}$ must be normal to the direction of $\mathbf{L}$, and the time rate of change of angular momentum, the torque, $\mathbf{N}$, is zero:

$$\frac{d\mathbf{L}}{dt} \equiv \mathbf{N} = \mathbf{r} \times \mathbf{F} = (\mathbf{r} \times \mathbf{r})\frac{1}{|\mathbf{r}|}F(r) = 0. \tag{3.8}$$

Consequently, the body will never be torqued out of its plane of motion. It is convenient to use polar coordinates to describe this system. Let r and ϕ denote the radial and azimuthal coordinates, with corresponding unit vectors $\hat{r}$ and $\hat{\phi}$. A sketch of the geometry reveals that the cartesian components of the unit vectors are $\hat{r} = (\cos\phi, \sin\phi)$ and $\hat{\phi} = (-\sin\phi, \cos\phi)$, which implies the identities:

$$\frac{d\hat{r}}{d\phi} = \hat{\phi}, \qquad \frac{d\hat{\phi}}{d\phi} = -\hat{r}. \tag{3.9}$$

Since the position vector is a product of two functions, $\mathbf{r} = r\,\hat{r}$, the velocity vector becomes a sum of two terms:

$$\begin{aligned} \mathbf{v} \equiv \frac{d}{dt}\mathbf{r} &= \frac{d}{dt}(r\,\hat{r}) = \dot{r}\hat{r} + r\frac{d\hat{r}}{d\phi}\dot{\phi}\,, \\ &= \dot{r}\hat{r} + r\dot{\phi}\hat{\phi}\,, \end{aligned} \tag{3.10}$$

where dots indicate time derivatives, and we have used (3.9) and the chain rule for differentiation. A second time derivative yields the acceleration vector:

$$\begin{aligned} \mathbf{a} \equiv \frac{d}{dt}\mathbf{v} &= \ddot{r}\,\hat{r} + \dot{r}\,\frac{d\hat{r}}{d\phi}\,\dot{\phi} + \dot{r}\,\dot{\phi}\,\hat{\phi} + r\,\ddot{\phi}\,\hat{\phi} + r\,\dot{\phi}\,\frac{d\hat{\phi}}{d\phi}\,\dot{\phi}\,, \\ &= \ddot{r}\,\hat{r} + \dot{r}\,\dot{\phi}\,\hat{\phi} + \dot{r}\,\dot{\phi}\,\hat{\phi} + r\,\ddot{\phi}\,\hat{\phi} - r\,\dot{\phi}^2\,\hat{r}\,, \\ &= (\ddot{r} - r\,\dot{\phi}^2)\,\hat{r} + (r\,\ddot{\phi} + 2\dot{r}\,\dot{\phi})\,\hat{\phi}\,. \end{aligned} \tag{3.11}$$

Thus, the radial and azimuthal components of the vector equation of motion, $\mu \mathbf{a} = \mathbf{F}$, are

$$\mu \ddot{r} - \mu r \dot{\phi}^2 = F(r) = -\frac{d\Phi}{dr}, \tag{3.12}$$

$$\mu r \ddot{\phi} + 2\mu \dot{r} \dot{\phi} = 0. \tag{3.13}$$

This is a system of two second-order, ordinary differential equations. The constants of motion that lead to the solutions for $r(t)$ and $\phi(t)$ are straightforward to find. First, multiply (3.13) by r, and then rearrange the equation into the form $d(\cdots)/dt$:

$$\mu r^2 \ddot{\phi} + 2\mu r \dot{r} \dot{\phi} = \frac{d}{dt}\left(\mu r^2 \dot{\phi}\right) = 0. \tag{3.14}$$

The constant of motion in (3.14) is the magnitude of the angular momentum, L:

$$|\mathbf{L}| \equiv L = \mu r^2 \dot{\phi} = \text{constant}. \tag{3.15}$$

Equation (3.15) will be used repeatedly in what follows to eliminate the term $\dot{\phi}$ whenever it appears. Our first opportunity is to eliminate $\dot{\phi}^2$ from (3.12), after which we multiply by the radial velocity $\dot{r}$:

$$\mu \dot{r} \ddot{r} - \frac{L^2}{\mu r^3} \dot{r} + \dot{r} \frac{d\Phi}{dr} = \frac{d}{dt}\left(\frac{1}{2}\mu \dot{r}^2 + \frac{L^2}{2\mu r^2} + \Phi\right) = 0. \tag{3.16}$$

The new constant of motion revealed in (3.16) is the total energy, E_{tot}:

$$E_{tot} = \frac{1}{2}\mu \dot{r}^2 + \frac{L^2}{2\mu r^2} + \Phi. \tag{3.17}$$

The total energy is the sum of the kinetic and potential energies:

$$K \equiv \frac{1}{2}\mu \dot{r}^2, \quad \Phi_e \equiv \frac{L^2}{2\mu r^2} + \Phi, \tag{3.18}$$

where Φ_e is the **effective potential energy**. The term $L^2/(2\mu r^2)$ arises from the centripetal acceleration term, $-\mu r \dot{\phi}^2$, in (3.12).

3.1.2 Kepler's 2nd Law

Kepler's 2nd Law is equivalent to the conservation of angular momentum. To see this, note that the differential area dA swept out by the radius vector when the body moves through a differential angle $d\phi$ is a triangle whose area is given by one-half its base times its height as

$$dA = \frac{1}{2}(r\, d\phi)\, r. \tag{3.19}$$

Consider the **areal velocity**, which is the area swept out by the radius vector per unit time. Conservation of angular momentum is equivalent to saying that areal velocity is constant. We divide the differentials in (3.19) by dt and use (3.15) to obtain

$$\frac{dA}{dt} = \dot{A} = \frac{1}{2} r^2 \dot{\phi} = \frac{1}{2}\frac{L}{\mu} = \text{constant}, \tag{3.20}$$

which is **Kepler's 2nd Law**. If the motion is periodic with period τ, then the total area enclosed by the orbit is given by:

$$A = \frac{1}{2}\frac{L}{\mu}\tau \,. \tag{3.21}$$

3.1.3 Formal solution

For problems with three or more bodies, chaotic trajectories are part of the solution space and it is impossible to write down a complete analytical solution. But, for the two-body problem, there are a sufficient number of constants of motion, namely L and E_{tot}, to solve for $r(t)$ and $\phi(t)$. First, we can use (3.17) to solve for $\dot{r}$ as

$$\dot{r} = \frac{dr}{dt} = \left[\frac{2}{\mu}\left(E_{tot} - \Phi(r) - \frac{L^2}{2\mu r^2}\right)\right]^{1/2} \,. \tag{3.22}$$

Next we may obtain an implicit formula for $r(t)$ in the form $t(r)$ by integrating (3.22) with respect to r:

$$t(r) = \int_{r_0}^{r} \left[\frac{2}{\mu}\left(E_{tot} - \Phi(r) - \frac{L^2}{2\mu r^2}\right)\right]^{-1/2} dr \,, \tag{3.23}$$

where r_0 is the radius at $t = 0$. In practice it is difficult to invert (3.23) to obtain an explicit formula for $r(t)$, and so (3.23) is referred to as the **formal solution**. Even in the case when $t(r)$ cannot easily be inverted, (3.23) provides a framework for studying the dynamics with powerful approximate techniques.

Assuming that we have found $r(t)$ from (3.23), we can then obtain $\phi(t)$ by integrating (3.15), which leads to:

$$\phi(t) = \phi_0 + \frac{L}{\mu}\int_0^t r^{-2}\, dt, \tag{3.24}$$

where ϕ_0 is the azimuthal angle at time $t = 0$. The four constants of integration, r_0, ϕ_0, L, and E_{tot}, can be determined from the initial position and velocity components, r_0, ϕ_0, $\dot{r}_0$, $\dot{\phi}_0$, and this completes the formal solution to the two-body central force problem. Note that we have not yet specified the form of $\Phi(r)$.

3.2 Inverse-Square Law

For an arbitrary central force the period of radial motion is not necessarily the same as the period of azimuthal motion. If it is, then the orbit is closed. In 1873, J. Bertrand showed that the only central forces that produce closed orbits for all bounded motions are the inverse-square law, $F(r) = -kr^{-2}\,\hat{r}$, and Hooke's Law, $F(r) = -kr\,\hat{r}$. The inverse-square law applies to the mutual attraction of masses or electrical charges, and Hooke's law applies to the small oscillations of springs. (For a proof of Bertand's theorem see Goldstein, 1980, Appendix A, 601.)

To derive Kepler's 1st and 3rd Laws, we will now focus on the attractive inverse-square-law force and its corresponding potential:

$$F(r) = -kr^{-2}\,, \quad \Phi(r) = -kr^{-1}, \tag{3.25}$$

where k is a constant of proportionality. In order to determine the nature of the orbits allowed by (3.25), it is convenient to set $r = u^{-1}$ and write the equations of motion in terms of u. With the following identities

$$\dot{r} = -\frac{1}{u^2}\frac{du}{d\phi}\,\dot{\phi} = -r^2\dot{\phi}\,\frac{du}{d\phi} = -\frac{L}{\mu}\frac{du}{d\phi}\,, \tag{3.26}$$

$$\ddot{r} = -\frac{L}{\mu}\frac{d^2u}{d\phi^2}\,\dot{\phi} = -\frac{L^2u^2}{\mu^2}\frac{d^2u}{d\phi^2}, \tag{3.27}$$

the equation of radial motion (3.12), with the centripetal acceleration written in terms of the angular momentum L using (3.15), becomes:

$$\frac{d^2u}{d\phi^2} + u = -\frac{\mu}{L^2}\frac{1}{u^2}\,F\left(\frac{1}{u}\right)\,, \tag{3.28}$$

after multiplying by $-\mu/(L^2u^2)$. For the case $L = 0$, (3.28) blows up, but (3.15) implies that this special case has constant ϕ and just describes straight-line motion through the origin.

Writing (3.28) for the inverse-square-law (3.25) yields

$$\frac{d^2u}{d\phi^2} + u = \frac{\mu k}{L^2}\,. \tag{3.29}$$

This is the same as the equation of a harmonic oscillator under the influence of a constant force, where the angle ϕ is playing the role usually played by time, t. The corresponding homogeneous equation is

$$\frac{d^2u}{d\phi^2} + u = 0\,, \tag{3.30}$$

which is solved in general by

$$u = u_0\,\cos(\phi - \phi_0)\,, \tag{3.31}$$

where u_0 and ϕ_0 are arbitrary constants. By inspection of the inhomogeneous equation (3.29), one solution is simply the constant solution

$$u = \frac{\mu k}{L^2}, \tag{3.32}$$

which means that the general solution to this problem is:

$$
\begin{aligned}
u &= \frac{\mu k}{L^2} + u_0 \cos(\phi - \phi_0)\,, \\
\Rightarrow \quad \frac{1}{r} &= \frac{\mu k}{L^2}\left[1 + e\cos(\phi - \phi_0)\right], \qquad (3.33)
\end{aligned}
$$

where $e \equiv u_0 L^2/(\mu k)$.

3.2.1 Kepler's 1st Law

The equation for a conic section in polar coordinates with one focus at the origin is

$$
r(\phi) = \frac{a(1-e^2)}{1 + e\cos(\phi - \phi_0)}\,.
$$

Thus, (3.33) is the polar-coordinate formula for a conic section with the origin at one focus. Depending on the value of the **eccentricity**, e, $r(\phi)$ traces out a circle ($e = 0$), an ellipse ($0 < e < 1$), a parabola ($e = 1$), or a hyperbola ($e > 1$). Corresponding to the extremes of the $\cos(\phi - \phi_0)$ factor in (3.33) are two **turning points** or **apsidal distances** $\{r_1, r_2\}$ that satisfy:

$$
\frac{1}{r_1} = \frac{\mu k}{L^2}(1+e)\,, \quad \frac{1}{r_2} = \frac{\mu k}{L^2}(1-e)\,. \qquad (3.34)
$$

For the case of a hyperbola there is only one physical turning point, r_1, since $e > 1$ makes $r_2 < 0$.

Consider the case that describes an ellipse, $0 < e < 1$. At the two turning points, the radial velocity $\dot{r}$ is zero, and therefore by the equation for E_{tot} (3.17):

$$
E_{tot} = \frac{L^2}{2\mu r^2} - \frac{k}{r}\,, \quad \Rightarrow \quad \frac{1}{r^2} - \frac{2\mu k}{L^2}\frac{1}{r} - \frac{2\mu E_{tot}}{L^2} = 0\,. \qquad (3.35)
$$

Solving (3.35) for r^{-1} using the quadratic formula yields

$$
\frac{1}{r_1} = \frac{\mu k}{L^2}\left(1 + \sqrt{1 + \frac{2E_{tot}L^2}{\mu k^2}}\right), \quad \frac{1}{r_2} = \frac{\mu k}{L^2}\left(1 - \sqrt{1 + \frac{2E_{tot}L^2}{\mu k^2}}\right). \qquad (3.36)
$$

By comparing (3.36) with (3.34), the relationship between energy and eccentricity becomes apparent:

$$
e = \sqrt{1 + \frac{2E_{tot}L^2}{\mu k^2}}\,. \qquad (3.37)
$$

The orbit type depends on E_{tot} in the following manner:

$E_{tot} = -\frac{\mu k^2}{2L^2}$,	$e = 0$,	circle,
$E_{tot} < 0$,	$0 < e < 1$,	ellipse,
$E_{tot} = 0$,	$e = 1$,	parabola,

$$E_{tot} > 0, \qquad e > 1, \qquad \text{hyperbola.}$$

Since the planets are in bound orbits with $E_{tot} < 0$, to the accuracy of this simple two-body model the planets are in elliptical orbits with the Sun at one focus, which is Kepler's 1st Law.

3.2.2 Kepler's 3rd Law

Notice from (3.36) that

$$\frac{1}{r_1} + \frac{1}{r_2} = \frac{2\mu k}{L^2}\,, \tag{3.38a}$$

$$\frac{1}{r_1} \times \frac{1}{r_2} = \left(\frac{\mu k}{L^2}\right)^2 \left[1 - \left(1 + \frac{2E_{tot}L^2}{\mu k^2}\right)\right] = -\frac{2\mu E_{tot}}{L^2}\,. \tag{3.38b}$$

With (3.38) we can find a simple relationship between E_{tot} and and the orbit's semi-major axis, a, which is one-half the sum of the apsidal distances

$$a = \frac{r_1 + r_2}{2} = \frac{1}{2}\left(\frac{1}{r_1} + \frac{1}{r_2}\right) r_1 r_2 = -\frac{\mu k}{L^2}\frac{L^2}{2\mu E_{tot}} = -\frac{k}{2E_{tot}}\,,$$

or

$$E_{tot} = -\frac{k}{2a}\,. \tag{3.39}$$

Notice that E_{tot} depends on a, but does *not* depend on e or L. Thus for elliptical orbits the semi-major axis depends only on energy. The area A of an ellipse is $\pi ab = \pi a^2(1 - e^2)^{1/2}$. Using this fact, together with (3.21), (3.37), and (3.39), it is a homework problem to show that the period τ satisfies

$$\tau^2 = 4\pi^2 a^3\,\frac{\mu}{k}\,. \tag{3.40}$$

Now, consider Newton's law of gravitation

$$F(r) = -\frac{Gm_1m_2}{r^2}, \quad \Phi(r) = -\frac{Gm_1m_2}{r}, \quad k = Gm_1m_2\,, \tag{3.41}$$

where G is the universal gravitation constant. With $k = G\,m_1m_2$, (3.40) yields:

$$\tau^2 = 4\pi^2 a^3\,\frac{1}{G(m_1 + m_2)}\,. \tag{3.42}$$

Since the mass of the sun is much larger than the mass of any of the planets, to a good approximation (3.42) says that the square of the period of revolution is proportional to the cube of the semi-major axis for the planets, which is Kepler's 3rd Law. The magnitude of the error introduced by neglecting the planetary mass is explored in the homework. Equation (3.42) allows us to "weigh" the total mass of a distant two-body system, like Pluto and its satellite Charon, by observing the system's period and semimajor axis. In passing, we note that an alternate way to write (3.42) is to introduce the average angular velocity, which planetary scientists call the **mean motion**, $n \equiv 2\pi/\tau$, in which case (3.42) becomes:

$$n^2 = \frac{G(m_1 + m_2)}{a^3} \,. \tag{3.43}$$

3.3 Kepler's Equation

To find the position of a planet in its orbit at a given time, $r(t)$, we need to invert $t(r)$ as given by (3.23). The factor L^2/μ in (3.23) can be rewritten for an elliptical orbit using $A = L\tau/(2\mu)$ and (3.42):

$$\begin{aligned} \frac{L^2}{\mu} &= \frac{(2\mu\pi ab)^2/\tau^2}{\mu} = 4\mu\pi^2 a^2 a^2 (1 - e^2) \frac{G(m_1 + m_2)}{4\pi^2 a^3} \,, \\ &= G m_1 m_2 a (1 - e^2) \,. \end{aligned} \tag{3.44}$$

Write (3.23) with $E_{tot} = -Gm_1m_2/(2a)$ from (3.39), $\Phi(r) = -Gm_1m_2/r$ from (3.41), and $L^2/\mu = Gm_1m_2\, a(1 - e^2)$ from (3.44):

$$\begin{aligned} t(r) &= \int_{r_0}^{r} \left\{ \frac{2}{\mu} \left(-\frac{Gm_1m_2}{2a} + \frac{Gm_1m_2}{r} - \frac{Gm_1m_2\, a(1 - e^2)}{2r^2} \right) \right\}^{-1/2} dr \,, \\ &= \Big\{ 2G(m_1 + m_2) \Big\}^{-1/2} \int_{r_0}^{r} \left\{ r - \frac{1}{2a} r^2 - \frac{a(1 - e^2)}{2} \right\}^{-1/2} r\, dr \,, \end{aligned} \tag{3.45}$$

where the starting time is taken such that r_0 corresponds to periapse (*i.e.*, minimum r). One carry-over from medieval astronomy is that angles are called **anomalies**, for example ϕ, defined such that $\phi = 0$ corresponds to periapse, is called the **true anomaly**. Before integrating (3.45), it is convenient to introduce a new angle E (not to be confused with energy) called the **eccentric anomaly**, which is defined by:

$$r = a(1 - e \cos E) \,. \tag{3.46}$$

Notice that periapse and apoapse correspond to $E = 0$ and $E = \pi$, respectively. By using (3.46) to replace r in favor of E in (3.45), employing the identity $\sin E/(1 - \cos^2 E)^{1/2} = 1$, multiplying by n, and using (3.44), the integral takes the simple form:

$$nt = \int_0^E (1 - e \cos E)\, dE \,, \tag{3.47}$$

which can be evaluated to give:

$$M \equiv nt = E - e\, \sin E \,, \tag{3.48}$$

were M is called the **mean anomaly**. Equation (3.48) is a transcendental equation called **Kepler's Equation**. Given t in the form of the mean anomaly, we must invert (3.48) to obtain E, and then use E in (3.46) to obtain r. The need to invert (3.48) to predict the locations of planets motivated some of the early work in numerical methods.

3.4 Tides and Resonances

With Newton's laws it is possible to explain tides in terms of the gravitational attraction. All planet-moon systems experience tidal interactions, several of which can be significant in terms of orbital and/or thermal evolution. A simple introduction to the concept of tides comes from consideration of the inverse square law. If a planet and moon are a distance a apart and the planet on which the tide is being analyzed has a radius R, then the force on the near side is proportional to $1/(a-R)^2$, while the force at the center of the planet is $1/a^2$ and the force on the farside varies as $1/(a+R)^2$. The differential force between the nearside and center is

$$\frac{1}{(a-R)^2} - \frac{1}{a^2} = \frac{a^2 - a^2 + 2aR - R^2}{a^2(a^2 - 2aR + R^2)}$$

or

$$\frac{1}{(a-R)^2} - \frac{1}{a^2} = \frac{2R - (R^2/a)}{a^3 - 2a^2R + aR^2}.$$

Because R is usually much less than $a$the first terms in the numerator and denominator on the RHS are the largest and the others can be ignored to first order. So the size of the tidal buldge varies as $2R/a^3$. The tide varies with the radius of the planet such that larger planets have larger tides than smaller planets for the same disturbing potential. And the tide varies with the cube of the distance of the moon so that a closer moon raises larger tides. Also note that the moon raises a tide both on the nearside and farside of the planet.

3.4.2 Active Volcanism on Io

In the outer solar system, orbital and rotational mechanics represents a mechanism for heating of small bodies. As an example of this process we consider the case of Jupiter's innermost Galilean satellite, Io. Io is the most volcanically active body in the solar system. This may seem surprising at first, because Io is about the size and density of Earth's Moon, and the Moon hasn't seen intense volcanism since early in its history. Even as the Voyager 1 spacecraft headed towards its closest approach with the Jovian system it was still widely expected that Io would not turn out to be all that unusual. In fact, the press conferences were over and the team scientists were heading home when, on March 8, 1979, navigation engineer Linda Morabito was examining an image of Io to facilitate a refinement of Voyager's trajectory and discovered an active volcanic plume. At first, she thought it might be another moon coming around the limb of Io, but she almost immediately realized that this was not the case.

We now know that Io is experiencing enormous tidal dissipation because it is locked in a three-way orbital resonance with its neighbors Europa and Ganymede. All three of Kepler's laws will be mentioned in the following discussion of the effects of tidal dissipation on satellites.

3.4.3 An Enigmatic Satellite

Even before Voyager 1 photographed active volcanoes on Io, there were occasional hints from Earth-based observations that Io was an unusual satellite. The following is a brief history of pre-Voyager studies of Io (after Nash *et al., Satellites*, p. 631-635):

1610. Io is discovered by Galileo.

1805. Laplace studies the 4:2:1 resonance that bears his name.

1927. Io has a pronounced variation in brightness with orbital phase angle.

1964. A report of an anomalous brightening of Io's surface as it emerged from behind Jupiter. Bursts of decametric radio emission from Jupiter are apparently controlled by Io's orbital position.

1968-69. Models of the electrodynamical and magnetospheric coupling between Io and Jupiter are published.

1970's. Spectral reflectance observations are made, mostly looking for evidence of a Moon-like surface with perhaps some water or ammonia frost.

1971. Io occults a bright star, providing an accurate determination of its radius.

1972. Discordant photometry at visible and infrared wavelengths.

1973. Io is determined to have dark poles and a relatively bright equator. Spectral evidence for sulfur on Io is discovered, and the expectation of water or ammonia frost is not upheld. Sodium D line emission is discovered in Io's spectrum. Pioneer encounters Jupiter and discovers an ionosphere and a thin atmosphere on Io.

1975. Ionized sulfur emission in the inner Jovian magnetosphere. More spectral evidence for sulfur on Io's surface.

1977. Strong absorption near 4 μm (later identified as SO_2) .

1979, before Voyager encounters. Observation of intense temporary brightening in the infrared from 2μm to 5μm. Interpretation that some of Io's surface is 600 K, compared with a daytime average of 130 K. Skepticism. Prediction of "...widespread and recurrent volcanism..." based on tidal dissipation and the Laplace resonance.

1979, Voyager encounters. Absence of impact craters. Intense surface coloration. Active volcanic plumes. SO_2 gas. Surface hotspots.

The following is a brief summary of the paper "Dynamic geophysics of Io," by A. Mc Ewen, J. Lunine, and M. Carr, in "Time-variable phenomena in the Jovian system," $NASASP-494$.

Among the observations made by the Voyager spacecraft in 1979 were active volcanoes up to 300 km high, hot spots more energetic than thousands of Yellowstones, and color and albedo changes over areas the size of Alaska. Changes in plume activity, surface features, and thermal emission have been observed on time scales ranging from hours to years.

Io and Earth's Moon have similar sizes and densities. However, the Moon has a global heat flow of about 0.02 Wm^{-2}, Io's heat flow is 1 to 3 Wm^{-2}.

Io's bulk composition is probably silicate, but the surface composition has been profoundly altered by volcanic resurfacing and outgassing. SO_2 has been positively identified as both a surface frost and a gas, with the frost possibly covering as much as 30% of the surface. The colors are thought to be caused by allotropes of sulfur, but there is some controversy over this idea.

The presence of 10-km high mountains implies a thicker lithosphere than originally envisioned by Peale *et al.*

Io's plumes erupt under conditions of pressure, temperature, and gravity that are very different from those on Earth. The plumes are probably driven by SO_2 and/or sulfur, in which multiple phase changes can occur, so the plume dynamics and thermodynamics are unlike terrestrial eruptions.

Nine eruption plumes were observed during the Voyager 1 encounter, and eight of these were re-observed by Voyager 2 four months later. Many of the plumes were 50-100 km high, but Pele, active during the first encounter and having the greatest height (≈ 300 km), had ceased activity sometime before the second encounter. The plume eruptions can be grouped into two major classes:

1. Prometheus-type eruptions: smaller (50-120 km high, 200-600 km diameter surface deposits), long lived (months to years), contain optically thick, dark jets, deposit bright white materials, erupt at velocities of $\approx 0.5\,\mathrm{km\,s^{-1}}$, and are concentrated at low latitudes in an equatorial band around the satellite.

2. Pele-type eruptions: large (≈ 300 km high, 1000-1500 km diameter surface deposits), are optically thin, deposit relatively dark red material, erupt at $\approx 1.0\,\mathrm{km\,s^{-1}}$, and occur in the region from longitude 240° to 360°.

A variety of ways of estimating the total resurfacing rate yield a range of 10^{-3} to $10\,\mathrm{cm\,yr^{-1}}$. The absence of impact craters on the visible surface of Io requires a rate greater than 0.1 $\mathrm{cm\,yr^{-1}}$. The energy required to bring internal material to the surface at this rate is less than 10% of the global average heat loss.

3.4.4 Orbital Resonances

The solar system is full of curious resonances and apparent coincidences between the orbital elements of its individual constituents. For example, it is quite common to find satellites, like Earth's Moon, that always keep the same face pointed towards their host planet. An even more synchronized state can hold in which "the day equals the month" and the satellite and planet rotate like a dumbbell, which is surmised to be the case for Pluto and Charon. Not all satellites are behaving so simply in the current solar system. The non-spherical shape of Saturn's satellite Hyperion allows for gravitational torques to cause it to tumble chaotically.

There are several competing effects that control the evolution of satellite orbits. Because planetary bodies are not point masses but are finite objects, and because the force of gravity varies as a function of distance, two mass elements on a satellite that are at different distances away from the planet feel slightly different gravitational attractions towards the planet. This differential force gives rise to **tides**, and usually to a permanent bulge on the satellite. Both the planet and its satellite raise tides on the other that are important over long time periods for the evolution of the pair.

Tides raised on a satellite by its host planet do not change the satellite's orbital angular momentum, but they do dissipate its orbital energy. Equations (3.39) and (3.41) imply that the orbital energy E_{tot} is related to the semi-major axis a by:

$$E_{\text{tot}} = -\frac{Gm_1m_2}{2a} \,. \tag{3.49}$$

The thermal energy produced by tidal dissipation is drawn from the orbital energy. This makes E more negative and corresponds to a decrease in a and a corresponding increase in the mean motion n.

The rate of change of a satellite's mean motion that results from tidal dissipation has been found to have the following form:

$$\frac{dn}{dt} \propto \frac{1}{a^8}\frac{e^2 k}{Q} \,, \tag{3.50}$$

where the **Love number** k and the **tidal dissipation parameter** Q, which is a measure of the energy dissipated per cycle, take into account the geophysical properties that control the energy dissipation rate. The dependence of tidal dissipation on eccentricity may be understood as follows. Consider the common case of a satellite in synchronous rotation (*e.g.,* Io, Earth's Moon). In the case of a circular orbit, such synchronous rotation implies that the same side of the satellite always faces the host planet. However, if the satellite's orbit is eccentric, then by Kepler's 2nd Law the satellite moves faster through the part of its orbit nearest to the planet, and slower through the part of its orbit away from the planet. Meanwhile, the rotation rate remains constant throughout the orbit. This means that the satellite's orientation relative to the planet is not locked, but slips back and forth slightly with each revolution. This nodding of the satellite about its equilibrium configuration causes strains to develop inside the satellite that produce heat and a loss of orbital energy. There is also a smaller effect due to the periodic changing of the tidal strain as the satellite moves closer to and farther from the planet in its eccentric orbit.

While the generation of heat inside the satellite causes a loss of orbital energy, there is a negligible torque produced on the satellite, and hence a negligible effect on the orbital angular momentum L. If L is a constant, then (3.44) implies

$$a \propto \frac{1}{1-e^2} = 1 + e^2 + e^4 + \ldots \tag{3.51}$$

Thus, a decrease in a implies a decrease in e, and orbits tend to become more circular with time for processes that conserve orbital angular momentum and dissipate orbital energy. In the absence of other effects, tidal heating would tend to stop the evolution, because as e reduces to zero, (3.50) implies that dn/dt reduces to zero.

There are competing effects that can frustrate this tendency. A satellite raises a tidal bulge on its host planet that is not necessarily directly underneath the satellite. Io raises a tide on Jupiter that always leads Io in its orbit because Jupiter rotates faster than Io orbits. Unlike the previous effect that is due to tides on the satellite, this effect is due to tides on the planet, and it does change the satellite's orbital angular momentum. In this most common case where the planet rotates faster than the satellite orbits, some of the planet's spin angular momentum is transferred to the satellite's orbital angular momentum. The torque on the satellite from the planetary bulge pulls the satellite forward, adding orbital energy and causing the orbit to expand, in competition with the previous effect. Another important effect is the action of orbital resonances between satellites. An important resonance exists between Io, Europa, and Ganymede that continuously pumps up the eccentricity of Io and prevents it from damping to zero.

3.4.5 Laplace 4:2:1 Resonance

The mean-motion resonance between the orbits of Io, Europa, and Ganymede was first carefully studied by Laplace (*Mecanique Celeste, vol 4*, 1805) and is called the Laplace resonance. The physical characteristics of the Galilean satellites are shown in Table 3.1.

Table 3.1

	Io	**Europa**	**Ganymede**
n (°/day)	203.4890	101.3747	50.3176
e_{forced} $(2:1)$	0.0041	0.0101	0.0006
e_{free}	0.00001	0.0009	0.0015

Adapted from Yoder (1979).

Let the subscripts 1, 2, and 3 refer to Io, Europa, and Ganymede, respectively. Then

$$n_1 - 2n_2 = (203.4890) - (202.3494) = 0.7396\,,$$

$$n_2 - 2n_3 = (101.3747) - (100.6352) = 0.7395\,,$$

$$n_1 - 3n_2 + 2n_3 \approx 0\,. \tag{3.52}$$

Without the forcing caused by this three-way mean-motion resonance, Io would have an eccentricity of only $e_1 = 0.00001$, which would produce negligible tidal heating. *Orbital resonances tend to amplify perturbations and to increase eccentricities.* The combined effect of the complete 4:2:1 resonance results in $e_1 = 0.0043$. This would normally still be a small value, but not for Io because of the strong tides it experiences because of its close proximity to Jupiter. In the 2 March 79 issue of *Nature*, Peale, Cassen, and Reynolds published a paper entitled "Melting of Io by Tidal Dissipation." In this paper, which came out just days before the Voyager 1 encounter of Jupiter on 5 March 79, the authors predicted:

> The implications of the orbital resonances of the inner three Galilean satellites are profound for the thermal state of Io. These calculations suggest that Io might currently be the most intensely heated terrestrial-type body in the solar system. The surface of the type of body postulated here has not yet been directly observed, and although the morphology of such a surface cannot be predicted in any detail, one might speculate that widespread and recurrent surface volcanism would occur, leading to extensive differentiation and outgassing.

Voyager 1 did indeed find widespread volcanism on Io — so much volcanism that the images show no evidence of impact craters, implying a fast resurfacing rate. The prediction and swift confirmation of Io's volcanism is one of the success stories in theoretical planetary science.

A picture has emerged that explains how the 4:2:1 resonance of the Galilean satellites may have come about. The paper "How tidal heating in Io drives the galilean orbital resonance locks" by C. Yoder (1979) was the first to convincingly explain the situation, which we summarize below:

1. Suppose that Io were formed well inside the orbit of Europa about 4.6×10^9 years ago. Any initial free (unforced) eccentricity in Io's orbit would be quickly damped out by the tide Jupiter raises on Io. Only modest tidal heating of Io occurs.

2. Io's orbit would then expand because of the effects of the leading tide it raises on Jupiter. No significant tidal heating yet.

3. Europa's orbit would also expand, but because of Io's greater mass and smaller semi-major axis, Io would spiral out faster.

4. Io would eventually approach the 2:1 commensurability with Europa.

5. Io's forced eccentricity increases until it reaches the critical value ≈ 0.0026. Europa's limiting eccentricity is ≈ 0.0014. The significance of reaching these numbers is that thereafter a resonant interaction causes Europa's orbit to expand at exactly half of Io's orbit.

6. This stable state is maintained until Europa encounters the 2:1 commensurability with Ganymede. Instead of dissipation in Europa repelling Ganymede, it is found that Io must work even harder using a three-body resonance to transfer angular momentum from Europa to Ganymede's orbit.

7. The vanishing of the difference frequency describes the presently observed three-body resonance.

Current research continues to refine and challenge this picture, and the final story has undoubtably not yet been told. Whatever the cause of the 4:2:1 resonance, its effects on Io are profound, and make it one of the most bizarre and interesting satellites in the solar system.

3.5 Chaotic Orbits

We have seen how the 4:2:1 commensurability in the mean motions $n = 2\pi/\tau$ between Jupiter's three innermost Galilean satellites, the so-called Laplace resonance, continuously forces Io's eccentricity to be higher than it otherwise would be, ultimately causing enough tidal dissipation to give rise to active volcanism. Mean-motion resonances also play a dominant role in the physics of the asteroid belt, where **chaotic dynamics**, rather than tidal heating, is the primary agent of change.

3.5.1 Structure in the Asteroid Belt

The distribution of orbital elements is not random in the asteroid belt. There are families of asteroids that share similar a-e-i elements. There are also conspicuous gaps and clumps of asteroids at certain semi-major axes from the Sun. In 1867, Daniel Kirkwood first noted that large gaps in the asteroid belt are associated with mean-motion commensurabilities with Jupiter. The two largest of these **Kirkwood gaps** fall on the 3:1 and 2:1 commensurabilities. Conversely, at the 3:2 commensurability there is a clumping of asteroids, called the **Hilda asteroids**. There is also a clumping of asteroids, the **Trojan asteroids**, that travel in the 1:1 commensurability 60° ahead and 60° behind Jupiter in its orbit. The Trojan asteroids are understood to be caught in the stable (L4, L5) Lagrange points in Jupiter's orbit, but what is the explanation for the Kirkwood gaps and for the Hilda asteroids?

If there had been only gaps at the commensurabilities, then the simple explanation might be Kirkwood's original suggestion that the gaps are the result of asteroids being pulled out of the resonance zones by the periodic tugs from Jupiter. We now know that this picture is too naive. We showed previously that the 2-body central force problem has a formal solution based on the two integrals of motion, energy $E_{\rm tot}$ and angular momentum L. However, the fact that the 3-body problem does not have such a formal solution provides the first hint that the dynamics of the solar system as a whole is likely to be much richer than is encompassed by Kepler's laws.

In 1892, Poincaré made two important discoveries about the nature of solar system dynamics. First, he found that the assumption of regularity of the motion implies the existence of a complete set of independent integrals of motion. However, he showed that these integrals of motion generally do not persist when there are perturbations, such as the gravitational perturbations that each planet makes on the others. Consequently, most conservative systems do not possess all the integrals of motion necessary for a complete solution. Poincaré also discovered that the motions near unstable periodic orbits possess almost unimaginable complexity. This was an early indication of the nature of chaotic orbits.

Poincaré's discoveries were not fully appreciated at first. Fermi (1923) tried to prove, on the basis of Poincaré's proof of the nonexistence of analytical integrals, that conservative (Hamiltonian) systems were generally **ergodic**, meaning that the planets will eventually pass through every point inside the space that is constrained by the conservation of energy and angular momentum. The regularity of the known planetary orbits seems to argue against this idea, however, and Fermi was shown to be incorrect. The truth lies somewhere between Kepler's elliptical orbits and Fermi's ergodic orbits.

3.5.2 KAM Theory

A theorem outlined by Kolmogorov in 1954 and proven independently by both Arnol'd and Moser in 1961, now called the KAM theorem, states that if the perturbations are sufficiently small and smooth, integrable motion remains in large measure quasi-periodic. The value of the proof of this theorem is *not* that it provides any practical method for handling real perturbations — the KAM theorem only handles extremely small perturbations of relative magnitude less than about 10^{-48} — but rather, as a proof by construction it demonstrates that quasi-periodic motions can indeed persist under perturbations. Arnol'd (1961) extended the proof to show that for solar systems with planetary masses, eccentricities, and inclinations sufficiently small, motions remain in large measure quasi-periodic.

As it happens, our solar system falls outside the range of validity of Arnold's extension to the KAM theorem, and this opens up the possibility that nonperiodic orbits might exist.

Because of the analytical intractability of nonlinear systems, modern work in solar system dynamics relies on numerical experiments. The difficulty is that the solar system has a wide range of timescales, from the planetary revolutions that are measured in years, to the large orbital changes that result from the cumulative effect of thousands or millions of years of perturbations. To avoid numerical instability, a computer must take small enough timesteps to resolve the highest-frequency components in a system, but often the lowest-frequency components are the most interesting. For example, the well-understood Keplerian component of a planet's orbit has a frequency that is around 1,000 times higher than the slow changes that result from the gravitational perturbations of the other planets.

3.5.3 Early Studies of Chaotic Systems

Two early numerical experiments that proved to be seminal to our understanding of nonlinear dynamics were the simplified weather model of Lorenz (1963), which has become the classic example of a dissipative chaotic system, and the nonlinearly coupled harmonic oscillator problem of Hénon and Heiles (1964), which has become the textbook example of a conservative chaotic system. Both models show clearly that complicated behavior can result from simple nonlinear equations of motion.

In order to aid in the analysis of nonlinear dynamical systems, objective tools have been developed to help with the characterization of chaos. It is observed that two nearby chaotic trajectories will diverge exponentially with time. This property of sensitive dependence on initial conditions is quantified by the **Lyapunov exponent**. A second important tool that is useful for simple problems is to plot the intersection of an orbit with some plane in its phase space, like using stroboscopic photography. Such a plot is called a **surface of section**. The difference between a quasi-periodic, or regular, orbit and a chaotic orbit in a surface of section is immediately obvious to the eye. As each new point intersects the surface of section, a regular orbit fills out a smooth curve, even though the points do not collect next to each other from one orbit to the next, whereas a chaotic orbit forms a jumble of dots.

A new intuition has emerged from the study of chaotic systems. One expects that a dynamical system with more than one degree of freedom will nearly always give rise to chaotic behavior for some initial conditions, and regular behavior for others, with the two occurring side by side in phase space. One also expects that resonances will nearly always be associated with chaotic zones. This turns out to be the case for the solar system's Kirkwood gaps.

For a complete summary of chaotic dynamics applied other problems in solar system dynamics, like the tumbling of Saturn's satellite Hyperion, see the review paper by Wisdom (1987).

3.5.4 Chaos and the Kirkwood Gaps

If we wish to understand the Kirkwood gaps, we first need to explore the extent of the chaotic region around each commensurability between Jupiter and the asteroid belt. It is difficult to predict analytically the extent of a chaotic region. As noted above, numerical experiments can also be prohibitively time-consuming. The first numerical time integrations of the solar system used approximate equations that were averaged to remove the rapidly varying orbital terms while leaving intact the resonant and long-period, or **secular**, terms. Small regions in phase space of chaotic orbits were found in these experiments, but the basic conclusion was that chaotic behavior was not that important. This conclusion was based on numerical integrations that were run for only 10,000 years.

Hénon and Heiles noted in their original paper that an important feature of a surface of section generated by a Hamiltonian system is that it is area preserving. That is, each time a small chord of trajectories intersects the surface of section, the area encompassed by that chord remains a constant. Hénon and Heiles suggested that significant progress might be made by studying an iterated area-preserving map of the section onto itself. In fact, much of what is known about general Hamiltonian systems has come from the study of the properties of area-preserving maps, without making reference to the underlying differential equations. In practical applications this mapping approach is limited because the explicit form of the map for a given dynamical problem is not usually known. If the map can be found it often leads to a thousand-fold increase in computational speed.

In 1979, B.V. Chirikov introduced a method for approximating the area-preserving maps of certain problems. In 1982, J. Wisdom applied Chirikov's idea to the problem of asteroid orbits near the 3:1 commensurability. As with previous methods, the strategy was first to remove the high frequency terms associated with orbital periods while retaining the resonant and secular terms. Then, new high-frequency terms made up of simple delta functions were added back into the system to mimic the original terms. The new equations are integrated across and between the delta functions, yielding the sought-after map of the phase space onto itself. Numerical experiments that use the map can be integrated over time periods that are a thousand times longer than was previously possible. The price paid is that the new equations are not identical to the old equations, but one hopes that the topology of the new phase space accurately reflects that of the original problem.

3.5.5 Changes in Eccentricity

When Wisdom first applied the fast-mapping technique to the simplified 2-dimensional elliptic problem, he found that an asteroid can spend a 100,000 years or longer in a low eccentricity orbit and then suddenly take large excursions into an highly eccentric orbit. This unexpected result was first suspected to be an artifact of the approximate method, but subsequent numerical integrations of the complete differential equations verified that the behavior is real. In some runs the eccentricity jumps all tend to reach the same eccentricity, but occur at irregular intervals. The more typical behavior shows bursts of irregular high-eccentricity behavior interspersed with intervals of irregular, low-eccentricity behavior, with an occasional eccentricity spike.

In the surfaces of section plotted in Figs. 9-12 of Wisdom (1987), the eccentricity is just the radial distance from the origin. Examination of these figures makes the origin of the strange intermittent behavior of the eccentricity clear. There exists a chaotic zone that surrounds the origin $e = 0$, and this chaotic zone has a very narrow branch that extends to eccentricities near 0.3. A trajectory wanders in the chaotic zone and every now and then it enters the narrow bridge to large eccentricity. As expected there are unstable periodic orbits that stand at the gateways ushering the trajectories one way or the other. The remarkable similarity of the eccentricity jumps is explained by the narrowness of that part of the chaotic region that extends to high eccentricity. The eccentricity varies irregularly because the chaotic zone is large.

To perform a million-year reversible integration in the face of exponential divergence of neighboring points, one would need to carry *200 digits* of accuracy on the computer, which is difficult even with today's supercomputers. Therefore, only the qualitative results of a long time integration can be trusted. The large eccentricity jumps discovered by the fast-mapping approximate techniques have now been repeatedly reproduced in conventional numerical integrations. In addition, the surfaces of section for the problem are consistent with the character of the eccentricity jumps, and a semi-analytical treatment has been found that even reproduces the shape of the chaotic zone. Next we shall see how these eccentricity jumps can explain the 3:1 Kirkwood gap.

3.5.6 Eccentricity and the 3:1 Kirkwood Gap

For an asteroid orbiting near the 3:1 commensurability at $a = 2.5\,\mathrm{AU}$, any eccentricity e above 0.3 will put the asteroid in a Mars-crossing orbit. It turns out that all of the 3:1 chaotic trajectories cross the orbit of Mars. This is also true of the quasi-periodic librators. Therefore, asteroids near the 3:1 commensurability can be efficiently removed from the asteroid belt by collisions or close encounters with Mars. Comparison of the outer boundary of the chaotic zone with the actual distribution of observed asteroids shows remarkably good agreement (see Wisdom, 1987, Fig. 13).

The efficiency of this asteroid-removal mechanism depends on how long an asteroid remains in a chaotic orbit before it is removed by Mars. It is estimated that Mars alone could have cleared the 3:1 Kirkwood gap in the age of the solar system. An element of this calculation is the average time spent by the asteroid as a Mars crosser, which unfortunately is just the type of question that is dangerous to ask of the numerical experiments.

Asteroids with eccentricities above 0.6 in the 3:1 chaotic zone are of practical importance because they are in Earth-crossing orbits. In order to establish whether or not this is an important source of meteorites, several 500,000 year numerical integrations of the complete equations of motion were carried out. The first four integrations attempted were encouraging, but failed to produce an Earth-crossing asteroid. The fifth attempt finally did reveal an asteroidal trajectory that reached Earth-crossing eccentricities. Subsequent calculations have confirmed that this mechanism can partially account for the population of Earth-crossing asteroids.

One of the biggest puzzles in the study of meteorites and asteroids has been that the most common meteorite type, the ordinary chondrite, which represents 80% of all meteorites, was not associated with any known counterpart in the asteroid belt. However, in 1993, R. Binzel and coworkers discovered that 3628 Božněmcová, a 7 km asteroid that orbits just beyond the outer edge of the 3:1 Kirkwood gap, has a visible and near-infrared reflectance spectrum that is similar to L6 and LL6 ordinary chondrites. It is not clear why more such asteroids have not yet been identified.

3.5.7 The 2:1 Kirkwood Gap and the Hilda Asteroids

The 2:1 Kirkwood gap and the 3:2 Hilda asteroids are two features of the main asteroid belt that still need to be explained. Also needing explanation is the sharp decline in the number of asteroids beyond the 2:1 resonance. The most obvious explanation for the latter is that asteroids beyond the 2:1 region are so close to Jupiter that their orbits are dynamically unstable, but this idea has not held up to close scrutiny. In fact, current numerical experiments do not indicate that the region between 2:1 and 3:2 is particularly unstable or chaotic. Longer and more realistic future calculations may alter this view. The instability of the region beyond the 3:2 resonance may be understood in terms of the resonance overlap criterion. The 2:1 Kirkwood gap is near the outer boundary of the main belt, but it is not clear whether there is a causal relationship. To answer these questions we must explore the dynamics of the 2:1 and 3:2 resonances. These turn out to be more complicated than the 3:1 resonance. The 3:1 resonance is relatively simple because it is far from the other resonances, and because the higher-order perturbation terms are not as important. Chaotic behavior for the 2:1 and 3:2 resonances begins to appear only after the second-order perturbation terms are included, which makes these resonances more difficult to study in detail.

3.5.8 Near-Earth Asteroids

There are three classes of near-Earth asteroids (NEA's). Asteroids with orbits that cross Earth's orbit are called **Apollo asteroids**. Surprisingly, over half a dozen of the low inclination Apollo asteroids require less rocket fuel for a return mission than is required for the Moon. Those asteroids with orbits that cross Mars' orbit but not Earth's orbit are called **Amors**, and those that remain inside of Earth's orbit are called **Atens**. The total number of observed near-Earth asteroids is now over 200. Shoemaker *et al.* [1979] have estimated that the actual populations include about 100 Atens, 700 ± 300 Apollos, and 1000-2000 Amors.

The lifetime of a near-Earth asteroid is only about 10^7-10^8 years because of the high probability that it will encounter Earth or Mars. In comparison, most main-belt asteroids can remain in stable orbits for the age of the solar system. This means that there must be a continual resupply of near-Earth asteroids. There appears to be a diversity of types, which indicates that there is more than one source region.

3.5.9 The Stability of the Solar System

The stability of the planetary orbits themselves is one of the oldest problems in celestial mechanics. The motions of the giant planets appear to be regular in a 210-million year integration, although a bit more complicated than one might have expected. On the other hand, Pluto has an orbit that is extraordinarily complicated. Besides the well-understood mean-motion resonance that prevents the close approach of Pluto and Neptune even though their orbits overlap, Pluto is known to participate in at least two other resonances: its argument of perihelion librates about 90°, and the frequency of the circulation of its ascending node is nearly commensurate with one of the fundamental frequencies in the motion of the giant planets. The latter resonance gives rise to strong variations in the eccentricity with a 137-million year period. In 1988 Sussman and Wisdom showed that Pluto's orbit is chaotic. In 1989, J. Laskar went further and showed that all nine planetary orbits are chaotic. He used an approximate technique, but in 1993 G. Sussman and J. Wisdom confirmed this result with a series of 100-million-year full calculations of the solar system.

Problems

3-1. *Kepler's 3rd Law.*

a) Verify (3.40).

b) What is the error in the period for Jupiter with the planetary mass set to zero? What about for Earth?

c) Pluto's mass remained unknown until 1979, when its satellite Charon was discovered. Charon orbits Pluto every 6.387 days at a distance of 1.96×10^7m. What is the total mass of the Pluto-Charon system? How does this compare with the mass of the Moon? With the mass of Neptune's satellite Triton?

d) Using mutual occultations between Pluto and Charon, their radii have been found to be 1151 and 593 km, respectively. What is the mean density of Pluto plus Charon? Is this what you would expect?

3-2. *Motion near a Potential Minimum.*

Assume that the effective potential $\Phi_e(r)$ for the two-body central force problem,

$$\Phi_e(r) \equiv \frac{L^2}{2\mu r^2} + \Phi(r)\,, \tag{3.53}$$

has a local minimum for the circular orbit $r = r_0$.

a) Treat $\Phi_e(r)$ as an arbitrary function of r and approximate it with a Taylor-series expansion about r_0. Then, apply $dE/dt = 0$ to (3.17) to show that for E slightly larger than $\Phi_e(r_0)$, r will execute approximately harmonic oscillations about r_0 with frequency ω satisfying:

$$\omega^2 = \frac{1}{\mu}\left(\frac{d^2\Phi_e}{dr^2}(r_0)\right)\,. \tag{3.54}$$

b) Apply (3.54) to the case of Newton's law of gravitation, and compare with Kepler's 3rd Law.

3-3. *Kepler's Equation and Kepler's 1st and 2nd Laws.*

a) On a piece of graph paper, plot an ellipse with $a = 8\,\text{cm}$ and $e = 0.8$ using the polar formula:

$$r = \frac{a(1-e^2)}{1 + e\cos\phi}\,. \tag{3.55}$$

b) Make a table with column headings: M, E, and r. Under the M column enter $0°$, $30°$, $60°$, $90°$, $120°$, $150°$, and $180°$, and the corresponding angles in radians. For each M in radians use whatever means are at your disposal to determine the E that satisfies Kepler's Equation (3.48), then use (3.46) to determine r.

c) Draw a line from the focus at the origin to the ellipse that is the length of each r entry. Estimate the area of each section thus produced. How do your results compare with Kepler's 2nd Law?

3-4. *Fun with chaos.* Chaotic dynamics plays a major role in the explanation of such phenomena as the Kirkwood gaps in the asteroid belt. In 1976, R.M. May wrote a review article "Simple mathematical models with very complicated dynamics," *Nature* **261** 459–67, in which he showed that even the simplest possible nonlinear iterative process:

$$x_{n+1} = 4a\,x_n(1 - x_n), \quad 0 < a \leq 1, \quad 0 \leq x \leq 1\,, \tag{3.56}$$

has complicated, chaotic behavior. Pick a small value for a, and plot the parabola $f(x) = 4ax(1-x)$ and the unit-slope line $f(x) = x$ on the unit square, that is, for x and $f(x) \in [0, 1]$. To iterate, start at some initial x_0, draw a vertical line upwards to the parabola, then draw a horizontal line over to the unit-slope line, then back to the parabola, then back to the unit-slope line, etc. This is a graphical way to keep feeding the new x back into (3.56), and is a nice way to illustrate the transition to chaos. Now, try this again with a larger a. Explore the parameter a until you discover values where the iteration settles into a two-cycle, that is, where it repeats after two iterations. Increase a until the iterations never settle; this is the chaotic behavior. In between these two states you should be able to find a's that lead to a four-cycle, an eight-cycle, and beyond. When you find an a large enough for chaotic behavior, start two trajectories very close to each other, and watch how quickly they diverge. After you have made a few sketches, it is acceptable to continue with a computer.

3-5. *Tides and the Roche limit.* Consider two small identical spheres of uniform density ρ_m and radius δr. The spheres are just touching each other, and are at a distance d from a large planet with density ρ_M and radius R. By equating the gravitational force of attraction between the touching spheres and the differential gravitational force (the tidal force) of the planet on the spheres, show that the spheres will be pulled apart by the planet's tidal forces if they venture closer to the planet than:

$$d \approx 2.5 \left(\frac{\rho_M}{\rho_m}\right)^{1/3} R\,. \tag{3.57}$$

This is a good estimate of the planet's **Roche limit**. More accurate assumptions about the properties of the small masses only result in small changes to the coefficient of (3.57). Use (3.57) to estimate the Roche limit for Saturn, and compare with the positions of Saturn's rings and satellites.

References

Belton, M.J.S., R.A. West, & J. Rahe, 1989, *Time-variable phenomena in the Jovian System*, NASA SP-494.

Binzel, R.P., T. Gehrels & M.S. Matthews, 1989, *Asteroids II*, University of Arizona Press.

Burns, J.A. & M.S. Matthews, 1986, *Satellites*, University of Arizona Press.

Goldstein, H., 1980, *Classical mechanics*, 3rd ed., Addison-Wesley.

Peale, S.J., P. Cassen, and R.T. Reynolds, 1979, Melting of Io by tidal dissipation, *Science* **203**, 892–894.

Symon, K., 1971, *Mechanics*, 3rd ed., Addison-Wesley.

Wisdom, J., 1987, Chaotic dynamics in the solar system, *Icarus* **72**, 241–275.

Yoder, C.F., 1979, How tidal heating in Io drives the galilean orbital resonance locks, *Nature* **279**, 767–770.