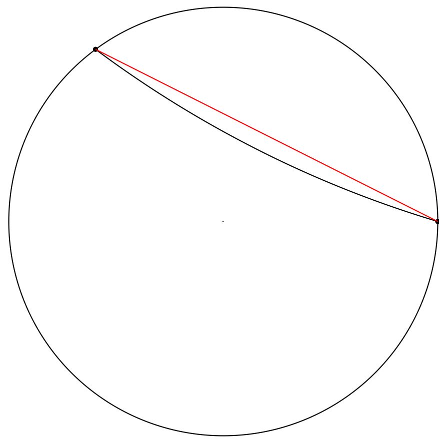
$$d\ell^2 = dr^2 + \frac{r^2}{1-\omega^2r^2/c^2}d\phi^2,$$

because their metersticks shrink in the tangential direction and it takes more of them to cover the distance we think of as $rd\phi$, though their metersticks agree with ours when measuring radial displacements.

The bugs will declare a curve to be a geodesic, or the shortest path between two points, if $\int d\ell$ is a minimum. Show that this requires that $r(\phi)$ satisfies

$$\frac{dr}{d\phi} = \pm\frac{r}{1-\omega^2r^2/c^2}\sqrt{\alpha^2r^2-1},$$

where α is a constant.



Straight lines to us and to the bugs, between the same two points.

2.12 Hamilton's Principle tells us that the motion of a particle is determined by the action functional being stationary under small variations of the path Γ in extended configuration space $(t,\vec{x})$. The unsymmetrical treatment of t and $\vec{x}(t)$ is not suitable for relativity, but we may still associate an action with each path, which we can parameterize with λ, so Γ is the trajectory $\lambda \to (t(\lambda),\vec{x}(\lambda))$.
In the general relativistic treatment of a particle's motion in a gravitational field, the action is given by $mc^2\Delta\tau$, where $\Delta\tau$ is the elapsed proper time, $\Delta\tau = \int d\tau$. But distances and time intervals are measured with a spatial varying metric $g_{\mu\nu}$, with μ and ν ranging from 0 to 3, with the zeroth component referring to time. The four components of extended configuration space are written x^μ, with a superscript rather than a subscript, and $x^0 = ct$. The gravitational field is described by the space-time dependence of the metric $g_{\mu\nu}(x^\rho)$. In this language, an infinitesimal element of the path of a particle corresponds to a proper time $d\tau = (1/c)\sqrt{\sum_{\mu\nu} g_{\mu\nu}dx^\mu dx^\nu}$, so

$$S = mc^2\Delta\tau = mc\int d\lambda\sqrt{\sum_{\mu\nu} g_{\mu\nu}(x^\rho)\frac{dx^\mu}{d\lambda}\frac{dx^\nu}{d\lambda}}.$$

(a) Find the four Lagrange equations which follow from varying $x^\rho(\lambda)$.

(b) Show that if we multiply these four equations by $\dot{x}^\rho$ and sum on ρ, we get an identity rather than a differential equation helping to determine the functions $x^\mu(\lambda)$. Explain this as a consequence of the fact that *any* path has a length unchanged by a reparameterization of the path, $\lambda \to \sigma(\lambda)$, $x'^\mu(\lambda) = x^\mu(\sigma(\lambda)$

(c) Using this freedom to choose λ to be τ, the proper time from the start of the path to the point in question, show that the equations of motion are

$$\frac{d^2x^\lambda}{d\tau^2} + \sum_{\rho\sigma}\Gamma^\lambda{}_{\rho\sigma}\frac{dx^\rho}{d\tau}\frac{dx^\sigma}{d\tau} = 0,$$

and find the expression for $\Gamma^\lambda{}_{\rho\sigma}$.

2.13 (a): Find the canonical momenta for a charged particle moving in an electromagnetic field and also under the influence of a non-electromagnetic force described by a potential $U(\vec{r})$.
(b): If the electromagnetic field is a constant magnetic field $\vec{B} = B_0\hat{e}_z$, with no electric field and with $U(\vec{r}) = 0$, what conserved quantities are there?

Chapter 3

Two Body Central Forces

Consider two particles of masses m_1 and m_2, with the only forces those of their mutual interaction, which we assume is given by a potential which is a function only of the distance between them, $U(|\vec{r}_1 - \vec{r}_2|)$. In a mathematical sense this is a very strong restriction, but it applies very nicely to many physical situations. The classical case is the motion of a planet around the Sun, ignoring the effects mentioned at the beginning of the book. But it also applies to electrostatic forces and to many effective representations of nonrelativistic interparticle forces.

3.1 Reduction to a one dimensional problem

Our original problem has six degrees of freedom, but because of the symmetries in the problem, many of these can be simply separated and solved for, reducing the problem to a mathematically equivalent problem of a single particle moving in one dimension. First we reduce it to a one-body problem, and then we reduce the dimensionality.

3.1.1 Reduction to a one-body problem

As there are no external forces, we expect the center of mass coordinate to be in uniform motion, and it behoves us to use

$$\vec{R} = \frac{m_1\vec{r}_1 + m_2\vec{r}_2}{m_1 + m_2}$$

as three of our generalized coordinates. For the other three, we first use the cartesian components of the relative coordinate

$$\vec{r} := \vec{r}_2 - \vec{r}_1,$$

although we will soon change to spherical coordinates for this vector. In terms of $\vec{R}$ and $\vec{r}$, the particle positions are

$$\vec{r}_1 = \vec{R} - \frac{m_2}{M}\vec{r}, \qquad \vec{r}_2 = \vec{R} + \frac{m_1}{M}\vec{r}, \qquad \text{where } M = m_1 + m_2.$$

The kinetic energy is

$$\begin{aligned} T &= \frac{1}{2}m_1\dot{r}_1^2 + \frac{1}{2}m_2\dot{r}_2^2 \\ &= \frac{1}{2}m_1\left(\dot{\vec{R}} - \frac{m_2}{M}\dot{\vec{r}}\right)^2 + \frac{1}{2}m_2\left(\dot{\vec{R}} + \frac{m_1}{M}\dot{\vec{r}}\right)^2 \\ &= \frac{1}{2}(m_1 + m_2)\dot{\vec{R}}^2 + \frac{1}{2}\frac{m_1 m_2}{M}\dot{\vec{r}}^2 \\ &= \frac{1}{2}M\dot{\vec{R}}^2 + \frac{1}{2}\mu\dot{\vec{r}}^2, \end{aligned}$$

where

$$\mu := \frac{m_1 m_2}{m_1 + m_2}$$

is called the **reduced mass**. Thus the kinetic energy is transformed to the form for two effective particles of mass M and μ, which is neither simpler nor more complicated than it was in the original variables.

For the potential energy, however, the new variables are to be preferred, for $U(\vec{r}_1 - \vec{r}_2) = U(\vec{r})$ is independent of $\vec{R}$, whose three components are therefore ignorable coordinates, and their conjugate momenta

$$\left(\vec{P}_{cm}\right)_i = \frac{\partial(T - U)}{\partial \dot{R}_i} = M\dot{R}_i$$

are conserved. This reduces half of the motion to triviality, leaving an effective one-body problem with $T = \frac{1}{2}\mu\dot{r}^2$, and the given potential $U(\vec{r})$.

We have not yet made use of the fact that U only depends on the *magnitude* of $\vec{r}$. In fact, the above reduction applies to any two-body system without external forces, as long as Newton's Third Law holds.

3.1.2 Reduction to one dimension

In the problem under discussion, however, there is the additional restriction that the potential depends only on the magnitude of $\vec{r}$, that is, on the distance between the two particles, and not on the direction of $\vec{r}$. Thus we now convert from cartesian to spherical coordinates (r, θ, ϕ) for $\vec{r}$. In terms of the cartesian coordinates (x, y, z)

$$\begin{array}{ll} r = (x^2+y^2+z^2)^{\frac{1}{2}} & x = r\sin\theta\cos\phi \\ \theta = \cos^{-1}(z/r) & y = r\sin\theta\sin\phi \\ \phi = \tan^{-1}(y/x) & z = r\cos\theta \end{array}$$

Plugging into the kinetic energy is messy but eventually reduces to a rather simple form

$$\begin{aligned} T &= \frac{1}{2}\mu\left[\dot{x}_1^2+\dot{x}_2^2+\dot{x}_3^2\right] \\ &= \frac{1}{2}\mu\Big[(\dot{r}\sin\theta\cos\phi+\dot{\theta}r\cos\theta\cos\phi-\dot{\phi}r\sin\theta\sin\phi)^2 \\ &\qquad +(\dot{r}\sin\theta\sin\phi+\dot{\theta}r\cos\theta\sin\phi+\dot{\phi}r\sin\theta\cos\phi)^2 \\ &\qquad +(\dot{r}\cos\theta-\dot{\theta}r\sin\theta)^2\Big] \\ &= \frac{1}{2}\mu\left[\dot{r}^2+r^2\dot{\theta}^2+r^2\sin^2\theta\dot{\phi}^2\right] \end{aligned} \tag{3.1}$$

Notice that in spherical coordinates T is a funtion of r and θ as well as $\dot{r}$, $\dot{\theta}$, and $\dot{\phi}$, but it is not a function of ϕ, which is therefore an ignorable coordinate, and

$$P_\phi = \frac{\partial L}{\partial \dot{\phi}} = \mu r^2\sin^2\theta\dot{\phi} = \text{ constant.}$$

Note that $r\sin\theta$ is the distance of the particle from the z-axis, so P_ϕ is just the z-component of the angular momentum, L_z. Of course all of $\vec{L} = \vec{r}\times\vec{p}$ is conserved, because in our effective one body problem there is no torque about the origin. Thus $\vec{L}$ is a constant[1], and the motion must remain in a plane perpendicular to $\vec{L}$ and passing through the origin, as a consequence of the fact that $\vec{r}\perp\vec{L}$. It simplifies things if we choose our coordinates so that $\vec{L}$ is in the z-direction. Then $\theta = \pi/2$, $\dot{\theta} = 0$, $L = \mu r^2\dot{\phi}$. The r equation of motion is then

$$\mu\ddot{r} - \mu r\dot{\phi}^2 + dU/dr = 0 = \mu\ddot{r} - \frac{L^2}{\mu r^3} + dU/dr.$$

This is the one-dimensional motion of body in an effective potential

$$U_{\rm eff}(r) = U(r) + \frac{L^2}{2\mu r^2}.$$

Thus we have reduced a two-body three-dimensional problem to one with a single degree of freedom, without any additional complication except the addition of a **centrifugal barrier** term $L^2/2\mu r^2$ to the potential.

Before we proceed, a comment may be useful in retrospect about the reduction in variables in going from the three dimensional one-body problem to a one dimensional problem. Here we reduced the phase space from six variables to two, in a problem which had four conserved quantities, $\vec{L}$ and H. But we have not yet used the conservation of H in this reduction, we have only used the three conserved quantities $\vec{L}$. Where have these dimensions gone? From $\vec{L}$ conservation, by choosing our axes with $\vec{L}\parallel z$, the two constraints $L_x = 0$ and $L_y = 0$ (with $L_z \neq 0$) do imply $z = p_z = 0$, thereby eliminating two of the coordinates of phase space. The conservation of L_z, however, is a consequence of an ignorable coordinate ϕ, with conserved conjugate momentum $P_\phi = L_z$. In this case, not only is the corresponding momentum restricted to a constant value, eliminating one dimension of variation in phase space, but the corresponding coordinate, ϕ, while not fixed, drops out of consideration because it does not appear in the remaining one dimensional problem. This is generally true for an ignorable coordinate — the corresponding momentum becomes a time-constant parameter, and the coordinate disappears from the remaining problem.

[1] If $\vec{L} = 0$, $\vec{p}$ and $\vec{r}$ are in the same direction, to which the motion is then confined. In this case it is more appropriate to use Cartesian coordinates with this direction as x, reducing the problem to a one-dimensional problem with potential $U(x) = U(r = |x|)$. In the rest of this chapter we assume $\vec{L} \neq 0$.

3.2 Integrating the motion

We can simplify the problem even more by using the one conservation law left, that of energy. Because the energy of the effective motion is a constant,

$$E = \frac{1}{2}\mu\dot{r}^2 + U_{\rm eff} = \text{ constant}$$

we can immediately solve for

$$\frac{dr}{dt} = \pm \left\{ \frac{2}{\mu}\left(E - U_{\rm eff}(r)\right) \right\}^{1/2}.$$

This can be inverted and integrated over r, to give

$$t = t_0 \pm \int \frac{dr}{\sqrt{2\left(E - U_{\rm eff}(r)\right)/\mu}}, \tag{3.2}$$

which is the inverse function of the solution to the radial motion problem $r(t)$. We can also find the orbit because

$$\frac{d\phi}{dr} = \frac{\dot\phi}{dr/dt} = \frac{L}{\mu r^2}\frac{dt}{dr}$$

so

$$\phi = \phi_0 \pm L \int_{r_0}^{r} \frac{dr}{r^2\sqrt{2\mu\left(E - U_{\rm eff}(r)\right)}}. \tag{3.3}$$

The sign ambiguity from the square root is only because r may be increasing or decreasing, but time, and usually ϕ/L, are always increasing.

Qualitative features of the motion are largely determined by the range over which the argument of the square root is positive, as for other values of r we would have imaginary velocities. Thus the motion is restricted to this allowed region. Unless $L = 0$ or the potential $U(r)$ is very strongly attractive for small r, the centrifugal barrier will dominate there, so $U_{\rm eff} \underset{r\to 0}{\longrightarrow} +\infty$, and there must be a smallest radius $r_p > 0$ for which $E \ge U_{\rm eff}$. Generically the force will not vanish there, so $E - U_{\rm eff} \approx c(r - r_p)$ for $r \approx r_p$, and the integrals in (3.2) and (3.3) are convergent. Thus an incoming orbit reaches $r = r_p$ at a finite time and finite angle, and the motion then continues with r increasing and the $\pm$ signs reversed. The radius r_p is called a **turning point** of the motion. If there is also a maximum value of r for which the velocity is real, it is also a turning point, and an outgoing orbit will reach this maximum and then r will start to decrease, confining the orbit to the allowed values of r.

If there are both minimum and maximum values, this interpretation of Eq. (3.3) gives ϕ as a multiple valued function of r, with an "inverse" $r(\phi)$ which is a periodic function of ϕ. But there is no particular reason for this period to be the geometrically natural periodicity 2π of ϕ, so that different values of r may be expected in successive passes through the same angle in the plane of the motion. There would need to be something very special about the attractive potential for the period to turn out to be just 2π, but indeed that is the case for Newtonian gravity.

We have reduced the problem of the motion to doing integrals. In general that is all we can do explicitly, but in some cases we can do the integral analytically, and two of these special cases are very important physically.

3.2.1 The Kepler problem

Consider first the force of Newtonian gravity, or equivalently the Coulomb attraction of unlike charged particles. The force $F(r) = -K/r^2$ has a potential

$$U(r) = -\frac{K}{r}.$$

Then the ϕ integral is

$$\begin{aligned} \phi &= \phi_0 \pm \int \frac{L}{\mu r^2} dr \left\{ \frac{2E}{\mu} + \frac{2K}{r} - \frac{L^2}{\mu^2 r^2} \right\}^{-1/2} \\ &= \phi_0 \pm \int \frac{du}{\sqrt{\gamma + \alpha u - u^2}} \end{aligned} \tag{3.4}$$

where we have made the variable substitution $u = 1/r$ which simplifies the form, and have introduced abbreviations $\gamma = 2\mu E/L^2$, $\alpha = 2K\mu^2/L^2$.

As $d\phi/dr$ must be real the motion will clearly be confined to regions for which the argument of the square root is nonnegative, and the motion in r will reverse at the turning points where the argument vanishes. The argument is clearly negative as $u \to \infty$, which is $r = 0$. We have assumed $L \neq 0$, so the angular momentum barrier dominates over the Coulomb attraction, and always prevents the particle from reaching the origin. Thus there is always at least one turning point, $u_{\max}$, corresponding to the minimum distance r_p. Then the argument of the square root must factor into $[-(u - u_{\max})(u - u_{\min})]$, although if $u_{\min}$ is negative it is not really the minimum u, which can never get past zero. The integral (3.4) can be done[2] with

[2]Of course it can also be done by looking in a good table of integrals. For example, see 2.261(c) of Gradshtein and Ryzhik[7].

the substitution $\sin^2\beta = (u_{\max} - u)/(u_{\max} - u_{\min})$. This shows $\phi = \phi_0 \pm 2\beta$, where ϕ_0 is the angle at $r = r_{\min}$, $u = u_{\max}$. Then

$$u \equiv \frac{1}{r} = A\cos(\phi - \phi_0) + B$$

where A and B are constants which could be followed from our sequence of substitutions, but are better evaluated in terms of the conserved quantities E and L directly. $\phi = \phi_0$ corresponds to the minimum r, $r = r_p$, the point of closest approach, or perigee[3], so $r_p^{-1} = A + B$, and $A > 0$. Let $\theta = \phi - \phi_0$ be the angle from this minimum, with the x axis along $\theta = 0$. Then

$$\frac{1}{r} = A\cos\theta + B = \frac{1}{r_p}\left(1 - \frac{e}{1+e}(1-\cos\theta)\right) = \frac{1}{r_p}\frac{1+e\cos\theta}{1+e}$$

where $e = A/B$.

What is this orbit? Clearly r_p just sets the scale of the whole orbit. From $r_p(1+e) = r + er\cos\theta = r + ex$, if we subtract ex and square, we get $r_p^2 + 2r_pe(r_p - x) + e^2(r_p - x)^2 = r^2 = x^2 + y^2$, which is clearly quadratic in x and y. It is therefore a conic section,

$$y^2 + (1-e^2)x^2 + 2e(1+e)xr_p - (1+e)^2r_p^2 = 0.$$

The nature of the curve depends on the coefficient of x^2. For

- $|e| < 1$, the coefficient is > 0, and we have an ellipse.
- $e = \pm 1$, the coefficient vanishes and $y^2 = ax + b$ is a parabola.
- $|e| > 1$, the coefficient is < 0, and we have a hyperbola.

All of these are possible motions. The bound orbits are ellipses, which describe planetary motion and also the motion of comets. But objects which have enough energy to escape from the sun, such as Voyager 2, are in hyperbolic orbit, or in the dividing case where the total energy is exactly zero, a parabolic orbit. Then as time goes to ∞, ϕ goes to a finite value, $\phi \to \pi$ for a parabola, or some constant less than π for a hyperbolic orbit.

[3]Perigee is the correct word if the heavier of the two is the Earth, perihelion if it is the sun, periastron for some other star. Pericenter is also used, but not as generally as it ought to be.

Let us return to the elliptic case. The closest approach, or **perigee**, is $r = r_p$, while the furthest apart the objects get is at $\theta = \pi, r = r_a = r_p(1+e)/(1-e)$, which is called the **apogee** or aphelion. e is the **eccentricity** of the ellipse. An ellipse is a circle stretched uniformly in one direction; the diameter in that direction becomes the **major axis** of the ellipse, while the perpendicular diameter becomes the **minor axis**.

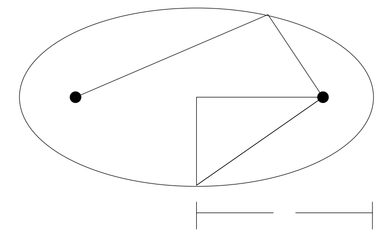
One half the length of the major axis is the **semi-major axis** and is denoted by a.

$$a = \frac{1}{2}\left(r_p + r_p\frac{1+e}{1-e}\right) = \frac{r_p}{1-e},$$

so

$$r_p = (1-e)a, \qquad r_a = (1+e)a.$$

Notice that the center of the ellipse is ea away from the Sun.



Properties of an ellipse. The large dots are the foci. The eccentricity is e and a is the semi-major axis.

Kepler tells us not only that the orbit is an ellipse, but also that the sun is at one focus. To verify that, note the other focus of an ellipse is symmetrically located, at $(-2ea, 0)$, and work out the sum of the distances of any point on the ellipse from the two foci. This will verify that $d + r = 2a$ is a constant, showing that the orbit is indeed an ellipse with the sun at one focus.

How are a and e related to the total energy E and the angular momentum L? At apogee and perigee, $dr/d\phi$ vanishes, and so does $\dot{r}$, so $E = U(r) + L^2/2\mu r^2 = -K/r + L^2/2\mu r^2$, which holds at $r = r_p = a(1-e)$ and at $r = r_a = a(1+e)$. Thus $Ea^2(1 \pm e)^2 + Ka(1 \pm e) - L^2/2\mu = 0$. These two equations are easily solved for a and e in terms of the constants of the motion E and L

$$a = -\frac{K}{2E}, \qquad e^2 = 1 + \frac{2EL^2}{\mu K^2}.$$

As expected for a bound orbit, we have found r as a periodic function of ϕ, but it is surprising that the period is the natural period 2π. In other words, as the planet makes its revolutions around the sun, its perihelion is always in the same direction. That didn't have to be the case — one could imagine that each time around, the minimum distance occurred at a slightly different (or very different) angle. Such an effect is called the **precession of the perihelion**. We will discuss this for nearly circular orbits in other potentials in section (3.2.2).

What about Kepler's Third Law? The area of a triange with $\vec{r}$ as one edge and the displacement during a small time interval $\delta\vec{r} = \vec{v}\delta t$ is $A = \frac{1}{2}|\vec{r}\times\vec{v}|\delta t = |\vec{r}\times\vec{p}|\delta t/2\mu$, so the area swept out per unit time is

$$\frac{dA}{dt} = \frac{L}{2\mu}.$$

which is constant. The area of an ellipse made by stretching a circle is stretched by the same amount, so A is π times the semimajor axis times the semiminor axis. The endpoint of the semiminor axis is a away from each focus, so it is $a\sqrt{1-e^2}$ from the center, and

$$\begin{aligned} A &= \pi a^2\sqrt{1-e^2} = \pi a^2\sqrt{1-\left(1+\frac{2EL^2}{\mu K^2}\right)} \\ &= \pi a^2 \frac{L}{K}\sqrt{\frac{-2E}{\mu}}. \end{aligned}$$

Recall that for bound orbits $E < 0$, so A is real. The period is just the area swept out in one revolution divided by the rate it is swept out, or

$$\begin{aligned} T &= \pi a^2 \frac{L}{K}\sqrt{\frac{-2E}{\mu}}\frac{2\mu}{L} \\ &= \frac{2\pi a^2}{K}\sqrt{-2\mu E} = \frac{\pi}{2}K(2\mu)^{1/2}(-E)^{-3/2} \qquad (3.5) \\ &= \frac{2\pi a^2}{K}\sqrt{\mu K/a} = 2\pi a^{3/2}(K)^{-1/2}\mu^{1/2}, \qquad (3.6) \end{aligned}$$

independent of L. The fact that T and a depend only on E and not on L is another fascinating manifestation of the very subtle symmetries of the Kepler/Coulomb problem.

3.2.2 Nearly Circular Orbits

For a general central potential we cannot find an analytic form for the motion, which involves solving the effective one-dimensional problem with $U_{\rm eff}(r) = U(r) + L^2/2\mu r^2$. If $U_{\rm eff}(r)$ has a minimum at $r = a$, one solution is certainly a circular orbit of radius a. The minimum requires $dU_{\rm eff}(r)/dr = 0 = -F(r) - L^2/\mu r^3$, so

$$F(a) = -\frac{L^2}{\mu a^3}.$$

We may also ask about trajectories which differ only slightly from this orbit, for which $|r-a|$ is small. Expanding $U_{\rm eff}(r)$ in a Taylor series about a,

$$U_{\rm eff}(r) = U_{\rm eff}(a) + \frac{1}{2}(r-a)^2 k,$$

where

$$\begin{aligned} k &= \left.\frac{d^2U_{\rm eff}}{dr^2}\right|_a \\ &= -\frac{dF}{dr} + \frac{3L^2}{\mu a^4} = -\left(\frac{dF}{dr}+\frac{3F}{a}\right). \end{aligned}$$

For $r = a$ to be a minimum and the nearly circular orbits to be stable, the second derivative and k must be positive, and therefore $F' + 3F/a < 0$. As always when we treat a problem as small deviations from a stable equilibrium[4] we have harmonic oscillator motion, with a period $T_{\rm osc} = 2\pi\sqrt{\mu/k}$.

As a simple class of examples, consider the case where the force law depends on r with a simple power, $F = -cr^n$. Then $k = (n+3)ca^{n-1}$, which is positive and the orbit stable only if $n > -3$. For gravity, $n = -2, c = K, k = K/a^3$, and

$$T_{\rm osc} = 2\pi\sqrt{\frac{\mu a^3}{K}}$$

agreeing with what we derived for the more general motion, not restricted to small deviations from circularity. But for more general n, we find

$$T_{\rm osc} = 2\pi\sqrt{\frac{\mu a^{1-n}}{c(n+3)}}.$$

[4]This statement has an exception if the second derivative vanishes, $k = 0$.

The period of revolution $T_{\rm rev}$ can be calculated for the circular orbit, as

$$L = \mu a^2 \dot\phi = \mu a^2 \frac{2\pi}{T_{\rm rev}} = \sqrt{\mu a^3 |F(a)|},$$

so

$$T_{\rm rev} = 2\pi \sqrt{\frac{\mu a}{|F(a)|}}$$

which for the power law case is

$$T_{\rm rev} = 2\pi \sqrt{\frac{\mu a^{1-n}}{c}}.$$

Thus the two periods $T_{\rm osc}$ and $T_{\rm rev}$ are not equal unless $n = -2$, as in the gravitational case. Let us define the **apsidal angle** ψ as the angle between an apogee and the next perigee. It is therefore $\psi = \pi T_{\rm osc}/T_{\rm rev} = \pi/\sqrt{3+n}$. For the gravitational case $\psi = \pi$, the apogee and perigee are on opposite sides of the orbit. For a two- or three-dimensional harmonic oscillator $F(r) = -kr$ we have $n = 1, \psi = \frac{1}{2}\pi$, and now an orbit contains two apogees and two perigees, and is again an ellipse, but now with the center-of-force at the center of the ellipse rather than at one focus.

Note that if ψ/π is not rational, the orbit never closes, while if $\psi/\pi = p/q$, the orbit will close after p revolutions, having reached q apogees and perigees. The orbit will then be closed, but unless $p = 1$ it will be self-intersecting. This exact closure is also only true in the small deviation approximation; more generally, Bertrand's Theorem states that only for the $n = -2$ and $n = 1$ cases are the generic orbits closed.

In the treatment of planetary motion, the precession of the perihelion is the angle though which the perihelion slowly moves, so it is $2\psi - 2\pi$ per orbit. We have seen that it is zero for the pure inverse force law. There is actually some precession of the planets, due mostly to perturbative effects of the other planets, but also in part due to corrections to Newtonian mechanics found from Einstein's theory of general relativity. In the late nineteenth century discrepancies in the precession of Mercury's orbit remained unexplained, and the resolution by Einstein was one of the important initial successes of general relativity.

3.3 The Laplace-Runge-Lenz Vector

The remarkable simplicity of the motion for the Kepler and harmonic oscillator central force problems is in each case connected with a hidden symmetry. We now explore this for the Kepler problem.

For any central force problem $\vec F = \dot{\vec p} = f(r)\hat e_r$ we have a conserved angular momentum $\vec L = \mu(\vec r \times \dot{\vec r})$, for $\dot{\vec L} = \mu \dot{\vec r} \times \dot{\vec r} + (f(r)/r)\vec r \times \vec r = 0$. The motion is therefore confined to a plane perpendicular to $\vec L$, and the vector $\vec p \times \vec L$ is always in the plane of motion, as are $\vec r$ and $\vec p$. Consider the evolution of $\vec p \times \vec L$ with time[5]

$$\begin{aligned} \frac{d}{dt}\left(\vec p \times \vec L\right) &= \dot{\vec p} \times \vec L = \vec F \times \vec L = \mu f(r) \hat e_r \times (\vec r \times \dot{\vec r}) \\ &= \mu f(r)\left(\vec r \hat e_r \cdot \dot{\vec r} - \dot{\vec r} \hat e_r \cdot \vec r\right) = \mu f(r)(\dot r \vec r - r \dot{\vec r}) \end{aligned}$$

On the other hand, the time variation of the unit vector $\hat e_r = \vec r/r$ is

$$\frac{d}{dt}\hat e_r = \frac{d}{dt}\frac{\vec r}{r} = \frac{\dot{\vec r}}{r} - \frac{\dot r \vec r}{r^2} = -\frac{\dot r \vec r - r\dot{\vec r}}{r^2}.$$

For the Kepler case, where $f(r) = -K/r^2$, these are proportional to each other with a constant ratio, so we can combine them to form a *conserved* quantity $\vec A = \vec p \times \vec L - \mu K \hat e_r$, called[6] the **Laplace-Runge-Lenz vector**, $d\vec A/dt = 0$.

While we have just found three conserved quantities in addition to the conserved energy and the three conserved components of $\vec L$, these cannot all be independent. Indeed we have already noted that $\vec A$ lies in the plane of motion and is perpendicular to $\vec L$, so $\vec A \cdot \vec L = 0$. If we dot $\vec A$ into the position vector,

$$\vec A \cdot \vec r = \vec r \cdot (\vec p \times (\vec r \times \vec p)) - \mu K r = (\vec r \times \vec p)^2 - \mu K r = L^2 - \mu K r,$$

so if θ is the angle between $\vec A$ and $\vec r$, we have $Ar\cos\theta + \mu K r = L^2$, or

$$\frac{1}{r} = \frac{\mu K}{L^2}\left(1 + \frac{A}{\mu K}\cos\theta\right),$$

[5]Some hints: $\vec A \times (\vec B \times \vec C) = \vec B(\vec A \cdot \vec C) - \vec C(\vec A \cdot \vec B)$, and $\hat e_r \cdot \dot{\vec r} = (1/r)\vec r \cdot \dot{\vec r} = (1/2r)d(r^2)/dt = \dot r$. The first equation, known as the **bac-cab** equation, is shown in Appendix A.1.

[6]by Goldstein, at least. While others often use only the last two names, Laplace clearly has priority.

which is an elegant way of deriving the formula we found previously by integration, with $A = \mu K e$. Note $\theta = 0$ is the perigee, so $\vec{A}$ is a constant vector pointing towards the perigee.

We also see that the magnitude of $\vec{A}$ is given in terms of e, which we have previously related to L and E, so $A^2 = \mu^2 K^2 + 2\mu E L^2$ is a further relation among the seven conserved quantities, showing that only five are independent. There could not be more than five independent conserved functions depending analytically on the six variables of phase space (for the relative motion only), for otherwise the point representing the system in phase space would be unable to move. In fact, the five independent conserved quantities on the six dimensional dimensional phase space confine a generic invariant set of states, or orbit, to a one dimensional subspace. For power laws other than $n = -2$ and $n = 1$, as the orbits do not close, they are dense in a two dimensional region of phase space, indicating that there cannot be more than four independent conserved analytic functions on phase space. So we see the connection between the existence of the conserved $\vec{A}$ in the Kepler case and the fact that the orbits are closed.

3.4 The virial theorem

Consider a system of particles and the quantity $G = \sum_i \vec{p}_i \cdot \vec{r}_i$. Then the rate at which this changes is

$$\frac{dG}{dt} = \sum \vec{F}_i \cdot \vec{r}_i + 2T.$$

If the system returns to a region in phase space where it had been, after some time, G returns to what it was, and the average value of dG/dt vanishes,

$$\left\langle \frac{dG}{dt} \right\rangle = \left\langle \sum \vec{F}_i \cdot \vec{r}_i \right\rangle + 2\left\langle T \right\rangle = 0.$$

This average will also be zero if the region stays in some bounded part of phase space for which G can only take bounded values, and the averaging time is taken to infinity. This is appropriate for a system in thermal equilibrium, for example.

Consider a gas of particles which interact only with the fixed walls of the container, so that the force acts only on the surface, and the sum becomes an integral over $d\vec{F} = -p d\vec{A}$, where p is the uniform pressure and $d\vec{A}$ is

an outward pointing vector representing a small piece of the surface of the volume. Then

$$\left\langle \sum \vec{F}_i \cdot \vec{r}_i \right\rangle = -\int_{\delta V} p\vec{r} \cdot d\vec{A} = -p\int_V \nabla \cdot \vec{r} dV = -3pV$$

so $\langle 2T \rangle = 3pV$. In thermodynamics we have the equipartition theorem which states that $\langle T \rangle = \frac{3}{2} N k_B \tau$, where N is the number of particles, k_B is Boltzmann's constant and τ the temperature, so $pV = N k_B \tau$.

A very different application occurs for a power law central force between pairs of particles, say for a potential $U(\vec{r}_i, \vec{r}_j) = c|\vec{r}_i - \vec{r}_j|^{n+1}$. Then this action and reaction contribute $\vec{F}_{ij} \cdot \vec{r}_j + \vec{F}_{ji} \cdot \vec{r}_i = \vec{F}_{ji} \cdot (\vec{r}_i - \vec{r}_j) = -(n+1)c|\vec{r}_i - \vec{r}_j|^{n+1} = -(n+1)U(\vec{r}_i, \vec{r}_j)$. So summing over all the particles and using $\langle 2T \rangle = -\langle \sum \vec{F} \cdot \vec{r} \rangle$, we have

$$\langle T \rangle = \frac{n+1}{2} \langle U \rangle.$$

For Kepler, $n = -2$, so $\langle T \rangle = -\frac{1}{2}\langle U \rangle = -\langle T + U \rangle = -E$ must hold for closed orbits or for large systems of particles which remain bound and uncollapsed. It is not true, of course, for unbound systems which have $E > 0$.

The fact that the average value of the kinetic energy in a bound system gives a measure of the potential energy is the basis of the measurements of the missing mass, or dark matter, in galaxies and in clusters of galaxies. This remains a useful tool despite the fact that a multiparticle gravitationally bound system can generally throw off some particles by bringing others closer together, so that, strictly speaking, G does not return to its original value or remain bounded.

3.5 Rutherford Scattering

We have discussed the $1/r$ potential in terms of Newtonian gravity, but of course it is equally applicable to Coulomb's law of electrostatic forces. The force between nonrelativistic charges Q and q is given[7] by

$$\vec{F} = \frac{1}{4\pi\epsilon_0} \frac{Qq}{r^3} \vec{r},$$

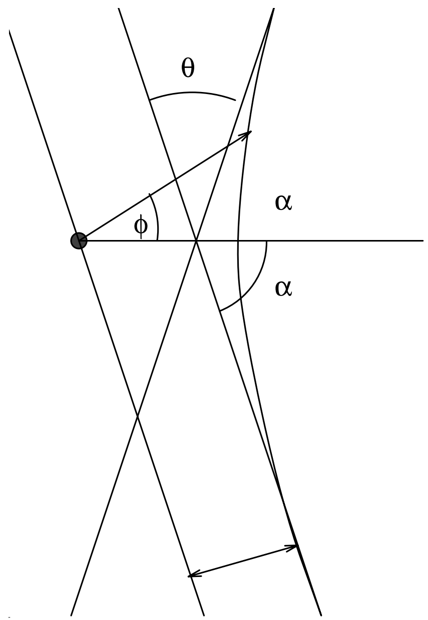
[7] Here we use S. I. or rationalized MKS units. For Gaussian units drop the $4\pi\epsilon_0$, or for Heaviside-Lorentz units drop only the ϵ_0.

and the potential energy is $U(r) = -K/r$ with $K = -Qq/4\pi\epsilon_0$.

Unlike gravity, the force is not always attractive ($K > 0$), and for like sign charges we have $K < 0$, and therefore U and the total energy are always positive, and there are no bound motions. Whatever the relative signs, we are going to consider scattering here, and therefore positive energy solutions with the initial state of finite speed v_0 and $r \to \infty$. Thus the relative motion is a hyperbola, with

$$\begin{aligned} r &= r_p \frac{1+e}{1+e\cos\phi} \\ e &= \pm\sqrt{1+\frac{2EL^2}{\mu K^2}}. \end{aligned}$$

This starts and ends with $r \to \infty$, at $\phi \to \pm\alpha = \pm\cos^{-1}(-1/e)$, and the angle θ through which the velocity changes is called the **scattering angle**. For simplicity we will consider the repulsive case, with $e < 0$ so that $\alpha < \pi/2$.



Rutherford scattering. An α particle approaches a heavy nucleus with an impact parameter b, scattering through an angle θ. The cross sectional area $d\sigma$ of the incident beam is scattered through angles $\in [\theta, \theta + d\theta]$.

We see that $\theta = \pi - 2\alpha$, so

$$\tan\frac{\theta}{2} = \cot\alpha = \frac{\cos\alpha}{\sqrt{1-\cos^2\alpha}} = \frac{|e|^{-1}}{\sqrt{1-|e|^{-2}}} = \frac{1}{\sqrt{e^2-1}} = \sqrt{\frac{\mu K^2}{2EL^2}}.$$

We have $K = Qq/4\pi\epsilon_0$. We need to evaluate E and L. At $r = \infty$, $U \to 0$, $E = \frac{1}{2}\mu v_0^2$, $L = \mu b v_0$, where b is the **impact parameter**, the distance by which the asymptotic line of the initial motion misses the scattering center. Thus

$$\tan\frac{\theta}{2} = |K|\sqrt{\frac{\mu}{\mu v_0^2(\mu b v_0)^2}} = \frac{|K|}{\mu b v_0^2}. \qquad (3.7)$$

The scattering angle therefore depends on b, the perpendicular displacement from the axis parallel to the beam through the nucleus. Particles passing through a given area will be scattered through a given angle, with a fixed angle θ corresponding to a circle centered on the axis, having radius $b(\theta)$ given by 3.7. The area of the beam $d\sigma$ in an annular ring of impact parameters $\in [b, b+db]$ is $d\sigma = 2\pi b|db|$. To relate db to $d\theta$, we differentiate the scattering equation for fixed v_0,

$$\frac{1}{2}\sec^2\frac{\theta}{2}d\theta = \frac{-K}{\mu v_0^2 b^2}db,$$

$$\begin{aligned} \frac{d\sigma}{d\theta} &= 2\pi b \frac{\mu v_0^2 b^2}{2K\cos^2(\theta/2)} = \frac{\pi \mu v_0^2 b^3}{K\cos^2(\theta/2)} \\ &= \frac{\pi\mu v_0^2}{K\cos^2(\theta/2)}\left(\frac{K}{\mu v_0^2}\right)^3\left(\frac{\cos\theta/2}{\sin\theta/2}\right)^3 = \pi\left(\frac{K}{\mu v_0^2}\right)^2\frac{\cos\theta/2}{\sin^3\theta/2} \\ &= \frac{\pi}{2}\left(\frac{K}{\mu v_0^2}\right)^2\frac{\sin\theta}{\sin^4\theta/2}. \end{aligned}$$

(The last expression is useful because $\sin\theta d\theta$ is the "natural measure" for θ, in the sense that integrating over volume in spherical coordinates is $d^3V = r^2 dr \sin\theta d\theta d\phi$.)

How do we measure $d\sigma/d\theta$? There is a beam of N particles shot at random impact parameters onto a foil with n scattering centers per unit area, and we confine the beam to an area A. Each particle will be significantly scattered only by the scattering center to which it comes closest, if the foil is thin enough. The number of incident particles per unit area is N/A, and the number of scatterers being bombarded is nA, so the number which get scattered through an angle $\in [\theta, \theta + d\theta]$ is

$$\frac{N}{A} \times nA \times \frac{d\sigma}{d\theta}d\theta = Nn\frac{d\sigma}{d\theta}d\theta.$$

We have used the cylindrical symmetry of this problem to ignore the ϕ dependance of the scattering. More generally, the scattering would not be uniform in ϕ, so that the area of beam scattered into a given region of (θ,ϕ) would be

$$d\sigma = \frac{d\sigma}{d\Omega}\sin\theta d\theta d\phi,$$

where $d\sigma/d\Omega$ is called the **differential cross section**. For Rutherford scattering we have

$$\frac{d\sigma}{d\Omega} = \frac{1}{4}\left(\frac{K}{\mu v_0^2}\right)^2 \csc^4\frac{\theta}{2}.$$

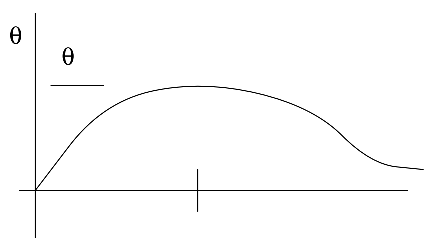
Scattering in other potentials

We see that the cross section depends on the angle through which the incident particle is scattered for a given impact parameter. In Rutherford scattering θ increases monotonically as b decreases, which is possible only because the force is "hard", and a particle aimed right at the center will turn around rather than plowing through. This was a surprize to Rutherford, for the concurrent model of the nucleus, Thompson's plum pudding model, had the nuclear charge spread out over some atomic-sized spherical region, and the Coulomb force would have decreased once the alpha particle entered this region. So sufficiently energetic alpha particles aimed at the center should have passed through undeflected instead of scattered backwards. In fact, of course, the nucleus does have a finite size, and this is still true, but at a much smaller distance, and therefore a much larger energy.

If the scattering angle $\theta(b)$ does run smoothly from 0 at $b = 0$ to 0 at $b \to \infty$, as shown, then there is an extremal value for which $d\theta/db|_{b_0} = 0$, and for $\theta < \theta(b_0)$, $d\sigma/d\theta$ can get contributions from several different b's,

$$\frac{d\sigma}{d\Omega} = \sum_i \frac{b_i}{\sin\theta}\left.\frac{db}{d\theta}\right|_i.$$

It also means that the cross section becomes infinite as $\theta \to \theta(b_0)$, and vanishes above that value of θ. This effect is known as **rainbow scattering**, and is the cause of rainbows, because the scattering for a given color light off a water droplet is very strongly peaked at the maximum angle of scattering.



Another unusual effect occurs when $\theta(b)$ becomes 0 or π for some nonzero value of b, with $db/d\theta$ finite. Then $d\sigma/d\Omega$ blows up due to the $\sin\theta$ in the denominator, even though the integral $\int(d\sigma/d\Omega)\sin\theta d\theta d\phi$ is perfectly finite.

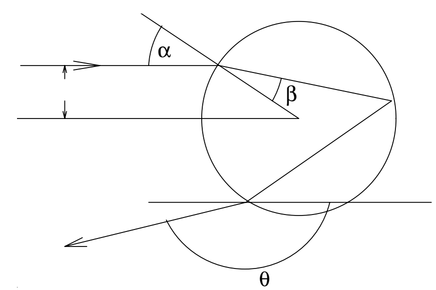
This effect is called **glory scattering**, and can be seen around the shadow of a plane on the clouds below.

Exercises

3.1 A space ship is in circular orbit at radius R and speed v_1, with the period of revolution τ_1. The crew wishes to go to planet X, which is in a circular orbit of radius $2R$, and to revolve around the Sun staying near planet X. They propose to do this by firing two blasts, one putting them in an orbit with perigee R and apogee $2R$, and the second, when near X, to change their velocity so they will have the same speed as X.

- (a) By how much must the first blast change their velocity? Express your answer in terms of v_1.
- (b) How long will it take until they reach the apogee? Express your answer in terms of τ_1
- (c) By how much must the second blast change their speed? Will they need to slow down or speed up, relative to the sun.

3.2 Consider a spherical droplet of water in the sunlight. A ray of light with impact parameter b is refracted, so by Snell's Law $n\sin\beta = \sin\alpha$. It is then internally reflected once and refracted again on the way out.
(a) Express the scattering angle θ in terms of α and β.
(b) Find the scattering cross section $d\sigma/d\Omega$ as a function of θ, α and β (which is implicitly a function of θ from (a) and Snell's Law).
(c) The smallest value of θ is called the rainbow scattering angle. Why? Find it numerically to first order in δ if the index of refraction is $n = 1.333 + \delta$
(d) The visual spectrum runs from violet, where $n = 1.343$, to red, where $n = 1.331$. Find the angular radius of the rainbow's circle, and the angular width of the rainbow, and tell whether the red or blue is on the outside.



One way light can scatter from a spherical raindrop.