Dynamics and Relativity

Stephen Siklos

Lent term 2011

Hand-outs and examples sheets, which I will give out in lectures, are available from my web site

www.damtp.cam.ac.uk/user/stcs/dynamics.html

Lecture notes, which I will not give out, are also available on my web site: they will appear after each lecture.

Chapter 0

Preliminaries

0.1 Course contents

Section 0: Preliminaries

0.1 Course plan
0.2 Some vector calculus.

Section 1: Newtonian dynamics: basic concepts

1.1 Newton's Laws
1.2 Dimensional Analysis

Section 2: Forces

2.1 Potentials
2.2 Conserved quantities (including angular momentum)
2.3 Friction
2.4 The Lorenz force
2.5 Gravitational force
2.6 Escape velocity

Section 3: Orbits

3.1 Motion in a plane
3.2 Orbits in a central force: the u-θ equation
3.3 Closed orbits
3.4 Stability of circular orbits
3.5 Orbits in an inverse square force
3.6 Rutherford scattering
3.7 Kepler's Laws
3.8 Orbits under more general forces

Section 4: Rotating Frames

4.1 Angular velocity
4.2 Rotating axes
4.3 Coriolis force
4.4 Centrifugal force

Section 5: Systems of particles

5.1 Equations of motion
5.2 Variable mass problems
5.3 Two-body problem
6.4 Moments of inertia
5.5 Inertia tensor
5.6 Motion of a rigid body

Section 6: Special Relativity

6.1 Concepts and laws
6.2 Space-time diagrams
6.3 Lorentz transformations
6.4 Length contraction and time dilation
6.5 Velocity composition
6.6 Proper time and the Minkowski metric
6.7 Four-vectors
6.8 Dynamical examples
6.9 *Doppler effect (if time allows)*

0.2 The course

It might seem odd to combine two seeming very disparate theories — Newtonian dynamics and special relativity — in one course. However, there are many ideas common to the two theories and thinking about the fundamental ideas behind Newtonian dynamics can light up one's understanding of relativity (and vice versa). It is very striking that the problems that Newton was wrestling with in the seventeenth century are exactly those occupying Einstein 220 years later.

Many students will have seen quite a lot of Newtonian dynamics at school and in the past have complained that there is little new for them in the course. The same students also complained that they understood the material better at school. Of course, what this means is that they could do the questions at school without really understanding what they were doing and, believing that they didn't need to work at the course, then found that they couldn't catch up. To try to prevent this happening, I have avoided setting questions on examples sheets that look like souped up A-level or STEP questions.

Many other students will have seen very little Newtonian dynamics at school and will find this course quite hard to start with. I advise these students to stick at it. Very soon, they should get to grips with what is after all mainly mathematics.

For the benefit of both sets of students, I have provided printed lecture notes. In these notes, I have put in more explanation and mathematical steps than is possible (or desirable) in large lectures, and I have tried also to include some deeper discussions and ideas beyond the course for those who really have covered some of the material. I will not distribute these notes in lectures: they can be seen on my web site `www.damtp.cam.ac.uk/user/stcs/dynamics.html` (which is presumably what you are now looking at).

0.3 Some vector calculus

Much of this course is formulated in terms of vectors. No doubt the following lemma will also be proved in the vector calculus course, but it is important enough to be proved twice.[1]

Lemma Let $\mathbf{u}(t)$ be a time-dependent vector[2] with components u_i $(i = 1,\ 2, 3)$ with respect to a set of fixed Cartesian axes. Then

$$\frac{d\mathbf{u}}{dt} = \left(\frac{du_1}{dt}, \frac{du_2}{dt}, \frac{du_3}{dt}\right).$$

Note This just says that vectors can be differentiated component by component, in Cartesians. Of course, we assume that the axes are not time-dependent (moving) and in Chapter 4 we deal with a case when this lemma does not hold: rotating Cartesian axes. Similarly, if the axes are not Cartesian, the lemma may not hold.[3]

Proof We can't make progress with this without knowing what differentiation of a vector means.

[1] Make the most of this: it is probably the last lemma-proof-corollary in the course.

[2] To be more precise, we want $\mathbf{u}$ to be a differentiable function $\mathbb{R} \to \mathbb{R}^3$.

[3] For example, in spherical polar coordinates and axes, the position vector of the particle is $(r, 0, 0)$, i.e. $r\widehat{\mathbf{r}}$, (NOT of course (r, θ, phi)), but the velocity is $(\dot r, r\dot\theta, r\sin\theta\dot\phi)$.

Taking the usual definition in terms of limits, we have

$$\frac{d\mathbf{u}}{dt} \equiv \lim_{h\to 0} \frac{\mathbf{u}(t+h)-\mathbf{u}(t)}{h} \qquad \text{(definition of differentiation)}$$

$$= \lim_{h\to 0} \frac{\begin{pmatrix} u_1(t+h) \\ u_2(t+h) \\ u_3(t+h) \end{pmatrix} - \begin{pmatrix} u_1(t) \\ u_2(t) \\ u_3(t) \end{pmatrix}}{h} \qquad \text{(same, in Cartesian components)}$$

$$= \lim_{h\to 0} \frac{\begin{pmatrix} u_1(t+h)-u_1(t) \\ u_2(t+h)-u_2(t) \\ u_3(t+h)-u_3(t) \end{pmatrix}}{h} \qquad \text{(usual rules for adding vectors)}$$

$$= \lim_{h\to 0} \begin{pmatrix} \dfrac{u_1(t+h)-u_1(t)}{h} \\ \dfrac{u_2(t+h)-u_2(t)}{h} \\ \dfrac{u_3(t+h)-u_3(t)}{h} \end{pmatrix} \qquad \text{(usual rule for multiplying vectors by scalars)}$$

$$\equiv \begin{pmatrix} \dfrac{du_1}{dt} \\ \dfrac{du_2}{dt} \\ \dfrac{du_3}{dt} \end{pmatrix} \qquad \text{(definition of differentiation)}$$

which is the required result.

An alternative approach might have been to write

$$\mathbf{u} = u_i \mathbf{e}_i \qquad (\mathbf{e}_i \cdot \mathbf{e}_j = \delta_{ij}, \ \text{summation convention applies})$$

and then differentiate to obtain the result immediately:

$$\begin{aligned} \frac{d\mathbf{u}}{dt} &= \frac{d(u_i\mathbf{e}_i)}{dt} \\ &= \frac{du_i}{dt}\mathbf{e}_i + \frac{d\mathbf{e}_i}{dt}u_i \\ &= \frac{du_i}{dt}\mathbf{e}_i \end{aligned}$$

(since the cartesian axis vectors $\mathbf{e}_i$ are fixed).

If the axes are *not* fixed, this result would not hold. For example, in plane polar coordinates, $\mathbf{r} = r\mathbf{e}_r$, and

$$\frac{d\mathbf{r}}{dt} = \frac{dr}{dt}\mathbf{e}_r + r\frac{d\mathbf{e}_r}{dt} = \frac{dr}{dt}\mathbf{e}_r + r\frac{d\theta}{dt}\mathbf{e}_\theta. \tag{1}$$

The second term can be obtained by converting to cartesian axes so that $\mathbf{e}_r = (\cos\theta, \sin\theta)$. Then differentiating with respect to t, using the chain rule, gives

$$\frac{d\mathbf{e}_r}{dt} = \frac{d\theta}{dt}\frac{d\mathbf{e}_r}{d\theta} = \dot\theta(-\cos\theta, \sin\theta) = \dot\theta\mathbf{e}_\theta.$$

The last term in (1) represents the compenent of velocity (if t is time) tangent to the circle $r =$ constant, as will be discussed in section 4.

Adopting this alternative approach assumes that we know how to differentiate the sum of vectors and the product of a scalar (u_i) and a vector ($\mathbf{e}_i$). We would have to start with a couple of lemmas:

$$\frac{d}{dt}(\mathbf{u}+\mathbf{v}) = \frac{d\mathbf{u}}{dt} + \frac{d\mathbf{v}}{dt}$$

and

$$\frac{d(\lambda\mathbf{v})}{dt} = \lambda\frac{d\mathbf{v}}{dt} + \frac{d\lambda}{dt}\mathbf{v}$$

where $\mathbf{u}$ and $\mathbf{v}$ are arbitrary t-dependent vectors and λ is and arbitrary t-dependent scalar. These are both easily proved from the definition of a derivative, applied to vectors.[4] For example:

$$\begin{aligned}\frac{d(\lambda\mathbf{v})}{dt} &= \lim_{h\to 0}\frac{\lambda(t+h)\mathbf{v}(t+h)-\lambda(t)\mathbf{v}(t)}{h}\\ &= \lim_{h\to 0}\frac{\lambda(t+h)\big(\mathbf{v}(t+h)-\mathbf{v}(t)\big)}{h} + \lim_{h\to 0}\frac{\mathbf{v}(t)\big(\lambda(t+h)-\lambda(t)\big)}{h}\\ &=\lambda\frac{d\mathbf{v}}{dt}+\frac{d\lambda}{dt}\mathbf{v}\end{aligned}$$

Corollary The derivatives of scalar and vector products obey the Leibnitz (product) rule.
Proof For scalar products, we have

$$\begin{aligned}\frac{d(\mathbf{u}\cdot\mathbf{v})}{dt} = \frac{d(u_i v_j \delta_{ij})}{dt} &= \frac{d(u_i v_i)}{dt} && \text{(using suffix notation and summation convention)}\\ &= \frac{du_i}{dt}v_i + u_i\frac{dv_i}{dt} && \text{(using Leibnitz for each term in the summation)}\\ &= \frac{d\mathbf{u}}{dt}\cdot\mathbf{v}+\mathbf{u}\cdot\frac{d\mathbf{v}}{dt} && \text{(using the above Lemma)}\end{aligned}$$

For vector products,

$$\frac{d(\mathbf{u}\times\mathbf{v})}{dt} = \frac{d\mathbf{u}}{dt}\times\mathbf{v} \;+\; \mathbf{u}\times\frac{d\mathbf{v}}{dt}.$$

This can be proved in the same way as for the scalar product, except the vector product has ϵ_{ijk} rather than δ_{ij}; this does not affect the proof, because the compenents of ϵ_{ijk} are constants, being 0 or ± 1.

[4] But then we might ask what exactly we mean by a vector that is a function of a parameter. It is like a set of Babushka dolls: everytime you go a layer deeper, another even deeper layer appears.

Chapter 1

Basic concepts

1.1 Newton's laws of motion

1.1.1 Statement of the laws

The laws governing the whole of this course except for the section on Special Relativity are Newton's three laws, so it seems right to bang them down on the table at the very beginning of the course.[1]

N1 Every particle remains at rest, or moves with constant speed in a straight line, unless acted upon by a force.

N2 The rate of change of momentum of a particle is equal (in magnitude and direction) to the force on the particle.

N3 To every action on a particle, there is an equal and opposite reaction.

In what follows, we will have a critical look at Newton's laws, and we will find that they are not nearly as straightforward as they might appear. The reason for doing this is not really to understand Newtonian dynamics; for most Newtonian purposes, we can get by without it. The reason is that we need a good understanding of the basis of Newtonian dynamics as context for special relativity, for which the same issues arise.

1.1.2 Newton's laws and inertial frames

Why, you may ask, did Newton feel the need to write down a first law that, on the face of it, is simply a special case (zero force) of his second law? According to Newton's biographer R.S. Westfall (*Never at Rest*), Newton came to his first law some 17 years after formulating his second law, which makes this question all the more in need of an answer. It is surprising how few people ask themselves this question; they should, because the answer is of fundamental importance.

One answer is that the first law is so counterintuitive that Newton thought an explicit statement was necessary; after all, we never see bodies moving with constant velocity *unless* there is a force acting (to balance air resistance, for example), and the idea that a force must act to maintain motion goes back to the Greeks. This may have been in Newton's mind but, if so, it seems strange that he should have taken so long to decide to write it down.

A more plausible explanation comes from asking another question, again one that occurs to few people: *when* does Newton's *second* law apply? After all, it certainly does not hold in a rotating axes.[2] The obvious answer is that Newton's second law holds in non-accelerating frames of

[1] We are used to the idea that Newton's laws are of groundbreaking and fundamental importance; less well known is that idea that his language was also groundbreaking. For example, in setting out his 'law' of gravity, Newton provided one relatively simple rule which applied equally and independently to every particle of matter in the universe and which by itself explained the motions of the planets and the comets, the fall of objects to the earth, and the tides. Newton called this a 'law', a word previously used in a legal context. Kepler's previously formulated 'laws' were not referred to as laws until after the time of Newton.

[2] As we will see in section 4, the appropriate equation is not $m\mathbf{a} = \mathbf{F}$ (mass times acceleration equals force) but

$$m\mathbf{a} = \mathbf{F} - m\ \ 2\boldsymbol{\omega} \times \mathbf{v} + \boldsymbol{\omega} \times (\boldsymbol{\omega} \times \mathbf{r}) + \dot{\boldsymbol{\omega}} \times \mathbf{r}$$

where $\boldsymbol{\omega}$ is the angular velocity vector of the rotating frame (see Chapter 4 for an explanation of the other quantities in this formula).

reference,[3] but what does that mean? Accelerating relative to what? Newton and others had been tussling with this last question for many years and the conventional answer was 'relative to *absolute space*'. But what it absolute space? Again there was a conventional answer: a frame determined by distant stars. The idea of absolute space was debated over and over again in the subsequent centuries and generally found wanting. Instead, Newton's first law can be thought of as setting the scene for the second law:[4] it defines a set of frames of reference (those in which the law holds); it is exactly in these frames that the second law holds.

Accordingly, we define an *inertial frame* to be a frame in which the statement N1 of Newton's first law holds; and we modify Newton's second law N2 by the insertion, at the beginning, of the words 'In any inertial frame'.

1.1.3 Forces and Newton's second law

What is meant in N2 by the 'force on the particle'? First thoughts suggest this is unproblematic: it is obvious when a force acts, what direction it acts in and what magnitude it has. On second thoughts, though, it is not quite so obvious. For example, the question of whether centrifugal force is 'real' and whether it acts inwards or outwards has bothered generations of students learning dynamics.

For theoretical physicists, there is no problem. They set up a theory (e.g. Newtonian dynamics or general relativity) and they say what forces act (e.g. the Newtonian gravitational force between two particles or the electromagnetic force between charged particles); there can then be no argument about what is or is not a force.

But for an ordinary person, a force is anything that has a tendency to make the particle move, so that an ordinary person in a car going round a corner feels a tendency to move outward which he or she calls the 'centrifugal force'. Its strength is determined in terms of what the person feels (G-force in an accelerating rocket, for example).[5]

There is no point in arguing about which point of view is 'right': it is just a matter of semantics. However, the ordinary person's point of view makes Newton's second law circular (tautological) and we theoretical physicists must therefore eschew it.

We therefore take the view that the force term in N2 is determined by, and explicitly given by, the physical model.

1.1.4 Newton's second law and varying mass problems

Often, Newton's second law is stated in terms of the acceleration ('Force equals mass times acceleration'), instead of the rate of change of momentum of a body. This is perfectly acceptable, because the mass of a particle, in Newtonian dynamics, is constant and the rate of change of momentum is exactly mass times acceleration:

$$\frac{d(m\mathbf{v})}{dt} = m\frac{d\mathbf{v}}{dt} = m\mathbf{a}\,.$$

However, later in the course, we will consider two situations where the mass is not constant but in which it is nevertheless possible to apply Newton's second law, provided we use the momentum formulation: in special relativity, where (slightly loosely speaking) the mass varies with speed; and for an open system of particles, which particles may enter or leave (a rocket, say).

1.1.5 Newton's third law

In a sense this is the most successful of the three laws: there is no known situation in which it does not apply. However, except in the case of two particles (or bodies) in contact which we consider in this course, the application may not be straightforward: for example, in the case of particles that are not in contact and are subject to forces that do not act along the line joining the two particles.[6]

[3] A *frame of reference* or *frame* for short is just a set of axes with respect to which the position and orientation of a body can be measured.

[4] Whether he intended this is not clear and really only of interest to historians of science.

[5] This is the point of view expressed in the well-known cartoon http://xkcd.com/123/ where the assumption is that a real force will appear in N2 as a result of a coordinate change and this force will crush Mr Bond. The physicists point of view is that Mr Bond would be crushed by the physical forces retaining Mr Bond. (The last line of the cartoon is a quotation from the 1964 film 'Goldfinger'. Bond is strapped to a table and just about to be cut in half by a laser. Bond: 'Do you expect me to talk?'. Goldfinger: 'No, Mr Bond, I expect you to die'.)

[6] I am thinking in particular of the case of charged particles in an electromagnetic field. One component of the force between them is proportional to $\mathbf{r}\times(\mathbf{r}\times\mathbf{a})$, where $\mathbf{r}$ is the position vector of one charged particle with respect to the other and $\mathbf{a}$ is the acceleration. This result might be proved in the Part II course called Electrodynamics.

1.1.6 Scope of Newton's laws

Newton's laws are not 'correct', in the sense that they will always give the right answer to any problem in dynamics: they work perfectly in idealised situations, and to a good approximation in more general circumstances, but require modification when these circumstances become too extreme. The circumstances in which they work well are as follows.

1. Particle-like bodies. Often, the word 'particle' in Newton's laws is replaced by 'body'. A particle is a point-like body with no internal structure, and this is strictly what Newton had in mind, though context is everything: for example, a planet is obviously not a particle but can be regarded as a particle for the purpose of working out its orbit because the size of the body is very small compared with the other length scales in the problem.[7]

   However, Newton's laws can be used to investigate the motion of composite bodies, such as rockets and solid spheres, provided all internal forces are taken into account, and this is the topic of chapter 5.

2. Small velocities. To be specific, we need $|\mathbf{v}|/c \ll 1$, where c is the speed of light, otherwise the effects of special relativity come into play.

3. Small masses. To be specific, we need $GM/Rc^2 \ll 1$, where G is Newton's gravitational constant and R is a typical length scale; otherwise general relativistic effects come into play.[8]

4. Massive particles.[9] Massless particles, such as photons, do not obey the Newton's laws despite the fact that they can be thought of as particles with energy and momentum.[10]

5. Quantum effects unimportant. This is a bit harder to quantify. For a particle moving in a potential (e.g. an electron orbiting the nucleus of an atom), quantum effects would be important if $pL \approx h$, where p is the momentum of the particle, L is the length scale over which the potential changes and h is Planck's constant (which is about 6.626068×10^{-34} m^2 kg/s — i.e. small). We therefore need momenta and length scales to be 'not too small'.

6. This should really be a footnote. It has fairly recently been suggested that N2 is not accurate for very small forces. The impetus behind this suggestion was the discovery that spiral galaxies do not rotate as they should: they appear to rotate rigidly, whereas one would expect the speed of the stars most distant from the centre to decrease as the reciprocal of distance (because the speed is given, according to N2, by $GM/r^2 = v^2/r$ where M is the mass of the central concentration and the right hand side is the centrifugal acceleration — see chapter 3 for an explanation of this equation). This is exactly what happens for planets in our solar system. According to MOND, or MOdified Newtonian Dynamics, the right hand side of N2 should, for very small forces only, be quadratic (not linear) in acceleration which, as is easily seen, gives constant rotational velocities. An alternative theory, which does not involve modifying N2, is that there is additional gravitating matter, called dark matter, in galaxies.

1.1.7 Absolute time

The idea of *absolute time* is central to Newtonian dynamics. To quote Newton: 'Absolute true and mathematical time, of itself, and from its own nature flows equably without regard to anything external, and by another name is called duration'. We are allowed to choose the origin of time and measure in hours instead of minutes but, having first agreed these conventions, the watches of all observers would show the same time; there would be no argument about whether two events occur simultaneously.

Newton's rather comfortable views on the nature of time are abandoned in special relativity: this is a (maybe *the*) fundamental difference between the two theories.

Clearly, this force does not act along the direction of the vector joining the two particles.

[7]The distance between the Sun and Earth, for example, and the distance over which the gravitational forces vary significantly (double in magnitude, say).

[8]The Schwarzschild radius of a body of mass M is $2GM/c^2$; this is the radius of a spherically symmetric black hole of mass M. On this length scale, the curvature of space-time would significantly effect the dynamics.

[9]In physics-speak, massive just means 'with mass', i.e. not massless, rather than 'as big as bus'.

[10]Laplace predicted the existence of black holes by assuming that photons obey the laws of Newtonian dynamics. The idea is that for a very heavy star, photons would not be able to escape (i.e. would not have the escape velocity — see section 2.6. Such 'black holes' would anyway not correspond to objects that we call black holes, for example because there would be no fixed event horizon.

1.1.8 Galilean transformations

A *Galilean transformation* is a transformation between two frames of reference (i.e. sets of coordinate axes) that preserves 'absolute space' and absolute time; where by absolute space I mean the modern interpretation in terms of inertial frames (i.e. frames in which Newton's first law holds).

Galilean transformations relate frames in which Newtonian physics applies, i.e. in which Newton's laws hold and any consequences (such as conservation of energy and momentum) apply.

Any Galilean transformation is a combination of the following transformations[11]:

- translations: $t \to t + t_0$, $\quad \mathbf{x} \to \mathbf{x} + \mathbf{a}$, where t_0 and $\mathbf{a}$ are constant (i.e. the same for all t and $\mathbf{x}$);
- rotations and reflections: $t \to t$, $\quad \mathbf{x} \to R\mathbf{x}$, where R is a constant matrix with[12] $R^T R = I$;
- Galilean boost: $t \to t$, $\quad \mathbf{x} \to \mathbf{x} + \mathbf{v}t$, where $\mathbf{v}$ is a constant vector.

We will be particularly interested in the third of these transformations. The effect of such a transformation is to give the axes a constant velocity with respect to the original axes. If you google 'Galilean transformation', on most of the sites only this one transformation will be given. We can use the rotational Galilean transformation to align the x-axis with the velocity vector $\mathbf{v}$, so that the transformation becomes

$$\begin{aligned} t' &= t \\ x' &= x + vt \\ y' &= y \\ z' &= z \end{aligned} \tag{1.1}$$

Galilean transformations form a group, as indeed they should: we would certainly expect each of the group axioms to hold. For example, if g is a Galilean transformation, then g^{-1} (getting back to the first frame) must also be a Galilean transformation. We can represent the group in terms of matrices acting on the coordinates of space-time:

$$\begin{pmatrix} t' \\ x' \\ y' \\ z' \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ v_1 & R_{11} & R_{12} & R_{13} \\ v_2 & R_{21} & R_{22} & R_{23} \\ v_3 & R_{31} & R_{32} & R_{33} \end{pmatrix} \begin{pmatrix} t \\ x \\ y \\ z \end{pmatrix} + \begin{pmatrix} t_0 \\ a_1 \\ a_2 \\ a_3 \end{pmatrix} \tag{1.2}$$

If we ignore the translations, the group properties can be very easily deduced from the rules of matrix multiplication.

It can be shown that Galilean transformations are the only transformations consistent with Newton's first law, but I don't think we should spend time worrying about this sort of thing. But in case you are interested, the following example shows how the argument might go.

1.1.9 Example: transformations that preserve Newtons's first law

For simplicity we restrict ourselves to the case of just one spatial dimension. We will also assume that time is absolute, and so unchanged in the transformation between frames.

Let the coordinate systems (x, t) and (x', t') of any two frames of reference S and S' be related by

$$x' = f(x, t), \qquad t' = t \tag{1.3}$$

for some function f. A particle follows a trajectory

$$x = x(t) \tag{1.4}$$

in S and its trajectory in S' is $x' = x'(t')$. We have

$$\begin{aligned} x' &= f(x(t), t) && \text{(from (1.3) and (1.4))} \\ &= f(x(t'), t') && \text{(since } t = t') \\ &\equiv x'(t'). && \text{((by definition))} \end{aligned}$$

[11]Scalings of the form $t \to kt$, $\mathbf{x} \to \lambda\mathbf{x}$ correspond to changes of units and are not included.

[12]Here, R is a 3×3 matrix, the superscript T denotes the transpose and I is the unit matrix.

By definition, the velocity v in S is $\dfrac{dx}{dt}$ and the velocity v' in S' is $\dfrac{dx'}{dt'}$. These velocities are related as follows:

$$\begin{aligned} v' &\equiv \frac{dx'}{dt'} && \text{(by definition)}\\ &= \frac{dx}{dt'}\frac{\partial f}{\partial x} + \frac{\partial f}{\partial t'}. && \text{(using the chain rule on } f(x(t'),t'))\\ &= v\frac{\partial f}{\partial x} + \frac{\partial f}{\partial t}. && \text{(again, since } t = t') \end{aligned}$$

Now suppose that S and S' are both inertial frames (so that Newton's first law holds in both frames), and that no force acts on the particle. Then both velocities must be constant (but not necessarily the same). That means $v\dfrac{\partial f}{\partial x} + \dfrac{\partial f}{\partial t}$ must be constant (independent of both x and t) for any value of v. Since f cannot depend on v, this can only be the case if both $\dfrac{\partial f}{\partial x}$ and $\dfrac{\partial f}{\partial t}$ are constant. This implies that $f(x,t)$ is of the form

$$f(t,x) = ax + vt + c,$$

where a and v are constants. This is the most general Galilean transformation of x in one space dimension (compare with the truncated version of (1.2)).

Note that in this example we have adopted the usual applied mathematicians' convention of using x to denote both a coordinate and a function, as in $x = x(t)$. This is convenient, but can be confusing: for example, x in the expression $\frac{dx}{dt}\frac{\partial f}{\partial x}$ means two different things: we are differentiating with respect to a coordinate x and differentiating a function $x(t)$.

We could avoid this. We would start by giving the trajectories in S and S' as $x = g(t)$ and $x' = h(t')$. The velocities in the two frames are $g'(t)$ and $h'(t')$, where prime denotes the derivative function. Now using the transformation $x' = f(x,t)$ gives

$$h(t') = f(g(t'),t').$$

Differentiating this with respect to t' using the chain rule gives

$$v'(t') = v(t')\frac{\partial f}{\partial x} + \frac{\partial f}{\partial t}$$

where the partial derivatives are to be evaluated at $x = g(t')$ and $t = t'$. Then we proceed as before.

This is all fine and very clear; but in a slightly more complicated problem, we soon forget what g, h, etc were supposed to represent; and if there are many more variables we soon run out of letters. But you can adopt whichever approach you feel comfortable with.

End of example

1.1.10 The principle of Galilean Relativity

Galileo considered (in 1632!) a thought experiment performed below decks on a ship on a completely calm sea. He decided that the laws of physics should be the same whether or not the ship was moving.[13] In our terminology, the Principle of Relativity states that the laws of physics are the same in all inertial frames. This principle applies in Newtonian physics, in Special Relativity and in General Relativity, the only difference being the groups of transformations that relate inertial frames in the different theories.

[13] This is how Galileo put it. *Shut yourself up with some friend in the main cabin below decks on some large ship, and have with you there some flies, butterflies, and other small flying animals. Have a large bowl of water with some fish in it; hang up a bottle that empties drop by drop into a wide vessel beneath it. With the ship standing still, observe carefully how the little animals fly with equal speed to all sides of the cabin. The fish swim indifferently in all directions; the drops fall into the vessel beneath; and, in throwing something to your friend, you need throw it no more strongly in one direction than another, the distances being equal; jumping with your feet together, you pass equal spaces in every direction. When you have observed all these things carefully (though doubtless when the ship is standing still everything must happen in this way), have the ship proceed with any speed you like, so long as the motion is uniform and not fluctuating this way and that. You will discover not the least change in all the effects named, nor could you tell from any of them whether the ship was moving or standing still.*

1.2 Dimensional analysis

1.2.1 Dimensions

Problems in dynamics (and all other areas of theoretical physics) involve quantities which are not pure numbers such as 2 or π, but have to be expressed in terms of some basic units, such as metres or kilograms. These quantities are said to have *dimensions* whereas pure numbers are said to be *dimensionless*. It is conventional to denote the dimensions of a quantity Q by [Q].

The basic dimensions for a problem in Newtonian dynamics are length (L), mass (M) and time (T). If the problem involves electric field another basic dimension is needed, which could be charge or current, but we will concentrate here on the simple case where there are just the three basic dimensions.[14] All other quantities arising in the problem have dimensions that can be expressed in terms of these basic dimensions. For example:

[Area] $=\mathrm{L}^2$,
[Speed] $= \mathrm{LT}^{-1}$ ('miles per hour')
[Force] $= \mathrm{MLT}^{-2}$ ('mass times acceleration')
[Newton's gravitational constant, G] $= \mathrm{M}^{-1}\mathrm{L}^3\ \mathrm{T}^{-2}$ (e.g. from the formula $F = \frac{Gm_1m_2}{r^2}$ — see section 2.4)

Of course, these dimensions correspond to the units in which the physical quantities are measured: square metres; metres per second; Newtons (= kilogram metres per second squared); cubic metres per kilogram per second squared).[15]

All equations in the problem must be dimensionally consistent: it makes no sense to add quantities with different dimensions. You cannot, for example, equate a quantity with dimensions of mass to a quantity with dimensions of length; even if they happened to be numerically equal, this would not be the case if the units were changed (from kilograms to grams, say).

This rather obvious observation has a trivial, though useful consequence: you can check the dimensions of each term in your equations to see if you have made any mistakes.[16] Sometimes, you don't know immediately what the dimensions of a quantity are, but you can always go back to the definition of that quantity which must reveal its dimensions.

1.2.2 Example: checking dimensions

Suppose x and y have dimensions of length. Then the equation

$$y = x^2 + e^x$$

can easily be seen to be wrong: some algebra must have gone awry. It is wrong for two reasons. First, the left-hand side has dimensions of length but the first term on the right-hand side has dimensions of length^2 (area) which does not make sense. Second, the exponential itself is meaningless, as follows immediately from the series expansion:

$$e^x \equiv 1 + x + \frac{1}{2!}x^2 + \cdots \tag{1.5}$$

which involves adding quantities with different dimensions. In order to make sense, the argument of the exponential has to be dimensionless, in which case the exponential is dimensionless.

End of example

1.2.3 Bridgman's theorem

By contrast, considerations of dimensional consistency can lead to deep insights into a given problem and indeed even to quantitative results when the only alternative route would involve solving intractable equations. This process is called *dimensional analysis.*

[14]Similarly, any other field — magnetic, or strong nuclear, for example — requires the introduction of at least one additional dimension, often generically called a charge. Thus mass could be thought of as gravitational charge.

[15]In the SI system of units, there are seven basic physical quantities: length L (unit: meter), mass M (unit: kilogram), time T (unit: second), electric current I (unit: ampere), temperature Θ (unit: kelvin), amount of substance N (unit: mole), and the luminous intensity, I_v (unit: candela). All other physical quantities are called derived quantities, as they are defined in terms of the basic quantities by means of defining equations.

[16]And this is a good habit to get into.

Before elaborating on the process, we need a simple and at first sight rather surprising preliminary result, namely that the dimensions of any physical quantity Q are expressible as products of powers of the basic dimensions, i.e.

$$[Q] = \mathrm{L}^\alpha \mathrm{M}^\beta \mathrm{T}^\gamma \ldots . \tag{1.6}$$

It certainly holds for the familiar examples (area, speed, force, G) given above.

To understand why this rule should hold, one has only to try to think of an alternative. How about an exponential? Clearly, this will not do; we cannot have e^L as already explained in equation (1.5). The same argument rules out all the familiar functions, such as trigonometric functions, hyperbolic functions and logs. In fact, suppose

$$[Q] = f(\mathrm{L}, \mathrm{M}, \mathrm{T}).$$

and suppose we can expand f in a Taylor series in L:

$$[Q] = \sum_{n=0}^{\infty} a_n L^n$$

where the coefficients a_n are independent of L. Then, since we are not allowed to add different powers of L, only one of the a_n can be non-zero. We can then expand successively in powers of M and T to obtain the required power law.

This argument does not quite prove the power law rule, because the Taylor series about $L = 0$ might not exist: for example, for $f(\mathrm{L}, \mathrm{M}, \mathrm{T}) = MLT^{-2}$ there is problem at $T = 0$. Similarly, it may be that the powers (i.e. indices) involved are fractional, in which case no Taylor series about the origin exists.

To fix the deficiency in this proof would take us a long way from our track.[17]

Equation (1.6) is a basic form of Bridgman's theorem. However, his theorem is normally expressed in a different and more useful way. We identify in any problem a set of *base quantities*. These quantities are not themselves dimensions, but the number of base quantities in the set is the same as the number of dimensions in the problem. In the case of Newtonian Dynamics, we need three base quantities; call them q_1, q_2 and q_3. We further require that each of L, M and T can be expressed in terms of $[q_1]$, $[q_2]$ and $[q_3]$.[18] Then Bridgman's theorem states that any physical quantity Q in the problem can be expressed in the form

$$Q = C q_1^\alpha q_2^\beta q_3^\gamma \tag{1.7}$$

where C is a dimensionless quantity. This is of course consistent with the previous result (1.6). It can be proved by means of scaling arguments, but again that would be too far off-piste for this discussion.

Armed with Bridgman's theorem, we can use dimensional analysis to obtain information about a problem without having to derive a complete solution. As we will see, considerable physical intuition may be required to obtain anything useful. It is best to study a simple example before discussing the general case.

1.2.4 The simple pendulum by dimensional analysis

Suppose we want to find out how the angular frequency ω of a simple pendulum varies when the length of the pendulum varies. We assume that the amplitude of the oscillations is small. Let the length of the pendulum be ℓ and the mass of the bob be m. The only other relevant quantity is the acceleration due to gravity, g. The three quantities m, ℓ and g are the base quantities for this problem. By Bridgman's theorem (1.7), ω must be expressible in terms of products of powers of these base quantities

$$\omega = C m^\alpha \ell^\beta g^\gamma, \tag{1.8}$$

where C is a dimensionless constant. The dimensions of the quantities in this equation are

$$[\omega] = \mathrm{T}^{-1}; \quad [m] = \mathrm{M}; \quad [\ell] = \mathrm{L}; \quad [g] = \mathrm{LT}^{-2}.$$

[17] There are various proofs, all involving scaling arguments. Roughly, the dimension $f(\mathrm{L}, \mathrm{M}, \mathrm{T})$ should mean the same dimension as the dimension $f(\alpha\mathrm{L}, \mathrm{M}, \mathrm{T})$, where α is any dimensionless number. Thus $f(\mathrm{L}, \mathrm{M}, \mathrm{T}) = kf(\alpha\mathrm{L}, \mathrm{M}, \mathrm{T})$, where k is a dimensionless number depending on α. This in turn implies that $f(\mathrm{L}, \mathrm{M}, \mathrm{T})$ has power-law dependence on L.

[18] We wouldn't for example choose q_1 and q_2 both to be lengths.

Note that M, L and T can be written in terms of $[m]$, $[\ell]$ and $[g]$, as required for Bridgman's theorem.

For equation (1.8) to make sense, we need

$$\mathrm{T}^{-1} = \mathrm{M}^{\alpha}\mathrm{L}^{\beta}(\mathrm{LT}^{-2})^{\gamma},$$

so, equating powers of the independent basic dimensions L, M and T respectively,

$$\begin{aligned} 0 &= \beta + \gamma, \\ 0 &= \alpha \\ -1 &= -2\gamma. \end{aligned}$$

Thus

$$\omega = C\ell^{-\frac{1}{2}}g^{\frac{1}{2}}. \tag{1.9}$$

This gives us the result we were seeking: the frequency varies as the inverse square root of the length.[19]

We cannot obtain C by dimensional analysis so we cannot go further than this. We could obtain C by experiment, or by solving the equations of motion.

You are probably impressed by the ease with which we obtained this excellent result: no differential equations to solve, no hard algebra. But we cheated! It was stated blithely that the only relevant quantities are mass of bob, length of pendulum and g; how could we know that the initial conditions play no role? In effect, we assumed that the frequency is independent of the amplitude. This is certainly true for very small oscillations because the motion is then simple harmonic and we have a good understanding of SHM gained by solving the equations — cheating!

So let us start again and proceed a little more carefully. First, does it even make sense talk about the frequency; i.e., is the motion periodic? It would not be periodic if there were friction in the suspension or if air resistance were a factor. But since there is no dissipation, our intuition tells us that, in this simple case, the motion is periodic.[20]

Now let us consider what other quantities could appear on the right side of equation (1.8). The pendulum itself has no more attributes (besides length and mass) and we are assuming there are no more external influences (besides gravity). The only additional quantities must therefore be related to initial displacement and angular speed, θ_0 and $\dot{\theta}_0$. Here, our intuition plays a further role. Clearly, although θ_0 and $\dot{\theta}_0$ can be varied independently, the subsequent motion is not in one to one correspondence with the sets of initial conditions: different initial conditions can give rise to the same subsequent motion. An increase in θ_0 could be traded off against a decrease of $\dot{\theta}_0$. We may surmise that it is just the total initial energy, E, rather than its individual components, that is needed to determine the motion.

We can still apply Bridgman's theorem, but with less success. As before, we choose m, ℓ and g as our base quantities, though now this choice is not forced: we could equally well have chosen m, ℓ and E, or other combinations of the four parameters.

Note that, with the extra quantity E in the reckoning, we can form a dimensionless quantity out of E, m, ℓ and g. Although there are many possible choices, they all depend on the unique combination $E/mg\ell$.

Now applying Bridgman's theorem gives the same result (1.9) as before, except that the undetermined dimensionless constant C can depend on the dimensionless parameter $E/mg\ell$. Thus

$$\omega = f(\lambda)\ell^{-\frac{1}{2}}g^{\frac{1}{2}}, \tag{1.10}$$

where $\lambda = E/mg\ell$ and f is an arbitrary function that, being dimensionless, cannot be determined by dimensional analysis. This result is not nearly as useful as the result for the case of small

[19] We could in fact have argued this from the symmetry properties of the differential equation. Since θ satisfies

$$\frac{d^2\theta}{dt^2} + \frac{g}{\ell}\theta = 0$$

a scaling of time $t \to \lambda t$ together with a scaling of the length of the pendulum $\ell \to \lambda^2\ell$ leaves the equation invariant, so the ratio of period squared to length is invariant.

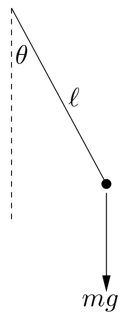
[20] The physical set-up doesn't have to be much more complicated for our intuition to be useless. The double pendulum consists of one pendulum attached to another. The configuration can be described by just two variables, for example, the angle that each pendulum makes with the vertical. The system is therefore governed by two coupled differential equations. For small oscillations, the system is linear and can be treated by matrix methods. If the ratio of the frequencies of the two normal modes is irrational, the system is not periodic. It is therefore only in special cases that the system is periodic. For oscillations that are not small, the system can be chaotic: the motion is extremely sensitive to initial conditions.

oscillations: because of the presence of the unknown function, we cannot now say what will happen if we double the length of the pendulum.

Finally in this example, we verify the above dimensional analysis by comparison with the complete solution.

To solve the problem analytically, we would have to integrate Newton's second law:

$$\ell\ddot{\theta} = -g\sin\theta.$$



The first integral (equivalent to conservation of energy) is

$$\frac{1}{2}\ell\dot{\theta}^2 - g\cos\theta = E/m\ell,$$

where $E/m\ell$ is just a constant of integration. We can identify E with the total energy provided that the potential energy is measured relative to the suspension point at height $\theta = \pi/2$.

Thus the period T is given by

$$\begin{aligned}\int_0^T dt &= 4\int_0^{\theta_0} \frac{d\theta}{\sqrt{2E/m\ell^2 + 2(g/\ell)\cos\theta}} \\ &= 4\sqrt{\frac{l}{g}}\int_0^{\theta_0} \frac{d\theta}{\sqrt{2\cos\theta - 2\cos\theta_0}} \qquad (1.11)\end{aligned}$$

where $\cos\theta_0 = -E/mg\ell$. (The factor of 4 is because the integral to θ_0 is a quarter of the full oscillation.) This makes sense, because $E/mg\ell$ is dimensionless and θ_0 is the maximum value of θ, i.e. the value of θ for which $\dot{\theta} = 0$.

In general, the integral on the right hand side of (1.11) cannot be evaluated in terms of elementary functions[21] though its value clearly depends only on the dimensionless parameter $E/mg\ell$. This is therefore exactly of the form (1.10) arrived at by dimensional analysis.

In the case of small oscillations, we can approximate both cosines in the denominator so (using $\cos x \approx 1 - \frac{1}{2}x^2$) that the integral becomes

$$\int_0^{\theta_0} \frac{d\theta}{\sqrt{\theta_0^2 - \theta^2}} = \int_0^1 \frac{dx}{\sqrt{1-x^2}} = \frac{\pi}{2},$$

where x is a dimensionless variable defined by $x = \theta/\theta_0$. Thus

$$\omega \equiv \frac{2\pi}{T} = \sqrt{\frac{l}{g}},$$

again as expected from dimensional analysis.

[21] It is an elliptic integral. It can be written as $K\ \ \sin\frac{1}{2}\theta_0)$ where K here is the *complete elliptic integral of the first kind* defined by

$$K(m) = \int_0^{\pi/2} \frac{d\theta}{\sqrt{1 - m^2\sin^2\theta}} \equiv \int_0^1 \frac{dt}{\sqrt{(1-m^2t^2)(1-t^2)}}.$$

The incomplete elliptic integral $F(\phi, m)$ has ϕ in the upper limit instead of $\frac{1}{2}\pi$. Its inverse is an elliptic function; this is a generalisation of the case $m = 0$ for which the inverse is a trigonometric function. To be precise, if the inverse function of $F(\phi, m) = u$ is $\phi = \mathrm{f}(u, m)$, so that $F(\mathrm{f}(u, m), m) = u$, then the Jacobi elliptic function sn is defined by $\mathrm{sn}(u, m) = \sin(\mathrm{f}\,(u, m))$. Elliptic functions are great fun, being generalisations of trigonometric functions, but the algebra can be complicated.

1.2.5 The method of dimensional analysis

We can summarise as follows. Suppose that we are trying to find a quantity Q which we know depends on quantities Q_1, Q_2, ..., Q_n. Suppose that the only basic dimensions are L, M and T (no electromagnetic fields, etc). By the power law rule, the dimensions of each of the Q_i are of the form (1.6).

If $n = 3$, we can expect to be able to set

$$Q = CQ_1^{\alpha}Q_2^{\beta}Q_3^{\gamma} \tag{1.12}$$

where C is a dimensionless constant and the three Greek letters can be evaluated by equating powers of the three dimensions.

If $n > 3$, then we can find $n - 3$ dimensionless parameters, $\lambda_1, \ldots \lambda_{n-3}$, each of which is a product of powers of Q_i. In this case, the best we can achieve from dimensional analysis is

$$Q = f(\lambda_1, \lambda_2, \ldots \lambda_{n-3})Q_1^{\alpha}Q_2^{\beta}Q_3^{\gamma},$$

where f is a function that can only be determined by further analysis of the differential equations, and again the three Greek letters α, β and γ can be evaluated by equating powers of dimensions. Of course one has to choose Q_1, Q_2 and Q_3 carefully, so that they are not functionally dependent (there is a danger that the product $Q_1Q_2Q_3$ depends only two of M, L and T). [22]

Note that the choice of the dimensionless parameters is not unique: we could, instead of λ_i use μ_i, where $\mu_i = g_i(\lambda_1, \ldots, \lambda_{n-3})$ and g_i are arbitrarily chosen functions. We could also have chosen the three independent Q_i differently: for example, we could equally well have written, with different f and Greek letters,

$$Q = f(\lambda_1, \lambda_2, \ldots \lambda_{n-3})Q_1^{\alpha}Q_2^{\beta}Q_4^{\gamma}.$$

If this is not all clear, it will become so when you try an example yourself.

1.2.6 G. I. Taylor and the energy of the atomic bomb

In 1950, Taylor published two articles on the energy of the first atomic bomb,[23] revealing information that had been previously classified. His estimate was that the blast was equivalent to 16.8 kilotons of TNT; the figure announced later by president Truman was 20k. Taylor's estimate was was astonishingly accurate, given that it was based on the basis of dimensional analysis and an old photograph.

Taylor argued that the energy of the blast was so great that the pressure and temperature of the air outside the shock are negligible compared with the pressure and temperature inside. He therefore identified the following relevant variables[24]

- E, the energy of the blast, to be determined — $[E] = \mathrm{ML^2T^{-2}}$;
- ρ, the density of the external air — $[\rho] = \mathrm{ML^{-3}}$;
- t, the time elapsed since the explosion — $[t] = \mathrm{T}$;
- $R(t)$, the radius of the roughly spherical expanding fireball — $[R] = \mathrm{L}$.

The idea is that the radius R at time t (i.e. the rate at which the fireball expands) should depend on only two parameters, namely, the energy of the explosion and the density of the external air. Thus

$$R(t) = f(E, \rho, t).$$

Since we are trying to find E, we write this in the form (using the power law rule)

$$E = C\rho^{\alpha}t^{\beta}R^{\gamma},$$

[22] This result, when set out precisely, is Buckingham's Π theorem. The ideas go back to Lord Rayleigh's book *The theory of sound* (1877). Buckingham's proof of the result involves under-determined systems of linear equations and the rank-nullity theorem from Linear Analysis, which all seems a bit heavyweight for this intuitive result. In his original paper, Buckingham called the dimensionless parameters Π_1, Π_2, ..., (instead of λ_1, λ_2, ...), which is why the theorem is named the Π theorem.

[23] The test for the two bombs dropped the following month on Japan.

[24] I meant to read Taylor's paper to see exactly why he didn't include the internal pressure and temperature, but I forgot. *Note to self: check it out.* Presumably they themselves would be determined by the other quantities and are not independent variables.

where C is a constant. Writing this equation in terms of dimensions gives

$$\mathrm{ML}^2\mathrm{T}^{-2} = (\mathrm{ML}^{-3})^\alpha \mathrm{T}^\beta \mathrm{L}^\gamma.$$

By equating coefficient we find $\alpha = 1$, $\beta = -2$ and $\gamma = 5$. Thus

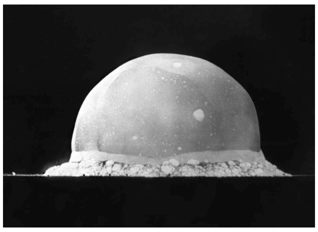
$$E = C\rho t^{-2} R^5.$$

Taylor estimated that the density of air was about 1.25 gram per cubic metre. The published photographs of the blast showed that the radius of the fireball was 80 metres after 0.006 seconds. Thus in joules

$$E \approx C \times 1.25 \times 0.006^{-2} \times 80^5 = C \times 1.13 \times 10^{14}.$$

Taylor then did some calculations to estimate C and so arrived at his value for E. There were further photographs giving the radius at a sequence of times, which verified that the expansion was governed by the power law (but these could not be used to determine C).

Note that to obtain his result, Taylor had not only to use his well-honed intuition to decide which variables were important but also had to to some calculations to determine the constant C. This could only be done by an expert in the field. Nevertheless, it was far easier than attempting to set up and solve the system of differential equations governing the explosion.[25]



One of the photographs used by Taylor to obtain his estimate for the yield of the Trinity test plutonium bomb. This one was taken 0.016 seconds after detonation. The radius is about 100 metres.

[25]Compare Taylor's approach with that of Enrico Fermi who dropped little pieces of paper into the path of the shock wave and observed how far they were blown; as might be expected, this gave a *very* rough answer.

Chapter 2

Forces

We consider here forces acting on a single particle, which may be an idealisation of an extended body. Force is what appears on the right hand side of Newton's second law, but one does use Newton's laws to determine whether a force acts: the force between bodies in any given theory is defined as part of the theory.[1] For example, in Newton's theory of gravity, the force between two particles is defined by the inverse square law (see section 2.4) and in electromagnetism the force on a charged particle is the Lorentz force (section 2.3).

We can distinguish two different sorts of forces: contact forces and non-contact forces. A non-contact force is one that acts at a distance. Examples are gravitational and electromagnetic forces. A non-contact force pervades the whole of space: it exists at every point, whether or not there is a particle at that point to feel it. Often, we will refer to such forces as *force fields*.

A contact force is one that the particle experiences by virtue of being in contact with another body. Examples are friction and normal reaction. In fact, for two bodies, these are the only examples: friction is the component of the force between two bodies that lies in the plane of contact; normal reaction is the component of force in the normal direction. Contact forces also occur when a particle moves through a fluid. Contact forces are caused by interactions between the atoms of the two bodies, so are really just convenient idealisations of non-contact forces.[2]

2.1 Potentials

A force is a vector. In three dimensions, it has three components and is therefore determined by three functions. In some very special, but very important cases, these three functions are related and can be expressed in terms of a single function called a *potential*. Potentials are immensely useful, because they are so much easier both to understand and to calculate.

2.1.1 Potentials in one dimension

We consider a force field $F(x)$ acting on a particle. The work done (WD) by the force in moving the particle from position x by an infinitesimal distance dx is, by definition, given by

$$\text{WD} = F(x)dx \qquad\qquad \text{('force times distance moved by force')}$$

The work done by the force in moving the particle directly[3] from x_0 to x is therefore

$$\int_{x_0}^{x} F(x')dx'.$$

The potential, $\phi(x)$, associated with $F(x)$ is defined by

$$\phi(x) - \phi(x_0) = -\int_{x_0}^{x} F(x')dx'. \tag{2.1}$$

[1] Recall that in section 1.2.2, we took the view that Newton's first law determines whether the frame is inertial, which of course assumes that we know whether a force is acting.

[2] 'Contact' has no meaning at the atomic level: atoms don't touch each other. They interact via non-contact forces such van der Waals forces. These forces are very short range compared with, say, gravitational forces which is the essential difference.

[3] No going backwards and then forwards again.

Thus the potential is a measure of the amount of work done on the particle and hence the ability of the particle itself to do work (i.e. to give back the work done on it[4].) Clearly, the potential is only defined up to an additive constant.

2.1.2 Example: uniform gravitational potential

It is often helpful to think of the familiar example of a uniform gravitational field. For a particle of mass m, the force field has magnitude mg and it acts downwards. The work done by the gravitational field when a particle falls from height z_0 to height z is

$$\int_{z_0}^{z}(-mg)dz = mg(z_0 - z)$$

(the minus sign in the integral arises because z is measured upwards but the force acts downwards). Thus

$$\phi(z) - \phi(z_0) = -mg(z_0 - z)$$

and $\phi(z) = mgz + \text{constant}$.

End of example

2.1.3 Total energy

If we differentiate the equality (2.1) with respect to x, we obtain

$$\frac{d\phi}{dx} = -F(x) \tag{2.2}$$

i.e.

'force equals minus gradient of potential'.

For a particle of mass m moving in a force field $F(x)$ with associated potential $\phi(x)$, we define the total energy E of a particle of mass m moving in the potential ϕ by

$$E = \tfrac{1}{2}m\dot{x}^2 + \phi(x) \tag{2.3}$$

the first term being the kinetic energy and the second being the potential energy.[5]

The total energy is conserved in the motion, i.e. independent of time:

$$\begin{aligned}
\frac{dE}{dt} &= m\dot{x}\ddot{x} + \frac{d\phi}{dt} \\
&= \dot{x}F(x) + \frac{d\phi}{dt} && \text{(using the equation of motion, namely Newton's second law)} \\
&= \dot{x}F(x) + \dot{x}\frac{d\phi}{dx} && \text{(by the chain rule)} \\
&= 0. && \text{(by equation (2.2))}
\end{aligned}$$

Thus the work done by the force contributes to the total energy of the particle, as one might expect. The minus sign in the definition (2.1) means that the potential can be thought of as a form of energy stored in the particle by virtue of its position in the force field which is reduced as the force does work on the particle.

We have shown above that total energy, defined by equation (2.3), is a *conserved quantity* when the force on the particle is derived from a time-independent potential according to (2.2). We will find other conserved quantities (such as momentum and angular momentum). To see if a quantity is conserved, all one has to do is differentiate it with respect to time and use the equations of motion (Newton's second law). Conserved quantities do not necessarily exist in more general situations. For example, there is no conserved quantity that could be interpreted as energy in General Relativity.[6]

A potential can provide an understanding of the motion of a particle without having to solve the equations of motion. This is illustrated in the following example.

---

[4]Rather loosely speaking: particles are structureless objects and can't really do anything

[5]In some situations, the potential is defined not in terms of the force, but in terms of the force on a particle of unit unit mass in a gravitational field, or the force on a particle of unit charge in an electric field.

[6]Conserved quantities are related (by Noether's theorem) to underlying symmetries of the theory. For Newtonian dynamics, the underlying symmetry is the Galilean group (see section 1.1.8). Energy conservation relates to the time translation, momentum conservation relates to spatial translations, and angular momentum conservation relates to rotations.

2.1.4 Example: particle in cubic potential

A particle of unit mass moves in a one-dimensional potential $\phi(x)$, where

$$\phi(x) = x^3 - 3x\,.$$

The force due to this potential is $-\dfrac{d\phi}{dx}$ ('minus the gradient of the potential'), so the equation of motion of the particle is

$$\frac{d^2x}{dt^2} \equiv \ddot{x} = -\frac{d\phi}{dx} = -3x^2 + 3\,. \tag{2.4}$$

Multiplying by $\dfrac{dx}{dt}$ and integrating with respect to time gives the first integral (the energy integral)

$$\tfrac{1}{2}\dot{x}^2 = -\phi(x) + E$$

where E is a constant of integration (the total energy). This first order differential equation can also be integrated in principle to obtain

$$\int \frac{dx}{\sqrt{2E - 2(x^3 - 3x)}} = t.$$

This is an elliptic integral — it cannot be expressed in terms of elementary functions, though its properties have been well-studied.[7]

A more illuminating approach comes from considering the equation of motion (2.4) to be that of a particle of unit mass rolling[8] under the action of gravity in a landscape the height of which above sea-level (say) is $\phi(x)$, as shown in the sketch. (Actually the height is $\phi(x)/g$ so that the gravitational potential is $g \times \phi(x)/g$; but let's just use units in which $g = 1$ so as not to complicate to picture.) Of course, what the particle does is to move along the x-axis, but because the equation of motion is exactly the same, we can translate the problem to that of the rolling particle.

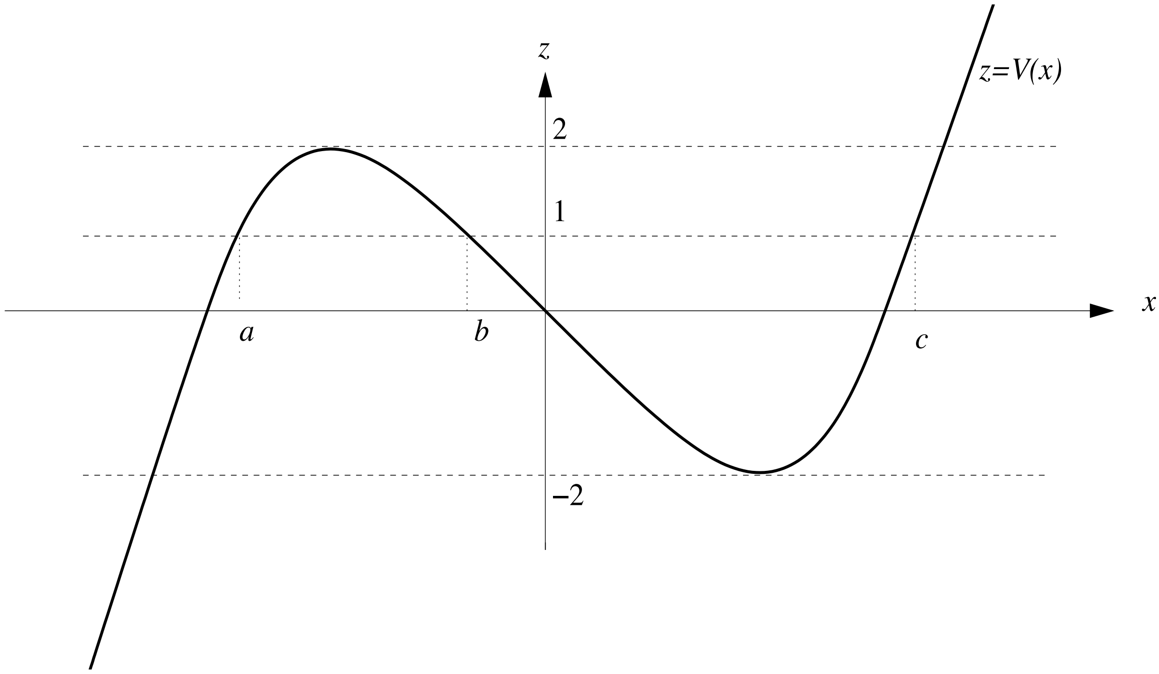
This approach works even for much more complicated potentials, where the integration approach would be unhelpful, and also for potentials that are functions of two variables.

The kinetic energy, and hence speed, of the particle is represented by the difference between the 'height' of the potential function and the fixed 'height' given by the total energy of the particle. At the points where these two heights coincide, the particle has zero speed but non-zero acceleration unless the point is a stationary point of the potential. For a smooth potential function, the particle will reverse when reaching such a point or, if it is a stationary point, will take an infinite amount of time to get there.[9]

[7] The result of doing the integral and then expressing x as a function of t gives an *elliptic function.* Elliptic functions are very beautiful mathematical objects, being doubly periodic in the complex plane: they satisfy a relation of the form $F(z + ma + nb) = F(z)$, where z is any complex number (actually, almost any, since elliptic functions generally have singularities), a and b are fixed complex numbers, and m and n are any integers. The functions can be thought of as existing on lattices in the complex plane, or on toruses. The closest familiar analogy are the trigonometric functions, which which are only singly periodic but can also be defined as the inverse of an integral similar to ours (though quadratic rather than cubic or quartic in the square root in the denominator).

[8] Actually, sliding since particles have no size and therefore cannot really be said to roll; but it is normal to call it rolling.

[9] You can easily check this assertion. For a smooth potential, by which I mean a potential with a Taylor series about each point, the motion of the particle very close to the stationary point is determined approximately by the first non-zero term of the Taylor series, i.e. by the equation $\dot{x}^2 = x^n$, where n is an integer greater than 1. Integrating this shows that the time taken to reach $x = 0$ is infinite.



From the diagram, we can see the following possibilities (there are many others), depending on the initial conditions. For convenience, the initial conditions are given in terms of x_0 and E, rather than x_0 and $\dot{x}_0$.

(i) $x_0 < a$, $\dot{x}_0 > 0$, $E = 1$.
In this case, the particle slows down until its velocity is reversed when $x = a$ (see diagram); it then goes off to $x = -\infty$.

(ii) $x_0 = a$, $E = 1$.
The particle, initially stationary, sets off towards $-\infty$, gathering speed.

(iii) $a < x_0 < b$, $E = 1$.
This is not possible: the particle does not have sufficient energy (classically) to exist in this part of the x-axis.

(iv) $b \le x_0 \le c$, $E = 1$.
The particle oscillates between b and c.

(v) $x_0 > c$, $E = 1$.
Again, not possible.

(vi) $E = 3$.
The particle ends up at $-\infty$ either directly if $\dot{x}_0 \le 0$, or after bouncing off the potential if $\dot{x}_0 > 0$.

(vii) $E = 2$, $x_0 = -1$. Note that the turning points of $\phi(x)$ are at ± 1. In this case the particle has no kinetic energy and just stays put. It is in unstable equilibrium, as is obvious from the diagram. This can be checked analytically. Let $x = -1+\epsilon$, where $\epsilon \ll 1$. Then, substituting into the equation of motion (2.4), we have

$$\frac{d^2}{dt^2}(-1+\epsilon) = -3(-1+\epsilon)^2 + 3 \approx +6\epsilon$$

so $\epsilon \approx \epsilon_0 \cosh \sqrt{6}(t-t_0)$, which grows grows exponentially. Small perturbations from the equilibrium will therefore in general become large, which means the equilibrium is unstable.

End of example

2.1.5 Potentials in three dimensions

As mentioned before, we cannot in general expect to be able to express all three components of a force $\mathbf{F}(\mathbf{r})$ in terms of a single potential. An obvious exception is a three-dimensional force that is essentially one-dimensional, such as the uniform gravitational field discussed in a previous example. There are in fact other exceptions, including many of the forces that arise in theoretical physics.

Following the treatment of the one-dimensional case we write the work done by the force in moving a particle from $\mathbf{r}$ to $\mathbf{r} + d\mathbf{r}$ as

$$\mathrm{WD} = \mathbf{F}.d\mathbf{r}.$$

The scalar product arises naturally here because we are only interested in the component of the force in the direction of motion. The work done in moving the particle from $\mathbf{r}_0 \to \mathbf{r}$ is

$$\int_{\mathbf{r_0}}^{\mathbf{r}} \mathbf{F}(\mathbf{r}').d\mathbf{r}' = \int_{t_0}^{t} \mathbf{F}(\mathbf{r}(t')).\frac{d\mathbf{r}}{dt'}dt' \tag{2.5}$$

where, in the second integral, t is a parameter (which could be time) along the path of integration. This is a line integral, and in general its value depends on the path joining $\mathbf{r_0}$ to $\mathbf{r}$. While the integral makes perfect sense as a measure of work done, it does not define a potential function of $\mathbf{r}$, because of the path dependence.

However, for some forces, the value of the integral does not in fact depend on the path. Such forces are said to be *conservative.* For conservative forces, we can define a potential $\phi(\mathbf{r})$ by[10]

$$\phi(\mathbf{r}) - \phi(\mathbf{r_0}) = -\int_{\mathbf{r_0}}^{\mathbf{r}} \mathbf{F}(\mathbf{r}').d\mathbf{r}'.$$

If the path is parameterised by t, we can differentiate this with respect to t, using the second from of the integral in (2.5), to obtain

$$\frac{d\phi}{dt} = -\mathbf{F}(\mathbf{r}).\frac{d\mathbf{r}}{dt} \tag{2.6}$$

By the chain rule,

$$\begin{aligned} \frac{d\phi}{dt} &= \frac{\partial \phi}{\partial x_i}\frac{dx_i}{dt} && \text{(summation convention)} \\ &= \boldsymbol{\nabla}\phi.\frac{d\mathbf{r}}{dt} \end{aligned}$$

Comparing this with (2.6) and remembering that $\dfrac{d\mathbf{r}}{dt}$, which is the tangent vector to the path, is arbitrary because the path is arbitrary (the value of the integral is the same for all paths), we have

$$\boldsymbol{\nabla}\phi = -\mathbf{F}(\mathbf{r}) \qquad \text{(‘force equals minus gradient of potential’)}$$

As in the one-dimensional case, we define the total energy E of a particle of mass m moving in the potential ϕ by

$$E = \tfrac{1}{2}m\dot{\mathbf{r}}.\dot{\mathbf{r}} + \phi \tag{2.7}$$

and again this is conserved:

$$\begin{aligned} \frac{dE}{dt} &= m\dot{\mathbf{r}}.\ddot{\mathbf{r}} + \frac{d\phi}{dt} \\ &= \dot{\mathbf{r}}.\mathbf{F} + \frac{d\phi}{dt} && \text{(Newton's second law)} \\ &= \dot{\mathbf{r}}.\mathbf{F} + \boldsymbol{\nabla}\phi.\dot{\mathbf{r}} && \text{(chain rule)} \\ &= 0 && \text{(using } \mathbf{F} = -\boldsymbol{\nabla}\phi) \end{aligned}$$

When the potential occurs in the definition of total energy, as in equation (2.7), it is called the *potential energy* of the particle.

2.1.6 Central forces

A force field $\mathbf{F}(\mathbf{r})$ is said to be *central* if it depends only on the distance between the point at which the force is acting (call it $\mathbf{r}$) and a fixed point. If we take the fixed point to be the origin of coordinate, we can write a central force in the form $\mathbf{F}(r)$, where r is the usual spherical polar coordinate.

---

[10] The definition of a conservative force is one for which the work done by the force is independent of the path for all paths between any two fixed points. We should specify that the fixed points and the paths must lie in some given volume V, which might be the whole of $\mathbb{R}^3$. A consequence (see below) is that there exists a function $\phi(\mathbf{x})$ such that $\mathbf{F} = -\boldsymbol{\nabla}\phi$ in V, and it is easily seen that, if such a function exists, $\mathbf{F}$ is conservative. As will be shown in the Vector Calculus course a necessary and sufficient condition for $\mathbf{F}$ to be conservative is $\boldsymbol{\nabla} \times \mathbf{F} = 0$ in V (i.e. $\mathbf{F}$ is curl-free, or irrotational).

A stronger definition of a central force, which is the one we will adopt, is that it acts towards or away from the fixed point. In this case, we can write the force in the form $f(r)\widehat{\mathbf{r}}$, where $\widehat{\mathbf{r}}$ is the unit vector in the radial direction.

For such a force, we can have hopes that it is conservative, since it depends on one function only; and our hopes are fulfilled. Recall[11] that

$$\boldsymbol{\nabla} r = \widehat{\mathbf{r}}.$$

Thus if we define $\phi(r)$ (up to an additive constant of integration) by $f(r) = -\dfrac{d\phi}{dr}$, we have

$$\mathbf{F}(\mathbf{r}) = f(r)\widehat{\mathbf{r}} = -\frac{d\phi}{dr}\boldsymbol{\nabla} r = -\boldsymbol{\nabla}\phi$$

using the chain rule for the last equality.

Therefore, any central force can be written in terms of a potential, and the total energy (potential plus kinetic) is conserved.

2.2 Friction

As mentioned earlier, friction is a contact force. It is a convenient way of describing the complicated interactions between the atoms of different bodies but is not itself a fundamental force[12]

There are two sorts of friction: *dry friction* and wet friction or *drag*. Dry friction occurs when two bodies are in contact; a particle resting or sliding on an inclined plane, for example. The governing equation is

$$F = \mu R,$$

where R is the normal reaction and μ is the coefficient of friction. This applies both in static friction (a body at rest) and sliding friction, though the coefficient of friction between two given bodies will be different in the two cases. This sort of friction is not particularly relevant to this course.

2.2.1 Fluid drag

Drag occurs when a body is moving through a fluid. Drag is velocity dependent, being normally either linear or quadratic in speed and parallel to velocity.

Linear drag is caused by the stickiness of the fluid and takes the form

$$\mathbf{F} = -k\mathbf{v}$$

where k is independent of velocity. *Stokes's law* for a spherical body gives $k = 6\pi\eta R$, where η is the viscosity and R is the radius of the body. Drag is approximately linear when viscous forces predominate, for example a rock in lava or a bacterium in water.

Quadratic drag takes the form

$$\mathbf{F} = -k|\mathbf{v}|\mathbf{v} \tag{2.8}$$

where k is now depends on the density of the fluid and the cross-sectional area of the body. This occurs when the resistance to motion is due to the body having to push the fluid aside, for example projectiles in air[13] and submarines.

Linear friction dominates when the speed of the body is small in the following sense:

$$\frac{\rho|\mathbf{v}|R^2}{\eta} \approx 1.$$

The dimensionless quantity on the left hand side is called the *Reynolds number*.[14]

---

[11] This is most easily verified using the expression for $\boldsymbol{\nabla}$ in polar coordinates:

$$\boldsymbol{\nabla} f = \frac{\partial f}{\partial r}\widehat{\mathbf{r}} + \frac{1}{r}\frac{\partial f}{\partial \theta}\widehat{\boldsymbol{\theta}} + \frac{1}{r\sin\theta}\frac{\partial f}{\partial \phi}\widehat{\boldsymbol{\phi}}.$$

Otherwise, one can conveniently consider $\boldsymbol{\nabla} r^2$ in Cartesians.

[12] The fundamental forces are gravitation, electromagnetic forces, and weak and strong nuclear forces.

[13] Including stones being dropped from leaning towers — see section 4.

[14] The Reynolds number measures the relative importance of inertial forces and viscous forces in fluid flow: if it is less than about 2000, the flow is laminar; if it is greater than about 4000 the flow is turbulent.

2.2.2 Example: vertical motion under gravity with quadratic friction

A particle of mass m moves vertically under the influence of gravity (assumed uniform) and a quadratic resistance force of magnitude mkv^2, where v is the velocity[15] of the particle. In what follows, we take z to measure distance vertically upwards. By definition, $v = \dfrac{dz}{dt}$, so $v > 0$ if the particle is moving upwards and $v < 0$ if the particle is moving downwards, and similarly $\dfrac{d^2z}{dt^2} > 0$ if v is (strictly) increasing.

(i) Upwards motion

The equation of motion is 'vector' equation — we have to worry about direction just as if we were working in three dimensions. Gravity acts downwards and, when the particle moves upwards, the resistive force also acts downwards. The equation of motion is therefore

$$m\frac{dv}{dt} = -mg - mkv^2, \qquad \text{i.e.} \qquad \frac{dv}{dt} = -g - kv^2.$$

Integrating this will give us v as a function of t. Alternatively, we could write the equation of motion, using the chain rule, as

$$v\frac{dv}{dz} = -g - kv^2 \qquad \text{i.e.} \qquad \tfrac{1}{2}\frac{d(v^2)}{dz} = -g - k(v^2) \tag{2.9}$$

which will give us v^2 as a function of height z.

Suppose the particle is projected upwards from $z = 0$ with speed V and we want to find the maximum height H. We can obtain the form of the expression for H by considering dimensions. The dimension of k is L^{-1}, as can be seen from the equation of motion (2.9). The other relevant quantities are g and V. Since that makes three in total, and they all involve only two dimensions, L and T, there is one dimensionless parameter, call it η. The choice of η is not, of course, unique; one possibility is $\eta = kV^2/g$ (and any function of this quantity would do). We therefore expect that H can be written in the form

$$H = k^{-1}f(\eta),$$

where the function $f(\eta)$ cannot be determined by dimensional analysis.

Integrating equation (2.9), and noting that the particle reaches its maximum height when $v = 0$, gives

$$\int_V^0 \frac{d(v^2)}{g + kv^2} = \int_0^H -2dz$$

so

$$2H = -\frac{1}{k}\log\left(\frac{g}{kV^2 + g}\right) = \frac{1}{k}\log\left(1 + \eta\right). \tag{2.10}$$

Now suppose that effect of friction is weak compared with the effect of gravity. Since the effect of friction is greatest at the point of projection, weak friction corresponds to $kV^2 \ll g$, i.e. $\eta \ll 1$. Expanding the log in the expression for H (2.10) gives

$$H = \frac{1}{2k}(\eta - \tfrac{1}{2}\eta^2 + \cdots) = H_o(1 - \tfrac{1}{2}\eta + \cdots),$$

where H_o is the height that the particle would have attained in the absence of friction. Note that we have to expand to second order in the small parameter to see the affect of friction.

(ii) Downwards motion – e.g. a raindrop

This time the equation of motion is of the particle

$$\frac{dv}{dt} = -g + kv^2,$$

because now the resistive force acts upwards.

Suppose the particle is dropped from (i.e. released from rest at) a great height. As we know from experience, there is a terminal speed which cannot be exceeded; in fact, as we shall see, it cannot be attained.

[15] Note: velocity not speed — it can be either positive or negative

We can find the speed v at any time t by integrating directly[16]:

$$\int_0^t dt = -\int_0^v \frac{dv}{g - kv^2}.$$

Setting $\sqrt{k}\, v = \sqrt{g} \tanh\theta$ gives

$$t = -\frac{1}{\sqrt{gk}} \tanh^{-1}(\sqrt{k/g}\, v) = \frac{1}{\sqrt{gk}} \tanh^{-1}(\sqrt{k/g}\, |v|)$$

(remember that v is negative for a falling particle).

It is a good idea to pause occasionally and check that all is dimensionally in order. Since $[k] = L^{-1}$ and $[g] = LT^{-2}$, we find that $[\sqrt{gk}] = T^{-1}$ and $[\sqrt{k/g}] = L^{-1}T$, which means that the above equation is dimensionally correct.

Thus

$$v = -\sqrt{g/k} \tanh(\sqrt{gk}t).$$

Note that as $t \to \infty$, $v \to \sqrt{g/k}$, though this speed is never attained. The quantity $\sqrt{g/k}$ is called the *terminal velocity* (more properly, the terminal speed).

The quantity $1/\sqrt{gk}$ is the only combination of parameters that has the dimension of time, so it must provide a timescale analogous to the half-life of exponential decay. My calculator give $\tanh 1 \approx 0.76$, so $1/\sqrt{gk}$ is the time taken for the particle to reach about 3/4 of the terminal velocity, starting from rest.

The terminal velocity for a sky diver in the free fall position (limbs outstretched) is about 55 metres per second — call it 50 metres per second. Taking $g = 10$ metres per second per second, we see that $k^{-1} = 250$ metres and the timescale is 5 seconds. Very roughly, the terminal velocity is proportion to the square root of its area (see the remark following equation (2.8) regarding the dependence of k on area). The terminal velocity for a mouse[17] is much smaller than for a human, so it is more likely to have a happy landing.[18]

End of example

2.2.3 Example: projectile with linear drag, using vectors

As mentioned above, typical projectiles in air are subject to quadratic drag, so the one we are thinking about here must be in water or maybe even treacle.

A particle of mass m is projected from the origin at velocity $\mathbf{u}$. The gravitational acceleration is denoted by $\mathbf{g}$ and the drag force is $-mk\mathbf{v}$, where k is a constant (the m is included here for convenience).

The equation of motion (Newton's second law) is

$$m\frac{d\mathbf{v}}{dt} = m\mathbf{g} - mk\mathbf{v}$$

i.e.

$$\frac{d\mathbf{v}}{dt} + k\mathbf{v} = \mathbf{g}.$$

[16] Note the lazy convention of not using a dummy variable in the integral; the only excuse for this is that there is no scope for confusion here.

[17] J.B.S. Haldane in his essay *On Being the Right Size* summarises the situation nicely but not delicately: 'You can drop a mouse down a thousand-yard mine shaft; and, on arriving at the bottom, it gets a slight shock and walks away, provided that the ground is fairly soft. A rat is killed, a man is broken, a horse splashes.'

[18] Curiously, cats survive big drops better than small drops according to a 1987 study from the Journal of the American Veterinary Medical Association wherein two vets examined 132 cases of cats that had fallen out of high-rise windows and were brought to the Animal Medical Center, a New York veterinary hospital, for treatment.

The vets postulated that cats sense acceleration, rather than speed. When a cat starts falling it begins accelerating at something close to g and it accordingly assumes its 'panic' posture: head tucked in; paws under body, arched back. This protects its vital organs, but unfortunately makes it more aerodynamic i.e. smaller k, corresponding to a bigger terminal velocity at which it is likely to be killed if it strikes a hard surface.

However, the acceleration reduces considerably as terminal velocity is approached, and the cat adopts a different strategy. This tends to happen after the cat has fallen about 8 storeys. It stretches out its legs and neck, like a flying fox, increasing its surface area, which increases k, decreases the terminal velocity, and so slows the cat down. At this lower terminal velocity it can survive the fall — from any height greater than 8 storeys, though apparently 32 stories is the highest on record. No cats were harmed in the making of this footnote.

We can solve this equation using an integrating factor, as if it were an ordinary (non-vector) differential equation. We first rewrite it as

$$\frac{d}{dt}\left(e^{kt}\mathbf{v}\right) = e^{kt}\mathbf{g}$$

then integrate and multiply by e^{-kt}:

$$\mathbf{v} = \frac{1}{k}\mathbf{g} + \mathbf{C}\,e^{-kt}$$

where $\mathbf{C}$ is a constant (vector) of integration which can be identified using the initial condition $\mathbf{v} = \mathbf{u}$ at $t = 0$. Thus

$$\mathbf{v} = \frac{1}{k}\mathbf{g} + (\mathbf{u} - \frac{1}{k}\mathbf{g})e^{-kt}.$$

This equation can be integrated directly to give $\mathbf{r}$:

$$\mathbf{r} = \frac{1}{k}\mathbf{g}t - \frac{1}{k}(\mathbf{u} - \frac{1}{k}\mathbf{g})e^{-kt} + \mathbf{d}$$

where $\mathbf{d}$ is a (vector) constant of integration which can be identified using the initial condition $\mathbf{r} = 0$ at $t = 0$. Thus

$$\mathbf{r} = \frac{t}{k}\mathbf{g} + \frac{1}{k}(\mathbf{u} - \frac{1}{k}\mathbf{g})(1 - e^{-kt}). \tag{2.11}$$

This is the complete solution. Choosing axes such that

$$\mathbf{r} = \begin{pmatrix} x \\ y \\ z \end{pmatrix} \qquad \mathbf{g} = \begin{pmatrix} 0 \\ 0 \\ -g \end{pmatrix} \qquad \text{and} \qquad \mathbf{u} = \begin{pmatrix} u\cos\alpha \\ 0 \\ u\sin\alpha \end{pmatrix},$$

the solution is

$$x = \frac{1}{k}u\cos\alpha\left(1 - e^{-kt}\right), \quad y = 0, \quad z = -\frac{gt}{k} + \frac{1}{k}\left(u\sin\alpha + \frac{g}{k}\right)\left(1 - e^{-kt}\right).$$

This looks a bit more complicated than the $k = 0$ case, but it is has some expected features. For very large t, in the sense $kt \gg 1$, the exponential terms can be ignored and the particle drops vertically at its terminal speed of g/k. The horizontal component of velocity has been completely eroded by the drag force.

For small k (i.e. $kt \ll 1$) we should retrieve the projectile-without-drag solution. At first sight, this limit looks bad because of the k in the denominator. However, if we expand the exponential in the vector form of the solution (2.11) as far as the quadratic terms we see that the limit is in fact defined (as it must be):

$$\begin{aligned} \mathbf{r} &= \frac{t}{k}\mathbf{g} + \frac{1}{k}(\mathbf{u} - \frac{1}{k}\mathbf{g})(1 - 1 + kt - \tfrac{1}{2}(kt^2) + \cdots) \\ &= \mathbf{u}t + \tfrac{1}{2}\mathbf{g}t^2 + O(kt). \end{aligned}$$

This is of course[19] the solution that we would have obtained by solving the equations of motion with $k = 0$.

End of example

2.3 Motion in an electromagnetic field

2.3.1 The Lorentz force

The *Lorentz force*[20] is the force experienced by a charged particle in an electromagnetic field. It is given by

$$\mathbf{F} = e\,(\mathbf{E} + \mathbf{v} \times \mathbf{B}) \tag{2.12}$$

[19]It is not quite 'of course'. We are asking if two processes commute: is solving an equation and taking a limiting value of the solution the same as taking the limit in the equation then solving it. This is the sort of thing one has to worry about when studying the theory of differential equations, and partial differential equations in particular, but we needn't let it detain us here.

[20]It was introduced by the Dutch physicist Hendrik Lorentz (after whom the Lorentz transformations of special relativity are named) in 1892, though a more or less equivalent equation occurs in the works of Maxwell thirty years earlier. Lorentz was awarded the Nobel prize in 1902 for his work on the Zeeman effect.

where e is the charge on the particle, $\mathbf{v}$ is the velocity of the particle, $\mathbf{E}$ is the electric field and $\mathbf{B}$ is the magnetic field. The term $e\mathbf{E}$ is called the electric force and the term $e\mathbf{v}\times\mathbf{B}$ is called the magnetic force.[21]

Generally, $\mathbf{E}$ and $\mathbf{B}$ depend on both position and time, though in the simple examples considered here, both these forces are likely to be uniform and constant. Note that the force is defined everywhere in space and time, regardless of whether a charge is present to experience the force.

A positively charged particle will be accelerated in the same direction as the electric field, but will curve perpendicularly to both the instantaneous velocity and the magnetic field according to the right-hand rule.

If we take the scalar product of equation (**??**) with $\mathbf{v}$, we see that

$$\mathbf{F}\cdot\mathbf{v} = e\mathbf{E}\cdot\mathbf{v}.$$

The left hand side of this equation is the rate at which work is done by the force $\mathbf{F}$ on the particle, and so we see that the magnetic field does not contribute at all the work done; it is all done by the electric part of the force field.

Like drag force, the electromagnetic force $\mathbf{F}$ depends on the velocity of the particle (explicitly), as well as on its position (implicitly, via the dependence of $\mathbf{E}$ and $\mathbf{B}$ on position) so it is not in general conservative; though it may be conservative in special cases[22]

2.3.2 Electric field of a point charge

The electric field of a particle of stationary charge q is given by

$$\mathbf{E} = \frac{q}{4\pi\epsilon_0}\frac{\mathbf{r}}{r^3}. \tag{2.13}$$

where ϵ_0 is a constant called the permittivity of free space. It relates the units of electric charge to the mechanical quantities M, L and T. Its value is $8.8541878210^{-12}\mathrm{m}^{-3}\mathrm{kg}^{-1}\mathrm{s}^4 A$, where A is the basic electric unit (amperes). Since $\boldsymbol{\nabla} r^n = n\mathbf{r}r^{n-2}$ (see the Vector Calculus course), we can write $\mathbf{E}$ as a gradient:

$$\mathbf{E} = -\boldsymbol{\nabla}\left(\frac{q}{4\pi\epsilon_0 r}\right).$$

The quantity $\dfrac{q}{4\pi\epsilon_0 r}$ is called the electrostatic potential for the point charge.

When $\mathbf{B}=\mathbf{0}$, as in the case for a stationary charge, the Lorentz force (2.12) is proportional to $\mathbf{E}$, so the force on a particle moving in the field of a point electric charge is conservative. The field of a point charge is very similar (identical really) to that of a point gravitational mass, as we shall see in the next section.

It is worth noting that since there are no free point magnetic charges (magnetic charges occur in pairs as in a bar magnet), there is no corresponding field for a point magnetic charge. The simplest magnetic field (that is not constant) is called a *dipole* field, which is the result of superposing a positive magnetic charge and a negative magnetic charge.

2.3.3 General motion of a charged particle in an electromagnetic field

In general, $\mathbf{E}$ and $\mathbf{B}$ are functions of both time t and position $\mathbf{r}$; in Cartesian coordinates, they are functions of x_i, where $\mathbf{r}=(x_1,x_2,x_3)$, and t. We assume that the electromagnetic fields are given and are not affected by the presence of a charged particle. [23]

Writing the trajectory of the particle as $\mathbf{r}(t)$, the equation of motion becomes

$$m\ddot{\mathbf{r}}(t) = \mathbf{F} = e\Big(\mathbf{E}\big(\mathbf{r}(t),t\big) + \dot{\mathbf{r}}(t)\times\mathbf{B}\big(\mathbf{r}(t),t\big)\Big),$$

which represents three coupled second-order ordinary non-linear (in general) differential equations with three dependent variables, and can in principle be solved, given suitable initial conditions.

[21] You don't have to know anything about electric and magnetic fields for this course: that comes in Part IB. The Lorentz force is included in this course as an exercise — an important exercise — in handling vector equations of motion.

[22] Electromagnetic fields are governed by the Maxwell equations, one of which is $\boldsymbol{\nabla}\times\mathbf{E} = -\frac{\partial\mathbf{B}}{\partial t}$. Thus the Lorentz force is conservative if (and only if) $\mathbf{B}=\mathbf{0}$.

[23] Of course, this is an idealisation: a moving charged particle will create its own electromagnetic fields and these might well affect the source (whatever it is) of the given $\mathbf{E}$ and $\mathbf{B}$.

2.3.4 Example: motion in a uniform electromagnetic field

Here we consider the case when the electromagnetic field is both constant (in time) and uniform (same at all points in space), so that

$$\frac{\partial \mathbf{E}}{\partial t} = \frac{\partial \mathbf{B}}{\partial t} = \mathbf{0}\,; \qquad \frac{\partial \mathbf{E}}{\partial x_i} = \frac{\partial \mathbf{B}}{\partial x_i} = \mathbf{0}\,.$$

Recall that a vector is (e.g.) time-independent if and only if its Cartesian components are time-independent.

The equation

$$\mathbf{F} \equiv e(\mathbf{E} + \dot{\mathbf{r}} \times \mathbf{B}) = m\ddot{\mathbf{r}}. \tag{2.14}$$

can be tackled in a number of ways. Below, we will solve it entirely in components and also entirely in vectors. Neither method is optimal: a judicious mixture would serve us better.

(i) Component method

The practical way to integrate the questions is to work in components; **BUT** it is essential to choose sensible axes. Since the lines of $\mathbf{B}$ are everywhere parallel, we can choose axes such that the z-axis is parallel to $\mathbf{B}$:

$$\mathbf{B} = (0, 0, B)$$

If $\mathbf{E}.\mathbf{B} = \mathbf{0}$, we can choose axes such that $\mathbf{E} = (E, 0, 0)$, but in general the best we can do (by rotating the x and y axes, which is the only freedom left after fixing the z axis) is

$$\mathbf{E} = (E_1, 0, E_3).$$

With this choice, the equations of motion (2.14) become

$$m\ddot{x} = eE_1 + eB\dot{y} \tag{2.15}$$
$$m\ddot{y} = \qquad - eB\dot{x} \tag{2.16}$$
$$m\ddot{z} = eE_3 \tag{2.17}$$

which can be solved by elementary means or by using matrices.

The solution to third equation (2.17) can be written down:

$$z = (e/2m)E_3 t^2 + at + b \tag{2.18}$$

where a and b are constants obtainable from initial conditions.

A neat way to solve the first two equations (2.15) and (2.16), which happens to work in this case, is to set $\xi = x + iy$, and add i times equation (2.16) to equation (2.15); of course, one could always do this to obtain a single complex equation containing both ξ and $\bar{\xi}$, but the special feature of our equations is that the result does not contain $\bar{\xi}$:

$$m\ddot{\xi} = eE_1 - ieB\dot{\xi}\,.$$

This can be integrated straight away:

$$\xi = pe^{-i\omega t} - iE_1 t/B + q$$

where $\omega = eB/m$ and the complex constants p and q can be obtained from the initial conditions.[24]

If the particle is initially at the origin, and moving in the y-direction, we find

$$\xi = p(e^{-i\omega t} - 1) - ikt\,,$$

where $k = E_1/B$ and p is real, so

$$x = p(\cos\omega t - 1)\,, \qquad y = -p\sin\omega t - kt\,.$$

This is roughly (exactly if $k = p$) a cycloid, so the motion of the particle is, somewhat counter-intuitively, a uniform acceleration parallel to $\mathbf{B}$ (but due to the component of the electric field parallel to $\mathbf{B}$) and cycloidal motion in the plane perpendicular to $\mathbf{B}$.

[24] ω is called the *Larmor frequency* after the physicist Joseph Larmor, senior wrangler in 1880, Lucasian Professor from 1903–1932. Larmor published the complete Lorentz transformations if special relativity in the Philosophical Transactions of the Royal Society in 1897 some two years before Hendrik Lorentz (1899, 1904) and eight years before Albert Einstein (1905). Larmor predicted the phenomenon of time dilation, at least for orbiting electrons, and verified that the FitzGerald-Lorentz contraction (length contraction) should occur for bodies whose atoms were held together by electromagnetic forces. This however was all in the context of an aether theory of space-time.

(ii) Vector algebra method

Now we resolutely refuse to use choose axes at all.

We first dot equation (2.14) with $\mathbf{B}$ to obtain

$$m\ddot{\mathbf{r}}.\mathbf{B} = e\mathbf{E}.\mathbf{B}.$$

This can be integrated directly since $\ddot{\mathbf{r}}.\mathbf{B} = \dfrac{d^2(\mathbf{r}.\mathbf{B})}{dt^2}$:

$$\mathbf{r}.\mathbf{B} = (e/2m)\mathbf{E}.\mathbf{B}t^2 + at + b\,. \tag{2.19}$$

This is equivalent, in the coordinate-dependent method, to the z-equation (2.18).

What now? To be sure that no information is lost, we should really next cross equation (2.14) with $\mathbf{B}$. We would then have taken first the component of the equation parallel to $\mathbf{B}$ and subsequently the component perpendicular to $\mathbf{B}$. That would be a systematic approach. We could also dot with $\dot{\mathbf{r}}$, and integrate:

$$\tfrac{1}{2}\dot{\mathbf{r}}.\dot{\mathbf{r}} = e\mathbf{E}.\mathbf{r} + \text{constant}$$

giving an energy-like conservation equation, which may or may not be helpful (it isn't particularly helpful for present purposes).[25]

However, the easiest way forward in this particular case is to integrate the vector equation once directly, giving:

$$m\dot{\mathbf{r}} = e\mathbf{E}t + e\mathbf{r}\times\mathbf{B} + \mathbf{C}$$

where $\mathbf{C}$ is a vector constant of integration.[26] Now that we have an expression for $\dot{\mathbf{r}}$, we can substitute it into the right hand side of the equation of motion (2.14) to obtain an equation of the form (the details are getting messy, so the constant vectors are just called $\mathbf{A}_i$):

$$\begin{aligned} m\ddot{\mathbf{r}} =& \mathbf{A_1} + \mathbf{A_2}t + (e^2/m)(\mathbf{r}\times\mathbf{B})\times\mathbf{B} \\ =& \mathbf{A_1} + \mathbf{A_2}t + (e^2/m)\left((\mathbf{r}.\mathbf{B})\mathbf{B} - (\mathbf{B}.\mathbf{B})\mathbf{r}\right) \\ =& \mathbf{A_3} + \mathbf{A_4}t + \mathbf{A_5}t^2 - (e^2B^2/m)\mathbf{r} \\ =& -(e^2B^2/m)\mathbf{r} + \text{ other stuff} \end{aligned}$$

where we have used in the penultimate equation the expression (2.19) for $\mathbf{r}.\mathbf{B}$. This is just the vector simple harmonic motion equation (or three individual simple harmonic motion equations if we wrote it out in Cartesian coordinates) with additional forcing terms. The solution to this equation can more or less be written down:

$$\mathbf{r} = \mathbf{C_1}\cos\omega t + \mathbf{C_2}\sin\omega t + \text{Particular integral}$$

in agreement with what was obtained rather more easily in components.

End of example

2.4 Gravitational forces

2.4.1 Newton's universal law of gravitation

Newton's law of gravitation (published in *Principia* in 1687)[27] states that the gravitational force experienced by a particle of mass m_2 due to a particle of mass m_1 at distance r has magnitude

$$\frac{Gm_1m_2}{r^2}. \tag{2.20}$$

This is the inverse square law of gravitational attraction. The constant G in this expression is *Newton's gravitational constant*, aka the *universal gravitational constant* or just 'big G'. It has a

[25] It is not exactly conservation of energy because there is no potential for the Lorentz force in this case, and hence no potential energy. It is instead a statement about the work done by the force on the particle: the element of force due to the magnetic field is perpendicular to the velocity and hence does no work.

[26] Recall that $\dfrac{d(\mathbf{r}\times\mathbf{B})}{dt} = \dot{\mathbf{r}}\times\mathbf{B}$.

[27] There is some controversy about whom credit for this law should be attributed. Certainly, Hooke, Halley and Christopher Wren had all discussed it. What is not controversial is that Newton demonstrated that planets would move in ellipses, in agreement with observations, if moving under the influence of an inverse square law.

value of $6.6730010^{-11}\mathrm{m^3kg^{-1}s^{-2}}$. Note the dimensions: $\mathrm{L^3M^{-1}T^{-2}}$. Its value is quite hard to measure because gravitational forces are comparatively very weak.[28] For example, the ratio of the strength of the gravitational force between a proton and an electron to the strength of the electrostatic force between a proton and an electron (2.13) the same distance apart is

$$\frac{G \times m_{proton} m_{electron}}{q_{proton} q_{electron}/4\pi\epsilon_0} = \frac{(6.6 \times 10^{-11}) \times (1.6 \times 10^{-27}) \times (9 \times 10^{-31})}{1.6 \times 10^{-19}) \times (1.6 \times 10^{-19})/4\pi \times 8.8 \times 10^{-12}} \approx 10^{-39}$$

Combinations of the form GM_S and GM_E, where M_S and M_E are the mass of the sun and the Earth, respectively, are much easier to determine: they can be deduced from the period and radius of the orbits of the Earth and the Moon (see section 3.3).

The gravitational force between to particles is *central* (which means that it is directed from one particle towards the other) and attractive, so can be expressed in vector form as

$$\mathbf{F}_{12} = -\frac{Gm_1m_2}{r^3}\mathbf{r} \tag{2.21}$$

where $\mathbf{r}$ is the vector from particle 1 to particle 2 and $\mathbf{F}_{12}$ is the force exerted *by* the particle of mass m_1 *on* the particle of mass m_2. . In more general notation,

$$\mathbf{F}_{12} = -\frac{Gm_1m_2}{|\mathbf{r}_2 - \mathbf{r}_1|^3}(\mathbf{r}_2 - \mathbf{r}_1)\,.$$

As proved in section 2.1, all central forces are conservative. The potential for the force (2.21) is given by (recall that $\boldsymbol{\nabla} r^n = nr^{n-2}\mathbf{r}$)

$$\phi(r) = -\frac{Gm_1m_2}{r}; \qquad (\boldsymbol{\nabla}\phi = -\mathbf{F})$$

In more general notation,

$$\phi_{12}(\mathbf{r}_1, \mathbf{r}_2) = -\frac{Gm_1m_2}{|\mathbf{r}_2 - \mathbf{r_1}|}.$$

We can take the gradient of this expression for the potential with respect to $\mathbf{r_2}$, regarding $\mathbf{r_1}$ as fixed, to obtain the force on particle 2 due to particle 1, or vice versa; the difference will only be minus sign (as expected from Newton's third law).

2.4.2 Important note

It is normal when considering gravitational potentials to omit the mass of the particle being acted on (the passive particle). Thus the gradient of the potential would give the acceleration of the passive particle not the force acting on it. I will distinguish between the two usages by using lower case ϕ for the potential which is equal to the potential energy of the particle (the gradient of the which gives the force), and upper case Φ for the potential more commonly used for gravitational and electric fields, the gradient of which gives the acceleration. Thus for a particle of mass M at the origin, the gravitational potential Φ is given by

$$\Phi(r) = -\frac{GM}{r}$$

whereas a particle of mass m moving in this potential would experience a force derived ('force = minus gradient of potential') from the potential function

$$\phi(r) = -\frac{GMm}{r}.$$

2.4.3 Addition of gravitational fields

Newtonian gravitational potentials are *linear* in the sense that the total potential due to two particles is just the sum of the potentials of the individual potentials. This is an observationally determined

[28] It was measured by Henry Cavendish in 1798 using a torsion balance.

result and does not hold for all types of potential.[29] Thus the potential at the point $\mathbf{r}$ due to point masses at points $\mathbf{r}_i$ $(i = 1,\ 2,\ \ldots)$ is

$$\Phi(\mathbf{r}) = -\sum_i \frac{Gm_i}{|\mathbf{r} - \mathbf{r}_i|}$$

and the total gravitational force on a particle of mass m at $\mathbf{r}$ is

$$-m\boldsymbol{\nabla}\Phi = -\sum_i \frac{Gm_i m(\mathbf{r} - \mathbf{r}_i)}{|\mathbf{r} - \mathbf{r}_i|^3}.$$

If all the masses are smeared out into a mass distribution with density $\rho(\mathbf{r})$, so that the mass in a volume dV of space is ρdV, the sums can be replace by a volume integrals[30] to obtain the total potential at the point $\mathbf{r}$:

$$\Phi(\mathbf{r}) = -\int \frac{G\rho(\mathbf{r}')\,dV'}{|\mathbf{r} - \mathbf{r}'|} \tag{2.22}$$

and the total gravitational force on a particle of mass m at $\mathbf{r}$ is

$$-\int \frac{Gm\rho(\mathbf{r}')\,(\mathbf{r} - \mathbf{r}')\,dV'}{|\mathbf{r} - \mathbf{r}'|^3}.$$

This is obtained by simply differentiating under the integral sign in (2.22), noting that $\mathbf{r}'$ is a dummy variable and is therefore a constant as far as differentiation with respect to $\mathbf{r}$ is concerned.

2.4.4 Gravitational field of a spherically symmetrical body

This is an important example: what we shall show is that the external gravitational field of a spherically symmetric body, such as a planet, of mass M is the same as that of a particle of mass M located at the centre of the body. **BUT** important though this result is, you should not regard the following calculation as being part of this course; it is really just an example of a volume integral as might be calculated in the Vector Calculus course, so you should stop reading this now, and come back to it when you are revising Vector Calculus. Though actually, the calculation is not very difficult.

We will demonstrate the result by evaluating the integral (2.22) to find the gravitational potential. Let the density of the body be $\rho(r)$ (which just depends on r, the distance from the centre, because the body is spherically symmetric), and let the radius of the body be a. We will calculate the gravitational field at a fixed point with position vector $\mathbf{R}$, a distance R from the centre, where $R \geq a$.

The first step is to choose coordinates. Obviously, we will use spherical coordinates, but the trick is to choose the polar direction $\theta = 0$ in the direction of $\mathbf{R}$. This means that for a position vector $\mathbf{r}$, the scalar product $\mathbf{R}.\mathbf{r} = Rr\cos\theta$. Further,

$$|\mathbf{R} - \mathbf{r}|^2 = (\mathbf{R} - \mathbf{r}).(\mathbf{R} - \mathbf{r}) = R^2 + r^2 - 2\mathbf{r}.\mathbf{R} = R^2 + r^2 - 2Rr\cos\theta.$$

[29] A more modern view (19th century — more modern than Newton) is that the forces of nature can be derived from potentials that satisfy Laplace's equation ($\nabla^2\phi = 0$). Since this equation is linear, solutions can be superposed. Einstein's equations for general relativity are non-linear and solutions cannot, in general, be added to obtain a new solution.

[30] You will come across volume integrals in the Vector Calculus course; we will not need to perform complicated integrals in this course.

Thus

$$
\begin{aligned}
\Phi(\mathbf{R}) &= -\int_{|\mathbf{r}|\le a} \frac{G\rho(r)dV}{\sqrt{R^2+r^2-2rR\cos\theta}} \\
&= -\int_0^a\int_0^\pi\int_0^{2\pi} \frac{G\rho(r)r^2\sin\theta d\phi d\theta dr}{\sqrt{R^2+r^2-2rR\cos\theta}} && \text{(remembering to put in the Jacobian)} \\
&= -\int_0^a\int_0^\pi \frac{2\pi G\rho(r)r^2\sin\theta d\theta dr}{\sqrt{R^2+r^2-2rR\cos\theta}} && \text{(doing the trivial } \phi\text{-integral)} \\
&= -\int_0^a\int_{-1}^1 \frac{2\pi G\rho(r)r^2 dr dc}{\sqrt{R^2+r^2-2rRc}} && \text{(setting } \cos\theta = c\text{)} \\
&= \int_0^a \frac{2\pi G\rho(r)r\left[\sqrt{R^2+r^2-2rRc}\right]_{c=-1}^{c=1} dr}{R} && \text{(doing the } c \text{ integral)} \\
&= -\int_0^a \frac{2\pi G\rho(r)r\left(|R+r|-|R-r|\right)dr}{R} && \text{(evaluating at } c=\pm1\text{)} \\
&= -\int_0^a \frac{4\pi G\rho(r)r^2 dr}{R} && \text{(using } r\le a\le R\text{)} \\
&= -\frac{GM}{R}
\end{aligned}
$$

as required.

End of example

2.5 Escape velocity

For a particle moving in a force field, the escape velocity is just the velocity that the particle must have to get out of the influence of the field; which normally means out to infinity. Often, one is thinking of projecting a particle from the surface of the Earth (say): the escape velocity tells you how fast you must project it for it to escape the Earth's gravitational pull.

For a general force field, the concept of escape velocity is not very helpful: the escape velocity would depend on the trajectory, and would not be possible to calculate without completely solving the equations of motion.

For a force field derived from a potential, such as a gravitational field, the concept is more useful because there is some chance that the escape velocity can be expressed in terms of the potential, without having to solve the equations of motion. If the particle has sufficient energy to overcome the potential it will escape. This is what is illustrated in the example on the motion of a particle in a cubic potential in section 2.1.

Even for a gravitational field, the concept only works well in the simplest case, the field of a single spherically symmetrical body such as (to good approximation) the Earth or the Sun. Even in the case of just two gravitating bodies, the escape velocity can depend critically on the direction of projection of the particle. For example, interplanetary probes use what is called the 'slingshot' effect to give the probe extra momentum by choosing the direction of projection so that the probe passes close to other planets. Voyager 1, which is now the furthest human-made object from Earth, is in the boundary zone between the Solar System and interstellar space. It gained the energy to escape the Sun's gravity completely by performing slingshot manoeuvres around Jupiter and Saturn. The energy gained was of course taken from the two planets, which perhaps slowed down[31] or moved further apart.

For a spherically symmetric planet of radius R, the gravitational potential at the surface is

$$-\frac{GM}{R}$$

[31]But not by much.

(see the example in section 2.5). Thus a particle of mass m projected with speed v from the surface has total energy E given by

$$E = \tfrac{1}{2}mv^2 - \frac{GMm}{R}$$

which is conserved. The potential energy of the particle if it escaped to infinity would be zero, so in order to have sufficient energy to escape,

$$E > 0,$$

i.e.

$$v > \sqrt{\frac{2GM}{R}} \equiv v_{\mathrm{esc}}.$$

The minimum value of v, namely v_{esc}, is the escape velocity; or, more properly, the escape speed since it is independent of direction.

Clearly a particle that has less than this speed on projection cannot escape and will fall back to the point of projection. A particle that has at least this speed will escape and eventually (taking infinite time) reach infinity. This last statement is perhaps not quite obvious: clearly, the particle can reach infinity if it is projected radially outwards, because it could only turn round if its speed (and hence its kinetic energy) reduced to zero which, by conservation of energy, is impossible if $E > 0$. But what if it is projected tangentially? As we shall see in chapter 3, it then follows a parabolic or hyperbolic path, again out to infinity.

2.6 Three kinds of mass

This is an extended footnote: interesting, I hope, and relevant; but not strictly part of the course.

We can recognise three different sorts of mass that arise in Newtonian dynamics:

- *Inertial mass*, which occurs in Newton's second law:

$$\text{force} = \text{inertial mass} \ \times \ \text{acceleration}.$$

- *Passive gravitational mass*, which measures the response of a particle to a gravitational field. For example, at the surface of the Earth, the vertical force on a particle is given by

$$\text{passive gravitational mass} \times g.$$

- *Active gravitational mass*, which measures the magnitude of the gravitational field produced by a massive body.

All three kinds of mass occur simultaneously in the formula for the acceleration $\mathbf{a}$ of particle 1 of inertial mass $m_{\mathrm{i}}^{(1)}$ and passive gravitational mass $m_{\mathrm{p}}^{(1)}$ moving with acceleration $\mathbf{a}$ in the gravitational field of particle 2 of active gravitational mass $m_{\mathrm{a}}^{(2)}$:

$$m_{\mathrm{i}}^{(1)}\mathbf{a} = -\frac{Gm_{\mathrm{p}}^{(1)}m_{\mathrm{a}}^{(2)}\mathbf{r}}{r^3}$$

The fact that we only use one kind of mass, that is, we assume that the three apparently different kinds of mass are the same, needs explanation.

2.6.1 Equality of active and passive gravitational mass

According to the law of universal gravitation, the gravitational force on particle 1 due to particle 2, $\mathbf{F_{12}}$, is given by

$$\mathbf{F_{12}} = \frac{Gm_{\mathrm{p}}^{(1)}m_{\mathrm{a}}^{(2)}(\mathbf{r_2} - \mathbf{r_1})}{|\mathbf{r_2} - \mathbf{r_1}|^{\mathbf{3}}}$$

and the gravitational force on particle 2 due to particle 1, $\mathbf{F_{21}}$ is given by

$$\mathbf{F_{21}} = \frac{Gm_{\mathrm{a}}^{(1)}m_{\mathrm{p}}^{(2)}(\mathbf{r_1} - \mathbf{r_2})}{|\mathbf{r_1} - \mathbf{r_2}|^{\mathbf{3}}}\ .$$

Newton's third law demands that these forces are equal in magnitude, so we require

$$\frac{m_{\mathrm{a}}^{(1)}}{m_{\mathrm{p}}^{(1)}} = \frac{m_{\mathrm{a}}^{(2)}}{m_{\mathrm{p}}^{(2)}}$$

and furthermore that this relationship holds for all particles. Since the ratio of active and passive gravitational masses is equal for all particles, we can choose it to be unity (which would just involve scaling G).

2.6.2 Equality of inertial and gravitational mass

This is more difficult: we have so far in this course encountered no law or principle that would determine or even suggest a relationship between inertial and gravitational mass. Nevertheless, inertial and gravitational mass have been found in a number of celebrated experiments to coincide to a very high degree. It is, for example, what Galileo was trying to demonstrate by (supposedly) dropping objects from the top of the leaning tower of Pisa.

If you slide a particle down a slope and measure the acceleration a, you have

$$m_{\mathrm{i}} a = m_{\mathrm{p}} g \sin\theta,$$

where θ is the angle of the slope to the horizontal. If the acceleration is found to be the same for different particles then the only varying quantity in the above equation, namely the ratio $m_{\mathrm{p}}/m_{\mathrm{i}}$ must in fact be the same for the different particles and, as before, it can be normalised to one.

The Hungarian physicist Eötvös[32] spend much of his working life demonstrating the equality of inertial and gravitational mass. His method was to suspend two heavy spheres made from different material from a torsion balance, which consists of a horizontal rod suspended from a fixed point by a quartz fibre attached to its midpoint. The two spheres experience the Earth's gravitational force and also a centrifugal force due to the rotation of the Earth (see chapter 3), The spheres were arranged so that the rod was exactly horizontal, which means that the gravitational masses balanced exactly. If the inertial masses did not balance, the rod would rotate. It didn't.

This experiment was improved by Robert Dicke, an American physicist, using the effect of the Sun's gravitational field, which would have given a 24-hour periodic oscillation if the inertial mass and gravitational masses were inequivalent; this is extremely sensitive and established the equivalence to an accuracy of 1 part in 10^{12}.

The equivalence of inertial and gravitational mass (the *principle of equivalence*) is a fundamental pillar of modern physics; without it General Relativity, which interprets gravitational forces as fictitious forces (i.e. like centrifugal forces) due to motion in a curved space-time, would collapse.

[32]Vásárosnaményi Bárö Eötvös Loránd, 1848–1919; his surname is pronounced, roughly, utvush (u as in 'put').

Chapter 3

Orbits

3.1 Motion in a plane

You might wonder why there is a section on motion in two dimensions when vector methods can be readily used to study motion in three dimensions — or an arbitrary number of dimensions. The answer is that in the case of a particle moving under the influence of a central force (such as the gravitational field of a massive uniform spherical body or the electric field of a point charge), the motion takes place in a plane. We will prove that in section 3.2.

For a particle moving in a central force, the obvious coordinates to use are plane polar coordinates, with the origin at the centre of the force: for example, at the Sun (regarded as a point) in the case of planetary orbits. Before studying such orbits, we need some preliminary results.

3.1.1 Angular variables

It is useful, before we move to more general motion in the plane, to review the motion of a particle on a circle. In this simple situation, we meet some concepts that will defined more formally later in this section and in later chapters.

For a particle moving in a circle of (constant) radius a, the usual kinematic variables — distance, speed, acceleration and momentum — are not very convenient. Instead, the position of the particle on the circle can easily be determined in terms of the obvious angular variable, θ. Similarly, the speed is $a|\dot{\theta}|$ which is determined in terms of $\dot{\theta}$, the *angular velocity*[1]. The angular component of acceleration is $a\ddot{\theta}$, which is determined by $\ddot{\theta}$, the *angular acceleration*.

We can write the kinetic energy in the form

$$\mathrm{KE} = \tfrac{1}{2}mv^2 = \tfrac{1}{2}ma^2\dot{\theta}^2 \equiv \tfrac{1}{2}I\dot{\theta}^2$$

where we have defined a quantity I, the *moment of inertia* of the particle, by

$$I = ma^2.$$

When angular velocity is used instead of (linear) velocity, the usual formula for kinetic energy holds provided mass is replaced by moment of inertia.[2] We will make extensive use of the idea of moment of inertia when we study the motion of rigid bodies in Chapter 5.

Instead of the (linear) momentum of the particle, we introduce a new quantity, the *angular momentum*, denoted by H, using the angular expressions corresponding to mass $\times$ velocity:

$$H = I\dot{\theta} \equiv ma^2\dot{\theta}. \tag{3.1}$$

Finally, suppose that there is a tangential force of magnitude F acting on the particle. Newton's second law in this situation is

$$F = m(a\ddot{\theta})$$

which we can write in the form

$$G \equiv aF = ma^2\ddot{\theta} \equiv I\ddot{\theta}$$

[1] Velocity rather than speed because it can be positive or negative. We are here considering essentially one-dimensional motion round a circle; in the next sections, we will investigate the velocity vector in two dimensions.

[2] Moment of inertia could be thought of as 'angular mass', though this would sound a little odd.

so the angular version of Newton's second law has G, or aF, on the left-hand side, which is called the *moment* of the force, or *torque*.

The table below summarises the correspondence.

Motion on a straight line	Motion on a circle
Displacement x	Angular displacement θ
Velocity $\dot{x} = v$	Angular velocity $\dot{\theta} = \omega$
Acceleration $\ddot{x} = \dot{v}$	Angular acceleration $\ddot{\theta} = \dot{\omega}$
Mass m	Moment of inertia $I = ma^2$
Momentum $m\dot{x} = mv$	Angular momentum $I\dot{\theta} = I\omega$
Kinetic energy $\frac{1}{2}mv^2$	(Rotational) kinetic energy $\frac{1}{2}I\omega^2$
Newton's second law $F = m\dot{v}$	Newton's second law $G = I\dot{\omega}$

The idea of angular variables need not be confined to simple motion in a circle, as we shall see as this course progresses.[3]

3.1.2 Acceleration: coordinate independent treatment

We will calculate the acceleration of a particle whose motion is confined to a plane in two ways: in this section, coordinate independently, using the tangent and normal vectors to the trajectory; and in the next section using plane polar coordinates.

Let $\mathbf{r}$ be the position vector of the particle at time t, with respect to a fixed origin, and let $\mathbf{t}$ be the unit tangent vector[4] defined by

$$\mathbf{t} = \frac{d\mathbf{r}}{ds}, \tag{3.2}$$

where s is arc-length.

The velocity vector $\dot{\mathbf{r}}$ is parallel to $\mathbf{t}$:

$$\dot{\mathbf{r}} = \dot{s}\frac{d\mathbf{r}}{ds} \equiv v\mathbf{t} \tag{3.3}$$

where $v = \dot{s}$ and we used the chain rule

$$\frac{d}{dt} = \dot{s}\frac{d}{ds}$$

for the first equality. We define the arc-length s to increase in the direction that the curve is traversed so that $v \geq 0$.

Differentiating equation (3.3) gives

$$\begin{aligned}\ddot{\mathbf{r}} &= \dot{v}\mathbf{t} + v\dot{\mathbf{t}} \\ &= \dot{v}\mathbf{t} + v^2\frac{d\mathbf{t}}{ds}\end{aligned} \tag{3.4}$$

using the chain rule again.

Now $\mathbf{t}\cdot\mathbf{t} = \mathbf{1}$ implies that

$$2\frac{d\mathbf{t}}{ds}\cdot\mathbf{t} = 0,$$

i.e., $\dfrac{d\mathbf{t}}{ds}$ is orthogonal to $\mathbf{t}$. Thus there is a scalar ρ such that

$$\rho\frac{d\mathbf{t}}{ds} = \mathbf{n},$$

or

$$\frac{d^2\mathbf{r}}{ds^2} = \frac{1}{\rho}\mathbf{n} \tag{3.5}$$

[3]The discussion above is a special case of a much grander scheme, which is presented in the Part II course Classical Dynamics. The starting point is a set of generalised coordinates q_i (here, just θ) and a function called the Lagrangian L (here, just the kinetic energy), which is expressed in terms q_i and the generalised velocities $\dot{q}_i$ (here, just $\dot{\theta}$). The generalised momenta p_i, (here, the angular momentum) are defined by $\partial L/\partial\dot{q}$ and Newton's laws translate to $\partial L/\partial q_i = -\dot{p}_i$, the left-hand side corresponding to a generalised force expressed as the gradient of a generalised potential. It is all very neat.

[4]There is an unfortunate duplication here in the use of the letter t; I will not mention time t again to prevent confusion.

where $\mathbf{n}$ is the normal to the curve and its direction is chosen to ensure that $\rho \geq 0$. (Recall that we are in two dimensions, so this last statement is unambiguous.)

Thus the acceleration of the particle (3.4) can be written in the form

$$\ddot{\mathbf{r}} = \dot{v}\mathbf{t} + \frac{v^2}{\rho}\mathbf{n}. \tag{3.6}$$

The first term is the rate of change of speed, which is the component of acceleration along the curve; the second term is the acceleration in the normal direction. We know that the acceleration of a particle moving in a circle of radius R is v^2/R, so we can identify ρ as the instantaneous radius of curvature of the trajectory.

You will have seen derivations similar to this in connection with the Serret-Frenet equations in Vector Calculus.

To understand this better, we consider points $\mathbf{r}(s)$ on the curve close to any given point $\mathbf{r}(0)$, so that the distance s along the curve is small. The Taylor series for $\mathbf{r}(s)$ is

$$\begin{aligned} \mathbf{r}(s) &= \mathbf{r}(0) + s\frac{d\mathbf{r}}{ds} + \tfrac{1}{2}s^2\frac{d^2\mathbf{r}}{ds^2} + O(s^3) \\ &= \mathbf{r}(0) + s\mathbf{t} + \tfrac{1}{2}s^2\mathbf{n}/\rho + +O(s^3) && \text{(using the definition (3.5))} \end{aligned}$$

where the derivatives, and $\mathbf{t}$, $\mathbf{n}$ and ρ, are evaluated at $s = 0$. This shows that to first order in s, the curve can be approximated by a straight line $\mathbf{r}(s) = \mathbf{r}(0) + s\mathbf{t}$, which is what we were expecting.

It is not so obvious from this equation that, to second order in s, the curve can be approximated by a circle. However, considering $\big|\mathbf{r}(s) - \mathbf{r}(0) - \rho\mathbf{n}\big|^2$ shows that to this approximation (i.e. ignoring s^3 and higher order terms) $\mathbf{r}(s)$ describes a circle of radius ρ with centre at $\mathbf{r}(0) - \rho\mathbf{n}$:

$$\begin{aligned} \big|\mathbf{r}(s) - \mathbf{r}(0) - \rho\mathbf{n}\big|^2 &= \big|s\mathbf{t} - \tfrac{1}{2}s^2\mathbf{n}/\rho - \rho\mathbf{n}\big|^2 && \text{(from the Taylor series)} \\ &= \big(s\mathbf{t} - \tfrac{1}{2}s^2\mathbf{n}/\rho - \rho\mathbf{n}\big)\cdot\big(s\mathbf{t} - \tfrac{1}{2}s^2\mathbf{n}/\rho - \rho\mathbf{n}\big) \\ &= \rho^2 + O(s^3), && \text{(recall that } \mathbf{n}\cdot\mathbf{t} = \mathbf{0}) \end{aligned}$$

as required.

3.1.3 Example: car on bridge

We investigate the possibility of a small car taking off (leaving the ground) as it goes over a bridge. The equation of motion, using the expression (3.4) for the acceleration, is

$$\mathbf{F} + \mathbf{R} + m\mathbf{g} = m\dot{v}\mathbf{t} + m\frac{v^2}{\rho}\mathbf{n}. \qquad \text{(force = mass times acceleration)}$$

The forces on the left-hand side are as follows. $\mathbf{F}$ is the force of friction between the road and the car tyres; it is what pushes the car along and it is in the direction $\mathbf{t}$. It is equal to the driving force provided by the car engine minus various losses due to friction of bearings and other moving parts. $\mathbf{R}$ is the normal reaction of the road on the car wheels; it is in the direction $\mathbf{n}$. Finally, $\mathbf{g}$ is, as always, the acceleration due to gravity, which is vertical. We have ignored air resistance. On the right-hand side, v is the speed of the car and ρ is the radius of curvature of the bridge.

The car will take off if the gravitational force is not sufficient to provide the acceleration required for the car to follow the curve of the bridge. We therefore look only at the normal component of the equation of motion:

$$mg\cos\theta - R = m\frac{v^2}{\rho}.$$

The normal reaction, of magnitude R, is the difference between the component of the gravitational force in the downwards normal direction and the normal acceleration required (times the mass), and when this vanishes the car is on the point of taking off. The maximum speed v_{max} is therefore given by

$$v_{\text{max}} = \sqrt{g\rho\cos\theta}.$$

The maximum speed depends on the shape of the bridge. At points where the radius of curvature is infinite, which means that the bridge is not curved at all, the car can go as fast as it likes. The most dangerous points are where the bridge is steep ($\cos\theta$ is small) and highly curved (ρ is small).

Putting in some typical figures: $\rho = 40$ metres, $\cos\theta = 1$ and $g = 10$ metres/sec/sec gives $v_{\text{max}} = 20$ metres/sec which is about 45 miles per hour.

End of example

3.1.4 Acceleration in polar coordinates

The previous calculation of acceleration used axes tied to the trajectory of the particle: namely the tangent and normal. Instead we use plane polar coordinates and axes. The axes still depend on the position of the particle (unlike Cartesian axes), as shown in the figure, but not on the direction of the trajectory.

Let $\widehat{\mathbf{r}}$ and $\widehat{\boldsymbol{\theta}}$ be unit vectors in the directions of r and θ increasing, respectively. In Cartesian axes,

$$\widehat{\mathbf{r}} = (\cos\theta, \sin\theta), \qquad \widehat{\boldsymbol{\theta}} = (-\sin\theta, \cos\theta)$$

so

$$\frac{\partial\widehat{\mathbf{r}}}{\partial\theta} = \widehat{\boldsymbol{\theta}}, \qquad \frac{\partial\widehat{\boldsymbol{\theta}}}{\partial\theta} = -\widehat{\boldsymbol{r}}. \quad \text{and} \quad \frac{\partial\widehat{\mathbf{r}}}{\partial r} = \frac{\partial\widehat{\boldsymbol{\theta}}}{\partial r} = 0,$$

Starting from $\mathbf{r} = r\,\widehat{\mathbf{r}}$, we find that the velocity can be written in the form

$$\dot{\mathbf{r}} \equiv \frac{d\mathbf{r}}{dt} = \frac{d(r\widehat{\mathbf{r}})}{dt} = \dot{r}\,\widehat{\mathbf{r}} + r\frac{d\widehat{\boldsymbol{r}}}{dt} = \dot{r}\,\widehat{\mathbf{r}} + r\frac{d\theta}{dt}\frac{d\widehat{\boldsymbol{r}}}{d\theta} = \dot{r}\,\widehat{\mathbf{r}} + r\dot{\theta}\widehat{\boldsymbol{\theta}}\,. \tag{3.7}$$

Differentiating again gives the acceleration:

$$\begin{aligned}
\ddot{\mathbf{r}} &= (\ddot{r}\widehat{\mathbf{r}} + \dot{r}\dot{\theta}\widehat{\boldsymbol{\theta}}) + (\dot{r}\dot{\theta}\widehat{\boldsymbol{\theta}} + r\ddot{\theta}\widehat{\boldsymbol{\theta}} - r\dot{\theta}^2\widehat{\mathbf{r}}) \\
&= (\ddot{r} - r\dot{\theta}^2)\widehat{\mathbf{r}} + (2\dot{r}\dot{\theta} + r\ddot{\theta})\widehat{\boldsymbol{\theta}} && (3.8) \\
&= (\ddot{r} - r\dot{\theta}^2)\widehat{\mathbf{r}} + \frac{1}{r}\frac{d}{dt}(r^2\dot{\theta})\widehat{\boldsymbol{\theta}}\,, && (3.9)
\end{aligned}$$

which is the required (and very important) result.

Note that both the radial and the angular component of acceleration in the expression (3.8) consist of two terms: the ‘obvious’ one and another that depends on first derivatives of the coordinates. Later, we will identify these extra terms as corresponding to the centrifugal and coriolis accelerations in a rotating frame. In the second expression (3.9), the quantity in the derivative will later be identified as the angular momentum (per unit mass).

3.2 Angular momentum

3.2.1 Definition

We can extend the idea of angular momentum discussed in section 3.1.1 to the case of an arbitrarily moving particle. We still need a fixed point corresponding to the centre of the circle in circular motion. This point can be chosen arbitrarily, but the angular momentum will depend upon the choice.

For a particle of mass m at position $\mathbf{r}$ (relative to a given though irrelevant origin) moving with velocity $\dot{\mathbf{r}}$, we define the angular momentum $\mathbf{H}$ about the fixed point $\mathbf{a}$ by

$$\mathbf{H} = (\mathbf{r} - \mathbf{a}) \times (m\dot{\mathbf{r}}). \tag{3.10}$$

Taking lengths of both sides gives

$$|\mathbf{H}| = mvd$$

where d is the shortest distance between the line of the momentum and the fixed point $\mathbf{a}$ (see diagram).[5] In general, given a vector quantity, the cross product with the position vector relative to $\mathbf{a}$ is called the *moment* of the quantity about $\mathbf{a}$. For example, the moment about the point $\mathbf{a}$ of a force $\mathbf{F}$ acting at a point $\mathbf{r}$ is

$$(\mathbf{r} - \mathbf{a}) \times \mathbf{F},$$

so angular momentum might equally be called *moment of momentum*.

[5] Note to self: need diagram.

3.2.2 Conservation of angular momentum

We now derive an important conservation law for angular momentum, corresponding to the conservation of linear momentum in directions orthogonal to the applied force. We have

$$\begin{aligned} \frac{d\mathbf{H}}{dt} &= m\dot{\mathbf{r}} \times \dot{\mathbf{r}} + m(\mathbf{r}-\mathbf{a}) \times \ddot{\mathbf{r}} && \text{(differentiating (3.10) using the product rule)} \\ &= m(\mathbf{r}-\mathbf{a}) \times \ddot{\mathbf{r}} && (\dot{\mathbf{r}} \times \dot{\mathbf{r}} = \mathbf{0}) \\ &= (\mathbf{r}-\mathbf{a}) \times \mathbf{F} && \text{(Newton's second law)} \\ &\equiv \mathbf{G} \end{aligned}$$

where the last equation defines the quantity $\mathbf{G}$, which is the moment of the force, or *torque*, acting on the particle.

For a central force, $\mathbf{F}$ is parallel to $\mathbf{r}$ (see section 2.1.6) so the moment of the force about the origin is zero. This the angular momentum about the centre of force is constant: taking $\mathbf{a} = \mathbf{0}$ in the definition of angular momentum, we have

$$\mathbf{G} = \mathbf{0} = \frac{d\mathbf{H}}{dt}.$$

This will turn out to be an important result when we come to consider orbits in a central force.

In general (whether $\mathbf{H}$ is constant or not), it follows from the definition (3.10) that

$$(\mathbf{r}-\mathbf{a}) \cdot \mathbf{H} = 0$$

i.e.

$$\mathbf{r} \cdot \mathbf{H} = \mathbf{a} \cdot \mathbf{H}.$$

In the case when $\mathbf{H}$ is a constant vector[6] this equation describes a plane with normal $\mathbf{H}$ containing the point $\mathbf{a}$, so in this case the motion of the particle lies entirely in this plane. Using the results of section 3.1.4, for motion in a plane, and taking $\mathbf{a} = \mathbf{0}$ for convenience, we have

$$\begin{aligned} \mathbf{H} &= m\mathbf{r} \times \dot{\mathbf{r}} \\ &= m(r\widehat{\mathbf{r}}) \times (\dot{r}\widehat{\mathbf{r}} + r\dot{\theta}\widehat{\boldsymbol{\theta}}\,) && \text{(using equation (3.7))} \\ &= mr^2\dot{\theta}\,\widehat{\mathbf{r}} \times \widehat{\boldsymbol{\theta}} \\ &= mr^2\dot{\theta}\,\widehat{\mathbf{z}} \end{aligned}$$

where $\widehat{\mathbf{z}}$ is the unit normal to the plane of motion. Note that the magnitude of this quantity is exactly the angular momentum of circular motion discussed in section 3.1.1.

3.3 Orbits in a central force

3.3.1 Equations of motion

We recall that, by definition, a central force $\mathbf{F}(\mathbf{r})$ can be written in the form

$$\mathbf{F}(\mathbf{r}) = f(r)\widehat{\mathbf{r}}\,. \tag{3.11}$$

The angular momentum, $\mathbf{H}$, of a particle about the centre of force, which is here the origin, is given by

$$\mathbf{H} = m\mathbf{r} \times \dot{\mathbf{r}}$$

and is constant for the force (3.11), as was shown in the previous section. The motion takes place in the plane given by $\mathbf{H} \cdot \mathbf{r} = \mathbf{0}$, i.e. in the plane spanned by $\mathbf{r}$ and $\dot{\mathbf{r}}$, which is intuitively obvious: this is the plane in which the force acts, so there is no component of acceleration taking the particle out of the plane.

The equation of motion in plane polar coordinates and axes (3.9) of a particle moving in the force field (3.11) is

$$(\ddot{r} - r\dot{\theta}^2)\widehat{\mathbf{r}} + \frac{1}{r}\frac{d}{dt}(r^2\dot{\theta})\widehat{\boldsymbol{\theta}} = \frac{1}{m}f(r)\widehat{\mathbf{r}}.$$

[6] Note that if $\mathbf{H}_1$ is the angular momentum of a particle about $\mathbf{a}_1$ and $\mathbf{H}_2$ is the angular momentum of the particle about $\mathbf{a}_2$, then $\mathbf{H}_1 - \mathbf{H}_2 = (\mathbf{a}_2 - \mathbf{a}_1) \times \mathbf{p}$, where $\mathbf{p}$ is the linear momentum of the particle. Therefore, the angular momentum about one point being constant does not imply that the angular momentum about another point is constant.

Taking components gives two equations for the two unknowns $\theta(t)$ and $r(t)$:

$$\ddot{r} - r\dot{\theta}^2 = \frac{1}{m} f(r) \tag{3.12}$$

$$\frac{d(r^2\dot{\theta})}{dt} = 0. \tag{3.13}$$

These are the equations of motion of a particle subject to a central force. There are two main approaches to integrating these equations: the straightforward approach is to eliminate $\dot{\theta}$ to give a second-order non-linear differential equation for r with t as the independent variable; the alternative approach is to change variable to obtain a linear differential equation with θ as the independent variable. These two approaches are described in the next two sections.

In both cases, we start by integrating (3.13):

$$r^2\dot{\theta} = h \tag{3.14}$$

where h is a constant, the *angular momentum per unit mass*. Now we use this equation to eliminate $\dot{\theta}$ from equation (3.12)

$$\ddot{r} - \frac{h^2}{r^3} = \frac{1}{m} f(r). \tag{3.15}$$

3.3.2 The r-t orbital equation

Equation (3.15) is seemingly simple, but the non-linear r^{-3} term makes it pretty intractable, as a second order differential equation, even without the force term on the right-hand side. However, it is a useful equation, and we can make some progress by obtaining a first integral:

$$\tfrac{1}{2}\dot{r}^2 + \frac{h^2}{2r^2} = \frac{1}{m}\int^r f(r')dr' \equiv -\Phi(r) + A, \tag{3.16}$$

where $\Phi(r)$ is the potential for $\mathbf{F}(\mathbf{r})/m$ (see section 2.1.6) and A is constant of integration. (Just differentiate equation (3.16) and compare with equation (3.15) to verify that it is correct.)

Rearranging the first integral (3.16) slightly gives

$$\tfrac{1}{2}\dot{r}^2 + \frac{h^2}{2r^2} + \Phi(r) = A,$$

which looks exactly like the usual equation of conservation of energy for a particle (of unit mass) moving in one-dimension, with the quantity

$$\frac{h^2}{2r^2} + \Phi(r) \tag{3.17}$$

acting as the potential. The first term in equation (3.17) is called the 'centrifugal barrier' potential and relates to the kinetic energy required to maintain, when the particle is at radius r, the constant angular momentum h. The quantity (3.17) is called the *effective* potential.

In the case of the gravitational field of a point particle, the effective potential is

$$\frac{h^2}{2r^2} - \frac{GM}{r}$$

and we can determine the motion of the particle qualitatively as in the example in section 2.1.4.

We can in principle integrate equation (3.17) again, starting with

$$\dot{r} = \pm\sqrt{2A - 2\Phi(r) - h^2/r^2}, \tag{3.18}$$

the choice of sign ($\pm$) being determined by the direction of motion (r increasing or decreasing). However, except in the case of a few special potentials, the function in square root will lead to a difficult integral.

When this integral can be evaluated, we obtain $r(t)$, i.e. r as function of t, and we can then find $\theta(t)$ from $r^2\dot{\theta} = h$ (equation 3.13). This give the *dynamical* solution $(r(t), \theta(t))$ of the equations of motion.

3.3.3 The u-θ orbital equation

We can instead find the *geometrical* solution $r(\theta)$ by means of a beautiful transformation of the differential equation (3.15) that makes it not just tractable but familiar. The transformation has two steps.

- Change the independent variable, using the chain rule, from t to θ. This seems to be a small mathematical step but the effect is to change fundamentally the way we look at the problem: it was a dynamical equation for r as a function of t; it becomes a geometrical equation for r as a function of θ, the solution of which is simply a plane curve.[7]
- Change the dependent variable from r to u defined by $u = r^{-1}$. The effect of this is to linearise the left-hand side.

The overall effect is to pass from a differential equation for $r(t)$ to a simpler differential equation for $u(\theta)$.

We work on these two steps together. First note that setting $r = u^{-1}$ in (3.14) gives

$$\dot{\theta} = hu^2. \tag{3.19}$$

Then

$$\begin{aligned}\frac{d}{dt} &= \frac{d\theta}{dt}\frac{d}{d\theta} && \text{(using the chain rule)}\\ &= hu^2\frac{d}{d\theta} && \text{(using (3.19))}\end{aligned}$$

so that

$$\frac{dr}{dt} = hu^2\frac{dr}{d\theta} = hu^2\frac{d(1/u)}{d\theta} = -h\frac{du}{d\theta} \tag{3.20}$$

and

$$\frac{d^2r}{dt^2} = \frac{d}{dt}\left(-h\frac{du}{d\theta}\right) = hu^2\frac{d}{d\theta}\left((-h\frac{du}{d\theta}\right) = -h^2u^2\frac{d^2u}{d\theta^2}. \tag{3.21}$$

We have assumed throughout that $\dot{\theta} \neq 0$, i.e. that the motion is not purely radial: clearly if θ is constant in the motion, it cannot be used to replace t as the parameter along the trajectory.

Substituting this into the equation of motion (3.15) gives

$$-h^2u^2\frac{d^2u}{d\theta^2} - h^2u^3 = \frac{1}{m}f(1/u) \tag{3.22}$$

i.e.

$$\frac{d^2u}{d\theta^2} + u = -\frac{f(1/u)}{mh^2u^2} \tag{3.23}$$

which is the *geometric orbital equation.*

The plan is to solve this equation for a given force $f(1/u)$, then reintroduce time via

$$\dot{\theta} = hu^2$$

which can now be integrated to give θ as a function of time, since u is known as a function of θ. Finally, now that we have θ as a function of time, we can find r as a function of time from $r = 1/u(\theta)$.

3.3.4 Kinetic energy

The expression for the kinetic energy of the particle in terms of u is a rather pleasing. We have:

$$\begin{aligned}\tfrac{1}{2}\dot{\mathbf{r}}\cdot\dot{\mathbf{r}} &= \tfrac{1}{2}(\dot{r}^2 + r^2\dot{\theta}^2) && \text{(using the formula (3.7))}\\ &= \tfrac{1}{2}(h^2u'^2 + h^2u^2) && \text{(using (3.20) and (3.19))}\\ &= \tfrac{1}{2}h^2(u'^2 + u^2), && (3.24)\end{aligned}$$

where the prime denotes differentiation with respect to θ. This is the kinetic energy per unit mass.

[7] This change of variable is possible because the equations are *autonomous*, which means that there is no explicit t dependence in the equations.

3.4 Circular orbits

3.4.1 Existence

Closed orbits are those for which $r(\theta + 2\pi n) = r(\theta)$ for some integer n. Note that n need not be equal to 1: it is possible to imagine an orbit given by, say, $r = 2 + \cos\frac{1}{2}\theta$, for which $n = 2$ (though this may correspond to a rather peculiar force law[8].

A very special case of a closed orbit is a *circular orbit*. For any attractive ($f(r) < 0$) central force, the orbital equation (3.23) admits circular orbits of any radius: the radius r, and hence u, is constant for a circular orbit so for any given u_0, it is only necessary to choose the value of h for which the orbital equation is satisfied:

$$u_0 = -\frac{f(1/u_0)}{mh^2u_0^2} \quad \text{or} \quad \frac{1}{r_0^3} = -\frac{f(r_0)}{mh^2}\,, \tag{3.25}$$

where $r_0 = 1/u_0$. This equation determines a unique value of h^2 for any value of the radius of the circular orbit. In the rotating frame, this just means that the angular velocity is such that the centrifugal force exactly balances the central force.[9]

3.4.2 Stability

In order to determine whether an orbit is stable, we consider the evolution of a small perturbation. Since the orbit is determined by two variables, $\theta(r)$ and $r(t)$, there are two modes of perturbation: radial and tangential. However, the tangential direction is associated with the angular momentum, which is constant, so the effect of perturbation in the tangential direction can be related to a small change in angular momentum. Since angular momentum is conserved, this small change neither grows nor decays: it just remains small. We shall therefore confine attention to the radial direction.

We can investigate the evolution of the perturbation using orbital equation either in the $r(t)$ form (3.15) or the $u(\theta)$ form (3.23). Perhaps the $r(t)$ form is easier to use.

Let $r = r_0 + \eta$, where r_0 is the (constant) radius of the circular orbit, which satisfies (3.25), and η is a small time-dependent perturbation.

Substituting into the orbital differential equation (3.15) gives

$$\begin{aligned}
\frac{d^2\eta}{dt^2} &= \frac{h^2}{(r_0+\eta)^3} + \frac{1}{m}f(r_0+\eta) \\
&= \frac{h^2}{r_0^3} - 3\frac{h^2}{r_0^4}\eta + \frac{1}{m}f(r_0) + \frac{1}{m}f'(r_0)\eta + O(\eta^2) && \text{(using Taylor series for both expansions)} \\
&= -3\frac{h^2}{r_0^4}\eta + \frac{1}{m}f'(r_0)\eta + O(\eta^2) && \text{(using (3.25))} \\
&= \left(3\frac{f(r_0)}{mr_0} + \frac{1}{m}f'(r_0)\right)\eta && \text{(using (3.25) again and ignoring higher order terms in } \eta)
\end{aligned}$$

We can achieve this result a bit more slickly by writing the orbital equation in the form

$$\ddot{r} = G(r),$$

where $G(r) = \dfrac{h^2}{r^3} + \dfrac{1}{m}f(r)$ and $G(r_0) = 0$. Then setting $r = r_0 + \eta$ and taking the first non-zero term of the Taylor series for $G(r_0 + \eta)$ gives

$$\ddot{\eta} = G'(r_0)\eta = \left(-3\frac{h^2}{r_0^4} + \frac{1}{m}f'(r_0)\right)\eta = \frac{1}{m}\left(3r_0^{-1}f(r_0) + f'(r_0)\right)\eta$$

as before.

[8] which could be found by substituting $u = (2 + \cos\frac{1}{2}\theta)^{-1}$ into the left hand side of the geometric orbit equation (3.23), and then expressing it all in terms of r again; whatever appears there is $-f(r)/mh^2$.

[9] We are assuming that the circular orbit is centred on the centre of force, i.e. on the origin of polar coordinate. It is possible to imagine circular orbits centred on some point other than $r = 0$. If we choose the coordinate θ such that the centre of the circle lies on the line $\theta = 0$, the equation of the circle is of the form

$$r^2 - 2rd\cos\theta + d^2 = a^2 \quad \text{i.e.} \quad (d^2 - a^2)u^2 - 2ud\cos\theta + 1 = 0.$$

Solving this quadratic equation to find u as function of $\cos\theta$ and substituting into the geometric orbit equation (3.23) will reveal the force law that would allow such a circular orbit.

Now if $3r_0^{-1}f(r_0) + f'(r_0) < 0$, this equation is simple harmonic; η will stay small and the orbit is stable to radial perturbations. Conversely, if $3r_0^{-1}f(r_0) + f'(r_0) > 0$, one of the solutions of the equation is a growing exponential and the orbit is unstable to radial perturbations.

In the case of a power law force, $f(r) = -kr^n$, where we require $k > 0$ since the force must be attractive,

$$3f(r_0) + r_0 f'(r_0) = -k(3+n)r_0^n$$

so the forces that allow stable circular orbits must have $n > -3$. This includes (fortunately) the inverse square law.

At this point, we can ask an interesting question: which forces $f(r)$ will provide not only stable but also closed orbits? The above equation for the perturbation η is linearised: it will give a necessary and sufficient condition for the orbit to be stable and a necessary condition for the orbit to be closed. Over many orbits, quadratic and higher order terms may become important and it is quite difficult to determine whether these terms will mean that the orbit is not in fact closed. There is a theorem due to Bertrand, which states (using second order perturbations) that the only force laws that permits closed and stable orbits are $f(r) \propto r^{-2}$ (inverse square) and $f(r) \propto r$ (Hooke's law)!

3.5 Orbits in an inverse square force

3.5.1 The orbits as conic sections

Let

$$f(r) = -\frac{mk}{r^2} \equiv -mku^2 \tag{3.26}$$

where

$$k = \begin{cases} GM & \text{gravitational force for a spherical body of mass } M \\ -\dfrac{qQ}{4\pi\epsilon_0 m} & \text{electrostatic force between two point electric charges } q \text{ and } Q. \end{cases} \tag{3.27}$$

Substituting this into the orbital equation (3.23) gives

$$\frac{d^2u}{d\theta^2} + u = \frac{k}{h^2} \tag{3.28}$$

and (magic!) we have an equation for which we can write down the solution using the standard complementary function plus particular integral method:

$$u = A\cos(\theta - \theta_0) + \frac{k}{h^2}, \tag{3.29}$$

where A can be chosen, without loss of generality, to be non-negative by setting $\theta_0 \to \theta_0 + \pi$ if necessary.

The largest value of u, which corresponds to the smallest value of r ('closest approach'), is given by

$$u_{\max} = A + \frac{k}{h^2} \tag{3.30}$$

and, without loss of generality, we choose the θ coordinate such that this corresponds to $\theta = 0$, which is equivalent to setting $\theta_0 = 0$ in the general solution (3.29):

$$u = A\cos\theta + \frac{k}{h^2}. \tag{3.31}$$

Now we set

$$\frac{|k|}{h^2} = \frac{1}{\ell} \tag{3.32}$$

and

$$A = \frac{e}{\ell} \tag{3.33}$$

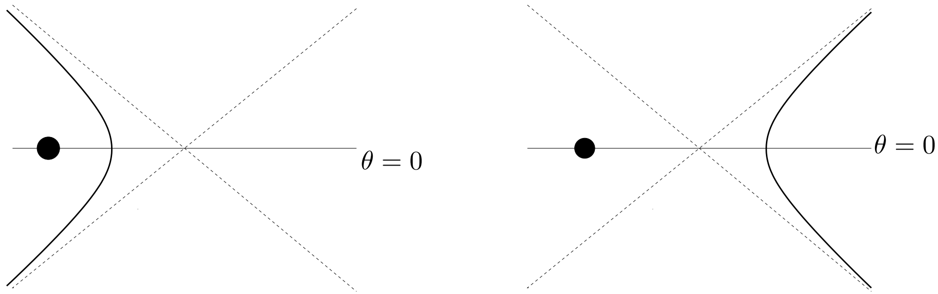
so that the orbital equation (3.30) becomes

$$u = \frac{1}{\ell}(e\cos\theta \pm 1) \qquad \text{or} \qquad r = \frac{\ell}{e\cos\theta \pm 1} \tag{3.34}$$

where the + sign corresponds to attractive forces (gravitation, force between unlike electric charges) and the − sign corresponds to repulsive[10] forces (force between like electric charges) as in the definition (3.27) of k.

In the form (3.34) is recognisable immediately as a conic section,[11] so the orbits are be hyperbolae, ellipses or parabolae as follows.

- Ellipse: $0 \le e < 1$, and the + sign; the circle is a special case corresponding to $e = 0$.
- Parabola: the borderline case $e = 1$.
- Hyperbola: $e > 1$, and either the + sign if the centre of the force is inside the hyperbola (corresponding to an attractive force) or the − sign if the centre of force is outside the hyperbola (corresponding to a repulsive force). The two cases are illustrated in the figures below.



The figures show the hyperbolic trajectories of a particle in an attractive inverse square potential (left-hand figure) and in a repulsive inverse square potential (right-hand figure). In both cases, the solid blob is the centre of the force and also the origin of polar coordinates with $\theta = 0$ corresponding to closest approach as shown. The dashed lines are parallel to the asymptotes.

You don't have to remember whether $e > 1$ corresponds to an ellipse or a hyperbola: it is obvious. The gross distinguishing feature is that an ellipse is bounded (r does not go off to infinity) and a hyperbola is unbounded. If $0 \le e < 1$, the factor $(e\cos\theta \pm 1)$ that appears in the solution (3.34) is non-zero for all θ so the orbit is bounded and hence an ellipse if we take the positive sign, corresponding to an attractive force, and we get nothing if we take the negative sign, because then $r < 0$ (which is expected: we could not imagine an elliptical orbit in a repulsive force).

The asymptotes of the hyperbolic trajectories are determined by $u \to 0$, so $\cos\theta = -1/e$ in the attractive case and $\cos\theta = 1/e$ in the repulsive case.

The total energy per unit mass is all cases is given by

$$\begin{aligned}
E &= \mathrm{KE} + \mathrm{PE} \\
&= \tfrac{1}{2}h^2(u'^2 + u^2) - ku && \text{(using (3.24) and (3.26))} \\
&= \tfrac{1}{2}h^2\big((-A\sin\theta)^2 + (A\cos\theta + k/h^2)^2\big) - k(A\cos\theta + k/h^2) && \text{(from (3.31))} \\
&= \tfrac{1}{2}h^2A^2 - \tfrac{1}{2}k^2/h^2 && (3.35) \\
&= \tfrac{1}{2}\left(\frac{h^2e^2}{l^2} - \frac{h^2}{l^2}\right) = \frac{h^2}{2\ell^2}(e^2 - 1). && \text{(using the definitions (3.33) and (3.32))}
\end{aligned}$$

We see that:

- the energy is constant (of course);
- the energy is positive if $e > 1$, which means that the particle has more than the escape velocity at each point on its trajectory — this corresponds to the hyperbolic orbits;
- the energy is negative if $0 \le e < 1$ meaning that the orbit is bound — this corresponds to the elliptic orbits;
- the energy is zero if $e = 1$, so the particle has exactly the escape velocity at each point on its trajectory — this corresponds to the parabolic orbits.

Thus everything fits together nicely.

[10]'Repulsive', like massive, is a word that physicists use in a very literal sense: a repulsive force is one that repels rather than one that is horrid.

[11]Conic sections are discussed in detail in section 3.10.

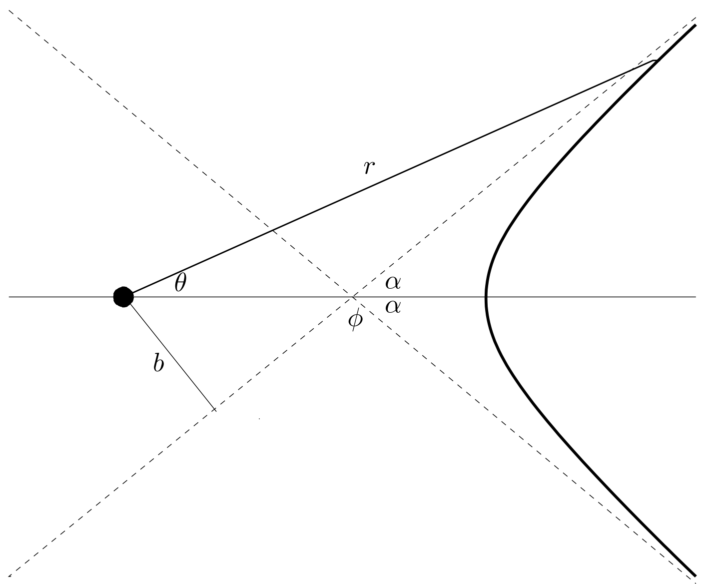
3.5.2 Rutherford scattering

A particle with a positive charge q and mass m moves in the electric field produced by a positive charge Q which is fixed at the origin. The mutual gravitational attraction of the charges is negligible compared with the repulsive electrostatic Coulomb force.

Initially, the particle is approaching from a very large distance (effectively $r = \infty$) at speed V along a path which, in the absence of Q, would pass a distance b from the origin: b is called the *impact parameter*. Our task is to find the angle through which the particle is deflected in terms of the parameters h and b.

The angular momentum per unit mass is given by $h = -Vb$, because $h = \pm|\mathbf{r} \times \dot{\mathbf{r}}|$ (angular momentum is 'moment of momentum' so this is just like taking the moment of a force: you multiply the magnitude of the force by the shortest distance between the line of action of the force and the point). The $\pm$ is to take into account the sense of the angular momentum, i.e. whether it is clockwise or anti-clockwise. In this case, the minus sign is correct because, as can be seen from the diagram, the moment is clockwise; to put it another way, $h = r^2\dot{\theta}$ and $\dot{\theta} < 0$.

Conservation of energy shows that when the particle has been bounced back by the repulsive field of the central charge and is heading out to ∞ its speed tends back to V; then conservation of angular momentum shows that the 'backwards' impact parameter is also b. The trajectory has a reflection symmetry as shown in the diagram.



The figure shows the hyperbolic trajectory of a particle undergoing Rutherford scattering. The distance b is the impact parameter. The angle α is equal to the asymptotic value of θ. The angle ϕ is the angle through which the particle, coming in from the top right, is deflected.

The orbital equation for like electric charges is

$$\frac{d^2u}{d\theta^2} + u = \frac{k}{h^2}\,,$$

where $k = -qQ/4\pi\epsilon_0 < 0$. The solution to this is

$$u = A\cos\theta + \frac{k}{h^2} \tag{3.36}$$

There are two unknowns in this solution, A and h, and these can be expressed in terms of V and b. However, for present purposes we are only trying to find α, which can be done easily by considering the velocity:

$$\begin{aligned}
\dot{\mathbf{r}} &= \dot{r}\,\widehat{\mathbf{r}} + r\dot{\theta}\,\widehat{\boldsymbol{\theta}} && \text{(velocity in polars)}\\
&= \dot{\theta}\frac{dr}{\mathrm{d}\theta}\,\widehat{\mathbf{r}} + \frac{h}{r}\,\widehat{\boldsymbol{\theta}} && \text{(chain rule and } h = r^2\dot{\theta})\\
&= -h\frac{du}{\mathrm{d}\theta}\,\widehat{\mathbf{r}} + hu\,\widehat{\boldsymbol{\theta}} && (u = 1/r)\\
&= hA\sin\theta\,\widehat{\mathbf{r}} + h(A\cos\theta + \frac{k}{h^2})\,\widehat{\boldsymbol{\theta}}\,. && \text{(from (3.36))}
\end{aligned}$$

As $\theta \to \alpha$, $\dot{\mathbf{r}} \to -V\,\widehat{\mathbf{r}}$ (there is a minus sign because the particle is coming inwards from $r = \infty$) so

$$hA\sin\alpha = -V \quad \text{and} \quad A\cos\alpha + \frac{k}{h^2} = 0\,.$$

Eliminating A from these two equations gives

$$\tan\alpha = \frac{Vh}{k} = \frac{-V^2 b}{k}.$$

The angle we are after is angle through which the particle is deflected, which is ϕ on the diagram. We have

$$\begin{aligned} \phi = \pi - 2\alpha \Longrightarrow \tfrac{1}{2}\phi &= \tfrac{1}{2}\pi - \alpha \\ \Longrightarrow \tan\tfrac{1}{2}\phi &= \cot\alpha = \frac{-k}{V^2 b} \end{aligned}$$

so we obtain the pleasingly simple result[12] that

$$\phi = 2\tan^{-1}\left(\frac{|k|}{V^2 b}\right).$$

The same result is obtained for the deflection of, say, an interstellar comet (i.e. not a periodic comet such as Halley's comet) by the sun. For a comet, though, V and b are probably not measurable but the distance d of closest approach to the sun at which $\theta = 0$, and the speed U at this point may be measurable.[13] We can then find h and A from

$$\frac{1}{d} = A + \frac{k}{h^2}; \quad h = dU.$$

3.5.3 Kepler's laws

Kepler[14] formulated his three laws in about 1605, on the basis of remarkably accurate observations by Tycho Brahe.[15]

They may be stated as follows:

K1: Each planet moves on an ellipse with the sun at one focus.

K2: The radius vector to the planet sweeps out equal areas in equal times.

K3: The period of the orbit is proportional to (mean radius)$^{\frac{3}{2}}$.

We have already established K1: the planets move in ellipses as a consequence of the inverse square law of gravitational attraction.[16]

[12]Ernest (later Lord) Rutherford used this result to explain the way that α-particles (positively charged) are scattered by atoms. In a series of experiments on gold leaf, he observed that scattering angles greater than $\pi/2$ were sometimes found, which he calculated was inconsistent with the current plum pudding model due to J.J. Thompson in which the electrons were the plums and the pudding was smeared out positive charge. He realised that his results were consistent with a model of the atom as a small, heavy positively charged nucleus surrounded by orbital electrons. This work was carried out in 1911, eight years before he came to Cambridge as head of the Cavendish laboratory (and three years after he was awarded the Nobel prize for *Chemistry* for his work on the physics of alpha particles.

[13]If one simply replaces V by c, the speed of light, in the deflection angle, one gets a formula that would give the deflection of star light by the sun if Newtonian dynamics applied to photons. Interestingly, Einstein's first effort at General Relativity, in 1911, gave exactly this result. It took into account the principle of equivalence (gravity is indistinguishable from acceleration) but not the fact that space-time is curved by the presence of massive bodies. He was impatient for observers to prove him right, but since the effect is too small to measure unless the light just grazes the sun, observations could only be made at a solar eclipse. Attempts to observe were thwarted first by cloudy conditions and then by the war. By the time of Eddington's expedition to Principe in 1919, Einstein had evolved his full theory of relativity, which gave a prediction of exactly twice the 'Newtonian' value, and this was triumphantly corroborated by Eddington (with suspicious accuracy). It is interesting to speculate how General Relativity would have fared without the hold-up to these observations.

[14]Johannes Kepler (1571-1630)

[15]1546 — 1601. Interesting factoid of the day: While a student, Tycho lost part of his nose in a duel. For the rest of his life, he was said to have worn a realistic replacement made of silver and gold, using a paste to keep it attached. When Tycho's tomb was opened in 24 June 1901 green marks were found on his skull, suggesting the false nose also had copper. It is possible that he had a number of different noses for different occasions.

[16]Curiously, there is only one other power law that gives rise to ellipses, namely $f(r) = r$, which is Hooke's law; but in this case the sun would be at the centre of the ellipse, not at the focus.

K2 is a consequence of the law of conservation of the law of conservation of angular momentum and therefore holds for any central force. This can be seen as follows. Let δA be the area swept out by the radius vector in time δt, during which the polar angle changes by $\delta\theta$. Then

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

by the usual geometrical argument. Thus

$$\frac{dA}{dt} = \tfrac{1}{2} r^2 \dot\theta \equiv \tfrac{1}{2} h.$$

Since h is constant, $\dfrac{dA}{dt}$ is constant, which is the required result.

It is not clear, and it seems that Kepler was not very clear about it either, what precisely he meant by 'mean radius'. He certainly would not have been able to average r over an orbit. Since planetary orbits are more or less circular, it doesn't much matter what it means. However, if we calculate the period of an orbit of a particle subject to an inverse square force, we see that K3 does in fact hold is a perfectly acceptable sense.

In order to calculate the period, T, we need an equation relating θ and t. Equation (3.19) is just the thing. We have

$$\dot\theta = hu^2$$

so

$$\begin{aligned} T \equiv \int_0^T dt &= \int_0^{2\pi} \frac{d\theta}{\dot\theta} = \int_0^{2\pi} \frac{d\theta}{hu^2} \\ &= \int_0^{2\pi} \frac{d\theta}{h(1+e\cos\theta)^2/\ell^2} && \text{(from the geometrical solution (3.34))} \\ &= \frac{\ell^2}{h} \frac{2\pi}{(1-e^2)^{\frac{3}{2}}} && \text{(take my word for this.)} \end{aligned}$$

The result of the last integral is of interest, though actually doing the integral is not.[17]

Now the smallest value of r (at the perihelion) $\theta = 0$, and the largest value of r (at the aphelion) $\theta = \pi$ satisfy (see the geometric form of the orbit (3.34))

$$\tfrac{1}{2}(r_{\text{min}} + r_{\text{max}}) = \tfrac{1}{2}\left(\frac{\ell}{1+e} + \frac{\ell}{1-e}\right) = \frac{\ell}{1-e^2}$$

so

$$T = \frac{2\pi \ell^{\frac{1}{2}}}{h}\left(\tfrac{1}{2}(r_{\text{min}} + r_{\text{max}})\right)^{\frac{3}{2}} = \frac{2\pi}{\sqrt{k}}\left(\tfrac{1}{2}(r_{\text{min}} + r_{\text{max}})\right)^{\frac{3}{2}}$$

using the definition (3.32) to eliminate h. This certainly counts as being proportional to (mean radius)$^{\frac{3}{2}}$.

3.6 Circular orbits and quantum mechanics

This little aside arises because Keppler's laws reminded me of another law, also based on observations, which is also relevant — but not very relevant, first because the law is wrong[18] and second because the effect is so small[19]. I have in mind Bohr's model of the atom, which postulated that the orbital angular momentum of an electron orbiting a nucleus is *quantised*, which means that it is only allowed to take certain values:

$$\text{Angular momentum} \equiv M_e R v = n\hbar\,. \tag{3.37}$$

In this formula, M_e is the mass of the Earth, R is the radius of the Earth's orbit (assumed circular), v is the speed of the Earth, n is an integer and $\hbar$ is Planck's constant (divided by 2π).

[17] The integral is doable by elementary means, starting for example with the substitution $t = \tan\frac{1}{2}\theta$; it takes about half an hour. It is quicker by means of a contour integral in the complex plane; about 10 minutes. Mathematica can do it in 0.47 seconds, but makes a terrible mess of it.

[18] We shouldn't be too discouraged by this: the results are correct even though the theory is wrong.

[19] You may be a bit discouraged when you see the magnitude of the effect

Together with the Newton's second law for a circular orbit:

$$\frac{GM_eM_s}{R^2} = \frac{M_ev^2}{R} \tag{3.38}$$

('gravitational force equals centrifugal force'), equation (3.37) show that R is also quantised; it can only take certain values. Eliminating v from (3.37) and (3.38) gives

$$R = \frac{n^2\hbar^2}{GM_e^2M_s}.$$

Substituting

$$\hbar = 1\times 10^{-34} kg\, m^2\, s^{-1} \tag{3.39}$$
$$G = 6.7\times 10^{-11}\, m^3\, kg^{-1}\, s^{-2} \tag{3.40}$$
$$M_s = 2\times 10^{30}\, kg \tag{3.41}$$
$$M_e = 6\times 10^{24}\, kg \tag{3.42}$$
$$R = 1.5\times 10^{10}\, m \tag{3.43}$$

shows that $n \approx 8\times 10^{70}$. There are therefore 10^{70} allowed values for the radius of orbits closer to the Sun than the Earth's orbit. The allowed orbits are tightly spaced: the difference in radius between the nth and $(n+1)$th allowed orbits is

$$\frac{[(n+1)^2 - n^2]\hbar^2}{GM_e^2M_s} \approx \frac{2n\hbar^2}{GM_e^2M_s} = \frac{2nR}{n^2} \approx 10^{-62}\, km.$$

3.7 Hohmann Transfer

*********Section under construction*************

A very important problem for the Apollo mission[20] was how to transfer the spacecraft from low orbit Earth orbit (the *parking orbit*) to the high Earth orbit of the Moon using as little fuel as possible. The solution to the problem turns out to be a *Hohmann transfer*, in which a roughly elliptical orbit interpolates between the two approximately circular orbits. This requires a burst of engine fire at both extremes of the roughly elliptical orbit, these two manoeuvres being called TransLunar Injection and TransEarth Injection.

To a good approximation, we can regard the roughly elliptical orbit as being a perfect ellipse with the Earth at one focus for most of the orbit and a perfect ellipse with the Moon at one focus for the remainder. The matching between the two ellipses takes place when the gravitational force of the moon is comparable with that of the Earth. The two accelerations are

$$\frac{GM_e}{r_e^2}, \quad \frac{GM_m}{r_m^2}$$

[20] The aim of the Apollo programme was to put a man on the moon. It was conceived in the presidency of Dwight D. Eisenhower but really got going after President Kennedy's 1961 address to a joint session of Congress promising to land a man on the Moon by the end of the decade. This was achieved by Apollo 11 in July 1969. The record for the greatest distance from the Earth is held by the three astronauts of the Apollo 13 mission, which malfunctioned on the outward journey. Instead of entering a bound (elliptical) moon orbit, a small adjustment was made that put the spacecraft into an unbound (hyperbolic) moon orbit which would automatically (without further use of the engines) return to Earth (*free return trajectory*). In the event, a small correction had to be made on the far side of the moon.

where r_e and r_m are the distances from the Earth and Moon respectively, shows that they will be roughly equal when

$$\frac{r_e}{r_m} = \sqrt{\frac{M_m}{M_e}} = 9$$

since $M_e = 81 \times M_m$.

Need numbers for apollo and parameters and the basic transfer equation. Should be able to do it geometrically as well as dynamically.

$$v_1 + \Delta v_1 = \sqrt{2\mu \frac{R_2}{R_1} \frac{1}{R_1 + R_2}}$$

The major axis is obviously $R_1 + R_2$ and the min and max disxtance are R_1 and R_2, which gives

$$\ell = \frac{2R_1 R_2}{R_1 + R_2}, \qquad e = \frac{R_2 - R_1}{R_2 - +_1}$$

3.8 General forces

The general orbital equation (3.23) may be tackled in two ways:

- direct integration of the second order differential equation;
- integration of the first order energy equation

$$\left(\frac{du}{d\theta}\right)^2 + u^2 = \phi(u)$$

  where $2\dfrac{d\phi}{du} = -\dfrac{f(1/u)}{mh^2u^2}$.

We may also find that some problems can be tackled using only the conservation of energy and angular momentum.

The problem with direct integration is that in general (in fact, except in the case of inverse square and inverse cube forces), the differential equation is non-linear in u and therefore very unlikely to have a solution in terms of standard functions. However, there are well-established approximation techniques that can be applied. The problem with the integral

$$\int \frac{du}{\sqrt{\phi(u) - u^2}}$$

arising from the first order equation is that it is again unlikely to yield standard functions and might be hard to approximate. It is exceptionally fortunate that we live in a universe to which the inverse law of gravitational attraction applies (at least approximately)!

3.8.1 Example: general relativistic orbits

The purpose of this example is to show how that presence of a small perturbation to an inverse square law affects elliptical orbits. We will find that the orbits remain almost elliptical but the axes of the ellipse precess (rotate) slowly. We will set this problem in the context of a famous historical calculation that provided a triumph for the newly hypothesised theory of General Relativity.

The solution of Einstein's field equations corresponding to a spherically symmetric gravitating body (the Sun in our case) turns out to be relatively simple.[21] It is possible to write down the geodesic equations, the solutions of which describe the trajectories of free particles, and of light rays. As in the Newtonian case, setting $u = 1/r$ makes the equation tractable, though it is not soluble in terms of elementary functions.[22] The equation is

$$\frac{d^2u}{d\theta^2} + u = \frac{GM}{h^2}\left(1 + \frac{3h^2u^2}{c^2}\right), \tag{3.44}$$

[21]The solution is named after Karl Schwarzschild (1873–1916) who discovered it while serving in the German army on the Russian front in 1915, which was the year in which Einstein first introduced general relativity. Schwarzschild died the year after (though not as a result of his military service).

[22]Multiplying by $du/d\theta$ and integrating gives an energy-like first integral of the form:

$$\tfrac{1}{2}\ \frac{du}{d\theta}^{\ 2} = E - \tfrac{1}{2}u^2 + \frac{GM}{h^2}u + \frac{GM}{c^2}u^3,$$

where E is a constant of integration. Taking square-roots and integrating gives an elliptic integral.

where h is as usual the angular momentum per unit mass. This differs from the Newtonian orbital equation only in the last term and is in the standard form for an orbit in a central force.

One interesting affect of the extra term is that for $h^2 < 12GM/c^2$, there are no circular orbits; this can easily be seen by setting $d^2u/d\theta^2 = 0$ and solving the resulting quadratic equation for u. This contrasts with the Newtonian case where for any angular momentum there is a circular orbit.[23]

For most astrophysical situations, the extra term is small. Nevertheless, planetary orbits have been observed for many centuries and even very small non-Newtonian affects are detectable.

We will calculate the advance of the perihelion of Mercury. We can estimate the magnitude of the extra term in equation (3.44) as follows. The mean radius of the orbit about the Sun is about 6×10^{10} metres. The period is 88 days, which is about 7.6×10^6 seconds, so $\dot\theta = 2\pi(7.6 \times 10^6)^{-1} \approx 10^{-6}$ radians per second. Thus

$$h = r^2\dot\theta \approx 36 \times 10^{20} \times 10^{-6} = 3.6 \times 10^{15} \quad \text{metres squared per second}$$

and

$$\frac{3h^2u^2}{c^2} = \frac{3 \times 3.6^2 \times 10^{30}}{36 \times 10^{20} \times 9 \times 10^{16}} \approx 10^{-7}$$

which is a very small perturbation. We can therefore approximate the solution by iteration.

The plan is to set

$$u(\theta) = u_0(\theta + \lambda u_1(\theta) + \lambda^2 u_2(\theta) + z \cdots ,$$

substitute this expression into the differential equation (3.44). We can do this term by term.

The unperturbed solution is the Newtonian solution

$$u_0(\theta) = \ell^{-1}(1 + e\cos\theta)$$

where $\ell = h^2/GM$.

To obtain the first iteration, we substitute the Newtonian solution $u_0(\theta)$ into the extra term on the right of equation (3.44) (setting $MG = h^2/\ell$):

$$\frac{d^2u_1}{d\theta^2} + u_1 = \frac{1}{\ell} + \frac{3h^2}{\ell^3c^2}(1 + 2e\cos\theta + e^2\cos^2\theta).$$

This we can solve by the usual particular integral/complementary function method. However, it is clear that only the $2e\cos\theta$ term is of interest: this is resonant and will give a non-periodic particular integral, whereas all the other terms are periodic (with periods either 2π or π) and cannot affect the perihelion advance; indeed, they cancel out when averaged over several orbits.

The corresponding term of the particular integral is

$$\frac{\lambda e}{\ell}\theta\sin\theta, \quad \text{where } \lambda = \frac{3h^2}{\ell^2c^2}.$$

Taking into account only this term of the particular integral gives the first iteration:

$$u_0 + \lambda u_1 \approx \ell^{-1}(1 + e\cos\theta + \lambda e\theta\sin\theta) \approx \ell^{-1}(1 + e\cos((1-\lambda))\theta)$$

where for the last approximation we have used $\cos(\lambda\theta) \approx 1$ and $\sin\lambda\theta \approx \lambda\theta$ (and ignored the periodic terms in u_1).

At the perihelions, $\cos(1-\lambda)\theta = 1$. If the first is when $\theta = 0$, then the second is when $(1-\lambda)\theta = 2\pi$, i.e. when $\theta \approx 2\pi(1+\lambda)$. The perihelion advance is therefore $2\pi\lambda$ radians per orbit.

Putting in the data for Mercury gives an advance of 43 arc second per century. Remarkably, it was known several decades before general relativity was formulated that out of a total observed precession of 5000 arc seconds per century, only 43 arc seconds are unexplained by Newtonian effects (such as the influence of other planets).

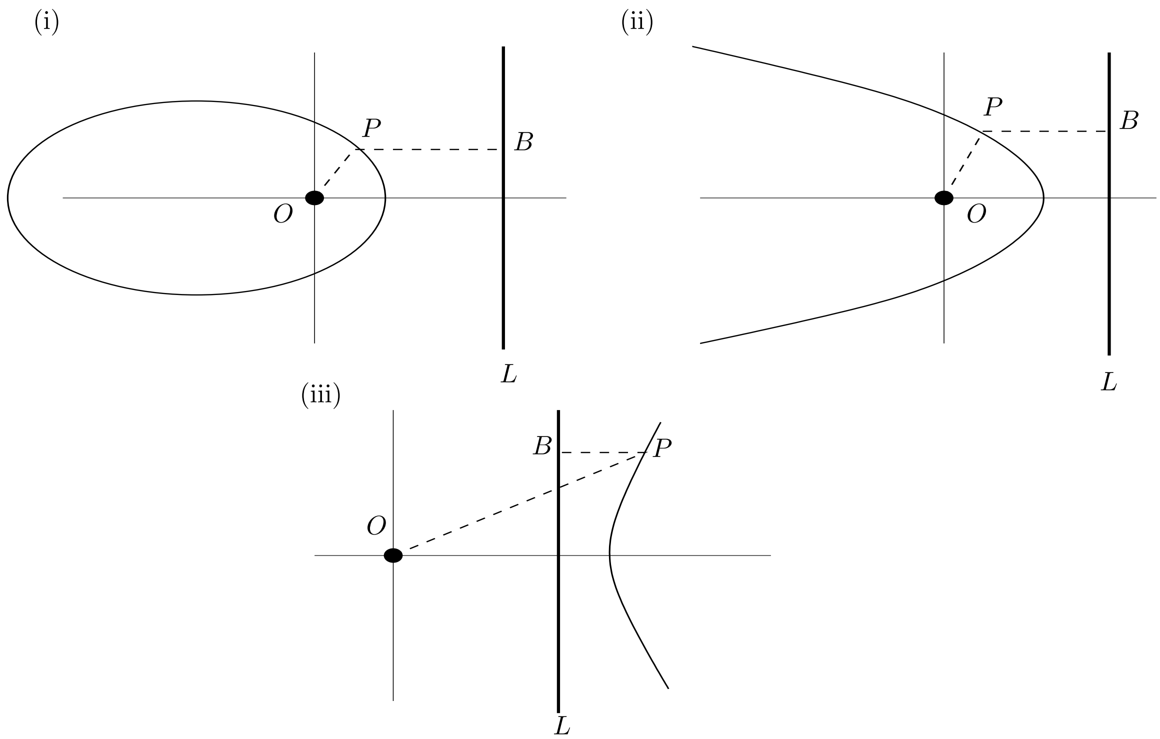
3.9 Conic sections

Conic sections are plane curves formed by the intersections of a plane in $\mathbb{R}^3$ with a double cone. There are three types, ignoring the degenerate cases of a point, a line, and a pair of lines, that arise if the plane passes through the apex of the cones. Suppose that the common axis of the cones is vertical and the semi-angle (the angle between a straight line on the surface of the cone and the axis) is α. Let the acute angle between the normal to the plane and the vertical be $\pi/2 - \theta$. Then the three cases are as follows.

[23] As shown in section **??**, for any attractive force circular orbit of any radius can be found, by choosing h appropriately.

- Ellipses, which are closed curves with $\theta > \alpha$. Circles are special cases with $\theta = \pi/2$.
- Hyperbolae, which are open curves with $\theta < \alpha$. Each consists of two branches, corresponding to the plane intersecting the two cones.
- Parabolae, which are open curves with $\theta = \alpha$.

That is how to picture conics, but it is easier to obtain properties of the conic sections by using a different, two-dimensional, defining property. Let O be the origin of Cartesian coordinates, and let L be a fixed line (called the *directrix* of the conic). The point P moves on a conic if $OP = ePB$, where B is the point on L closest to P and e is a fixed positive number called the *eccentricity* of the conic, as shown in the diagrams.



We can work out the Cartesian and polar equations of conics as follows. Let the equation of the directrix be $x = \ell/e$, where ℓ is a (fixed) positive constant.[24] Then the point (x, y) lies on the conic defined by e and ℓ if

$$x^2 + y^2 = e^2(\ell/e - x)^2$$

It is easy to rearrange this to obtain, after a translation along the x-axis, the three standard forms corresponding to $e < 1$ (ellipse), $e > 1$ (hyperbola) and $e = 1$ parabola:

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$$
$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1$$
$$4x - cy^2 = 0$$

where

$$a^2 = \frac{\ell^2}{(1-e^2)^2}, \qquad b^2 = \frac{\ell^2}{|1-e^2|}, \qquad c = -\ell/2.$$

Much more useful to us are the equations in polar coordinates. Setting $\sqrt{x^2+y^2} = +r$ and $x = r\cos\theta$ gives, in cases (i) and (ii) in the diagram,

$$r = e(\ell/e - r\cos\theta) \qquad \text{i.e.} \qquad r = \frac{\ell}{1 + e\cos\theta} \tag{3.45}$$

[24] ℓ is the length of the line parallel to the directrix from the focus (here, the origin) to the curve; the line is called the *semi latus rectum* from the Latin half + side + straight.

and in case (iii)

$$r = e(r\cos\theta - \ell/e) \qquad \text{i.e.} \qquad r = \frac{\ell}{-1 + e\cos\theta}\,. \tag{3.46}$$

Note that it is easy to distinguish between the closed (and hence bounded) ellipses and the open (and hence unbounded) hyperbolae/parabolae: r can go to infinity only if the denominator of equation (3.45) goes to zero, which can only happen when $e \geq 1$. In the case $e > 1$, the hyperbolae tend to the pair of straight lines (the *asymptotes*) given by $\cos\theta = -1/e$ in case (ii) (given by equation (3.45) and $\cos\theta = 1/e$ in case (iii) (given by equation (3.46)).

3.10 Some useful numbers

Mass of Earth
Mass of Sun
Newton's Gravitational constant (G)
Radius of Earth
Earth's orbital aphelion
Earth's orbital perihelion

Chapter 4

Rotating Frames

4.1 Angular velocity

4.1.1 The concept of angular velocity

Recall that in Chapter 3, we discussed the motion of a particle confined to a circle. In this situation, the concept of linear velocity was replaced with angular velocity ($\dot{\theta}$). The purpose of this section is to extend this concept to the case of more general motion.

The term *angular velocity* can be used in relation either to a particle that is rotating about an axis or to a set of axes that is rotating relative to another set of axes. The meaning of the term is the same in both cases. In the latter case, we might equally think about a rigid body with the axes embedded in it.

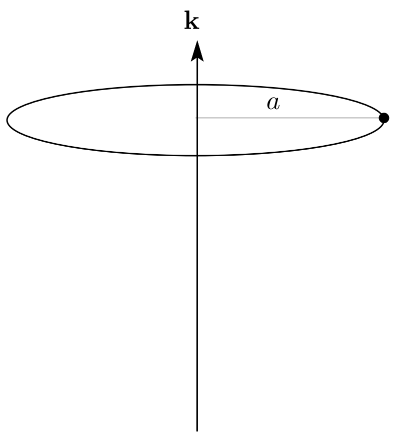
Although we are primarily thinking of a particle rotating about a 'fixed' axis, or axes that are rotating with respect to a 'fixed' set of axes, it doesn't have to be this way; and in any case, we have to define what we mean by 'fixed'. A 'fixed' axis in a laboratory is actually rotating because the Earth is rotating; and even if this is taken into account, as we shall in the later sections of this chapter, there is still the rotation of the Earth about the Sun, and the Sun's rotation about the centre of the Galaxy, and the Galaxy's rotation in the local group, and so on.

The importance of one set of axes being 'fixed' only arises when we come to using Newton's second law, and then the meaning of 'fixed' becomes clear: we mean 'inertial'.

4.1.2 Angular velocity of a rotating particle

We consider the case of a particle rotating about a fixed axis. By a fixed axis, we mean one which is fixed in a given set of axes. Let $\mathbf{k}$ be a unit vector along the axis. By 'rotating about the axis $\mathbf{k}$' we mean that the particle is moving on a circle centred on the axis in a plane whose normal is parallel to $\mathbf{k}$.

Let a be the radius of the circle and let $\dot{\theta}$ be the angular speed of the particle. Then the speed of the particle is $|a\dot{\theta}|$. The direction of the velocity is in the plane perpendicular to $\mathbf{k}$ and, in the diagram below, into the paper.



We will now derive an important coordinate-independent formula for the velocity of the particle. Our first step is to choose cartesian axes, even though our aim is eventually to abandon them. We now assume, for convenience, that our given set of axes are such that $\mathbf{k} = (0, 0, 1)$, with the origin at a point on the axis of rotation a distance h from the plane of motion. In these axes, the position vector of the particle is given by

$$\mathbf{r} = (a\cos\theta,\ a\sin\theta,\ h).$$

We can differentiate this to obtain the velocity of the particle:

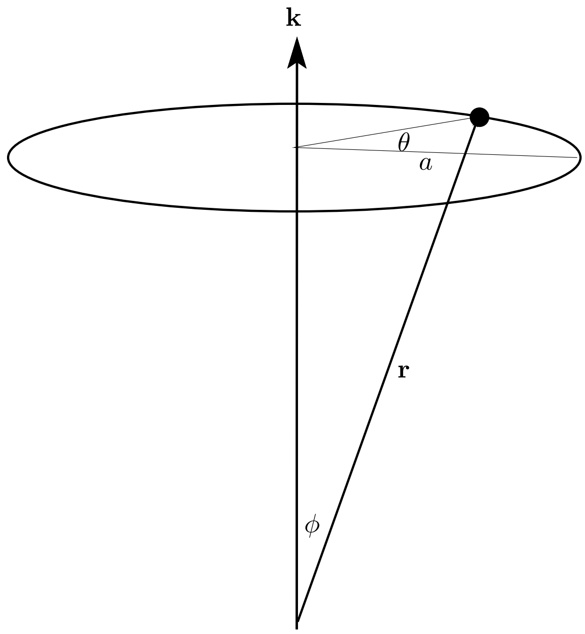
$$\dot{\mathbf{r}} = (-a\dot\theta\sin\theta, a\dot\theta\cos\theta, 0) \equiv \boldsymbol{\omega}\times\mathbf{r} \tag{4.1}$$

where $\boldsymbol{\omega} = \dot\theta\mathbf{k}$. This is the required formula. The vector quantity $\boldsymbol{\omega}$ is called the *angular velocity* of the particle,[1] and its magnitude, $|\dot\theta|$, is the angular speed of the particle.

This all makes sense. The particle is moving anticlockwise round the circle, which is indeed in the direction $\mathbf{k}\times\mathbf{r}$, as can be seen using the right hand rule[2]. One can also think of the direction of motion being a given by the action of a corkscrew[3] in the direction of $\boldsymbol{\omega}$. Furthermore, if ϕ is the angle between $\mathbf{k}$ and $\mathbf{r}$, as show in the diagram below,

$$|\dot{\mathbf{r}}| = |\boldsymbol{\omega}||\mathbf{r}|\sin\phi = a|\dot\theta|$$

as expected.



4.1.3 Rotating axes

Now we consider the case of rotating axes. Let $\mathbf{e}_i$ be a set axis vectors which are which are moving relative to a given (non-rotating) set of axes[4] in such a way that their origins always coincide. Let

[1] Note that the direction of the angular velocity vector is along the axis, not in the direction of the velocity of the particle.

[2] The right hand rule was invented for use in electromagnetism by British physicist Zachariah William Cole in the late 1800s: if you point your thumb along the first vector of a cross product and your index finger along the second vector, your middle finger will point in the direction of the cross product.

[3] The following is an extract from a very engaging footnote in James Clerk Maxwell's book *A treatise on Electromagnetism, vol 1.*

The combined action of the muscles of the arm when we turn the upper side of the right hand outwards, and at the same time thrust the hand forwards, will impress the right-handed screw motion on the memory more firmly than any verbal definition. A common corkscrew may be used as a material symbol of the same relation.

Professor W.H. Miller has suggested to me that as the tendrils of the vine are right-handed screw and those of the hop left-handed, the two systems of relations in space might be called those of the hop and of the vine respectively.

The system of the vine, which we adopt, is that of Linneaus, and that of screw-makers in all civilised countries except Japan. Screws like the hop tendril are used for the couplings of railway carriages, and for the fittings of wheels on the left side of ordinary carriages, but are always called left-handed screws by those who adopt them.

[4] At the back of our minds, we are thinking of the given set of axes as being 'fixed' in some sense — for example, fixed with respect to the laboratory — though as discussed in section 4.1.1 they are not fixed in the sense of motionless. We will later think of these axes as being inertial axes.

P be any point that is fixed in the rotating axes, i.e. has position vector of the form $b_i\mathbf{e}_i$ (remember we are using using summation convention), where each b_i is independent of time. A theorem of Euler states that the instantaneous motion of P relative to the origin is a rotation about an axis through the origin (which is obvious, since P is a fixed distance from the origin) and that the axis is the same for all such points (which is not obvious).

Thus, taking P to be a unit distance along the basis vector $\mathbf{e}_i$,

$$\dot{\mathbf{e}}_i = \boldsymbol{\omega} \times \mathbf{e}_i \tag{4.2}$$

as in (4.1) above, where again $\boldsymbol{\omega}$ is the angular velocity vector. Note that this is the *instantaneous* motion. The motion at each time is of this form, but the axis is in general different for different times; i.e. $\boldsymbol{\omega}$ can be a function of time.

For an arbitrary vector $\mathbf{b} = b_i\mathbf{e}_i$, not necessarily fixed in the rotating axes, we have

$$\dot{\mathbf{b}} = \dot{b}_i\mathbf{e}_i + b_i\dot{\mathbf{e}}_i \tag{4.3}$$

$$= \dot{b}_i\mathbf{e}_i + b_i\boldsymbol{\omega} \times \mathbf{e}_i \tag{4.4}$$

$$= \dot{b}_i\mathbf{e}_i + \boldsymbol{\omega} \times \mathbf{b} \tag{4.5}$$

This is a very important result. The left hand side is the rate of change of the vector $\mathbf{b}$ measured in the inertial frame. The right hand side is the rate of change as seen in the rotating axes plus a contribution due to the motion, underfoot, of the axes.

4.1.4 Velocity in rotating axes

For a concrete example, let $\mathbf{b} = \mathbf{r}$, where $\mathbf{r}$ is the position vector of a particle relative to a fixed origin. In the rotating axes, let $\mathbf{r} = x_i\mathbf{e}_i$, where the components x_i may depend on t. Applying (4.5) gives

$$\dot{\mathbf{r}} = \dot{x}_i\mathbf{e}_i + \boldsymbol{\omega} \times \mathbf{r}. \tag{4.6}$$

The term on the left of this equation is the velocity of the particle in the non-rotating axes. The first term on the right is the velocity of the particle as measured in the rotating axes (which rotate with angular velocity $\boldsymbol{\omega}$ with respect to the non-rotating axes); the second term takes into account the relative motion of the axes.

As we have mentioned many times, if the non-rotating axes are inertial, then the rotating axes are non-inertial and so the transformation relating the two sets of axes is not Galilean (though a *constant* rotation is Galilean). For a Galilean transformation given by $\mathbf{r}' = \mathbf{r} + \mathbf{v}t$, $t' = t$, the velocity $\mathbf{u}$ transforms according to:

$$\mathbf{u}' = \mathbf{u} + \mathbf{v}$$

(just differentiate the transformation law with respect to t or t').

The transformation law for rotating axes is $\mathbf{r}' = R\mathbf{r}$, where R is a time-dependent rotation matrix, and we have shown above that the velocity transforms according to

$$\mathbf{u} = \mathbf{u}' + \boldsymbol{\omega} \times \mathbf{r}'.$$

In section 4.2 we show how R is related to $\boldsymbol{\omega}$, but this is not essential for present purposes.

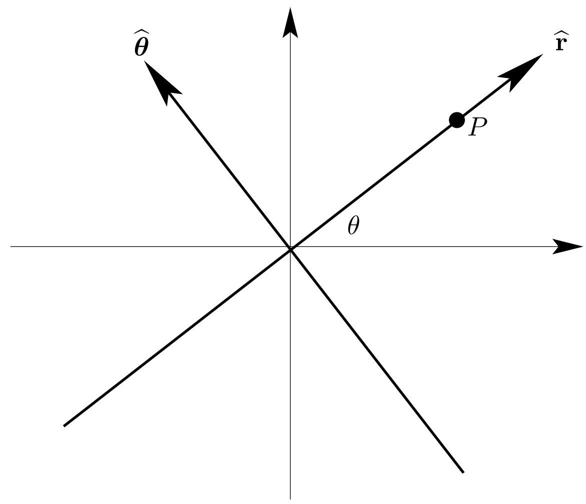
4.1.5 Plane polar coordinates

Plane polar axes provide a rather peculiar example of rotating axes. We consider a particle moving in plane. In this case, the particle always lies (by definition) on one axis, namely the $\widehat{\mathbf{r}}$ axis, and the axes rotate with the position of the particle to ensure that this is so. Since we need the angular velocity is perpendicular to the plane of rotation, we add a third axis in this direction (giving cylindrical polar axes). Thus the position of a particle in plane polar (the rotating) axes is $(r, 0, 0)$, i.e. $r\widehat{\mathbf{r}} + 0\widehat{\boldsymbol{\theta}} + 0\widehat{\mathbf{z}}$, and the velocity *with respect to the rotating axes* is simply $(\dot{r}, 0, 0)$. The axes rotate with angular velocity $\boldsymbol{\omega} = (0, 0, \dot{\theta})$.

The velocity of the particle with respect to the non-rotating axes, using equation (4.6) is

$$(\dot{r}, 0, 0) + (0, 0, \dot{\theta}) \times (r, 0, 0) = (\dot{r}, r\dot{\theta}, 0), \tag{4.7}$$

just as it should be.



4.2 Euler's theorem

This section is interesting, I hope, and relevant; but it is not part of the course.

Let $\mathbf{e}_i$ be axes the origin of which is a fixed point. Then it is clear that the point a unit distance along $\mathbf{e}_1$ satisfies (4.2)

$$\dot{\mathbf{e}}_1 = \boldsymbol{\omega}_1 \times \mathbf{e}_1$$

for some angular velocity vector $\mathbf{e_1}$: the point is a constant distance from the origin so it is instantaneously rotating. Similar equations hold for $\dot{\mathbf{e}}_2$ and $\dot{\mathbf{e}}_3$. These equations do not define the angular velocities uniquely: we could add any multiple of $\mathbf{e}_i$ to $\boldsymbol{\omega}_i$ without changing $\dot{\mathbf{e}}_i$.

Euler's theorem states that the three angular velocity vectors $\boldsymbol{\omega}_i$ are equal in the sense that there is a single angular velocity $\boldsymbol{\omega}$ such that

$$\dot{\mathbf{e}}_i = \boldsymbol{\omega} \times \mathbf{e}_i;$$

this is equation (4.2). It can easily be verified that the orthogonality relations between the basis vectors $\mathbf{e}_i \cdot \mathbf{e}_j = \delta_{ij}$ are preserved by (4.2).

We can obtain (4.2) from a more fundamental starting point (i.e. without quoting Euler's theorem). Suppose that the rotating axes $\mathbf{e}_i$ are related to non-rotating axes $\mathbf{i}_j$ by the linear transformation

$$\mathbf{e}_i = R_{ij}\mathbf{i}_j \tag{4.8}$$

where R is a time-dependent orthogonal matrix (by virtue of the orthonormality of the basis vectors: $\mathbf{e}_j \cdot \mathbf{e_k} = \delta_{jk} = \mathbf{i}_j \cdot \mathbf{i}_k$).

Since R is orthogonal, it satisfies (by definition)

$$RR^T = I,$$

where the superscript T denotes transpose. Differentiating this gives[5]

$$\dot{R}R^T + R\dot{R}^T = 0$$

i.e.

$$\dot{R}R^T + (\dot{R}R^T)^T = 0, \tag{4.9}$$

using the general result $(AB)^T = B^T A^T$. Equation (4.9) says that the matrix $\dot{R}R^T$ is antisymmetric. It follows that we can find a vector $\boldsymbol{\omega}$ such that

$$(\dot{R}R^T)_{ij} = \epsilon_{ijk}\omega_k, \tag{4.10}$$

where $\boldsymbol{\omega} = \omega_k \mathbf{e}_k$, i.e. ω_k are the components of $\boldsymbol{\omega}$ in the rotating frame. We can find the components of $\boldsymbol{\omega}$ explicitly by hitting equation (4.10) with ϵ_{ijm}:

$$\epsilon_{ijm}(\dot{R}R^T)_{ij} = \epsilon_{ijm}\epsilon_{ijk}\omega_k = 2\delta_{mk}\omega_k = 2\omega_m.$$

[5] Note that $\dfrac{d(R^T)}{dt} = \left(\dfrac{dR}{dt}\right)^T$.

We will now show that $\boldsymbol{\omega}$ is precisely the angular velocity of the rotating axes defined above (hence the suggestive use of the same symbol for both). We start by differentiating equation (4.8) and then meander through some rather pleasing vector algebra, as follows.

$$\begin{aligned}
\dot{\mathbf{e}}_i &= \dot{R}_{ij}\mathbf{i}_j && \text{(since } \mathbf{i}_j \text{ are fixed vectors)}\\
&= \dot{R}_{ij}R^{-1}_{jk}\mathbf{e}_k && \text{(inverting (4.8))}\\
&= \dot{R}_{ij}R^{T}_{jk}\mathbf{e}_k && \text{(since } R \text{ is orthogonal)}\\
&= \epsilon_{ikm}\omega_m\mathbf{e}_k && \text{(by (4.10))}\\
&= \epsilon_{ikm}(\boldsymbol{\omega}.\mathbf{e}_m)\mathbf{e}_k && \text{(inverting } \boldsymbol{\omega} = \omega_m\mathbf{e}_m)\\
&= \tfrac{1}{2}\epsilon_{ikm}(\boldsymbol{\omega}.\mathbf{e}_m)\mathbf{e}_k - \tfrac{1}{2}\epsilon_{ikm}(\boldsymbol{\omega}.\mathbf{e}_k)\mathbf{e}_m && (\epsilon_{ikm} \text{ is antisymmetric in } k \text{ and } m)\\
&= \tfrac{1}{2}\epsilon_{ikm}\boldsymbol{\omega}\times(\mathbf{e}_k\times\mathbf{e}_m) && \text{(standard vector identity)}\\
&= \boldsymbol{\omega}\times\mathbf{e}_i && \text{(the vectors } \mathbf{e}_i \text{ form an orthonormal basis)}
\end{aligned}$$

which is the required result.

We can verify this result in the simple case of axes rotating about a fixed axis $\mathbf{k}$, which we take to be $\mathbf{e}_3$. Thus

$$R = \begin{pmatrix} \cos\theta & \sin\theta & 0\\ -\sin\theta & \cos\theta & 0\\ 0 & 0 & 1\end{pmatrix}$$

and

$$\dot{R} = \begin{pmatrix} -\dot{\theta}\sin\theta & \dot{\theta}\cos\theta & 0\\ -\dot{\theta}\cos\theta & -\dot{\theta}\sin\theta & 0\\ 0 & 0 & 0\end{pmatrix}$$

giving

$$\dot{R}R^T = \begin{pmatrix} 0 & \dot{\theta} & 0\\ -\dot{\theta} & 0 & 0\\ 0 & 0 & 0\end{pmatrix}$$

which is indeed anti-symmetric. Furthermore, comparing with the definition of $\boldsymbol{\omega}$ in equation (4.10), we see that in this case

$$\boldsymbol{\omega} = \begin{pmatrix} 0\\ 0\\ \dot{\theta}\end{pmatrix}$$

as expected.

4.3 Newton's second law in rotating axes

We are familiar with rotating axes; at least, we should be, since we live in axes spinning with the Earth[6]. We are also familiar with some effects produced by rotating axes, such as the rotation of hurricanes, anticlockwise in the Northern hemisphere and clockwise in the Southern hemisphere.

Recall that Newton's second law applies specifically in inertial frames, in fact, in exactly those frames for which the first law holds; therefore, definitely not in rotating axes. This means that it is usually best to use frames related by Galilean transformations. Sometimes, it is far more convenient to do the calculations in a non-inertial frame. In such cases, the thing to do is to apply the second law in an inertial frame and then transform to the non-inertial frame.

We start with the basic equation, for a single particle,

$$\mathbf{F} = m\mathbf{a}^{\text{in}}, \tag{4.11}$$

where $\mathbf{a}^{\text{in}}$ is the acceleration relative to the inertial frame, and ask how each term in this equation will appear in a rotating frame.

Taking the easiest term first, we can say that the mass of the particle is an intrinsic property of the particle and it does not matter which frame we measure it in.[7]

[6] Not to mention the Earth's rotation round the Sun, or the Sun's rotation with the galaxy.

[7] In Special Relativity, we have to be more careful: the *rest mass* of the particle (i.e. the mass measured in a frame in which the particle is at rest) is intrinsic to the particle, but the mass in a moving frame depends on the speed of motion.

Next we consider the force term. Since force is a physical vector quantity, it can be written in the form $\mathbf{F} = F_i \mathbf{e}_i$ and, because of the vector nature of force, this holds good in any set of axes. For example, the gravitational force (in Newtonian dynamics[8]) that the particle feels does not depend on the axes it is measured in. [9]

Like force, the position vector from the origin can be written in the form

$$\mathbf{r} = x_i \mathbf{e_i} \tag{4.12}$$

where x_i are the coordinates of the particle and this form also holds good in any axes with the same origin.

However, the velocity of the particle does not transform so easily. In an inertial frame $\mathbf{i}_k$,

$$\mathbf{r} = x_k^{\text{in}} \mathbf{i}_k$$

and

$$\dot{\mathbf{r}} = \dot{x}_k^{\text{in}} \mathbf{i}_k \equiv \mathbf{v}^{\text{in}}$$

which is the velocity with respect to the inertial frame. In a rotating frame $\mathbf{e}_i$, we have (using (4.2))

$$\begin{aligned}\dot{\mathbf{r}} &= \dot{x}_i \mathbf{e}_i + x_i \dot{\mathbf{e}}_i \\ &= \dot{x}_i \mathbf{e}_i + x_i \boldsymbol{\omega} \times \mathbf{e}_i \\ &= \dot{x}_i \mathbf{e}_i + \boldsymbol{\omega} \times \mathbf{r} \\ &\equiv \mathbf{v} + \boldsymbol{\omega} \times \mathbf{r},\end{aligned}$$

where $\mathbf{v}$ is the velocity in the rotating frame. Thus

$$\mathbf{v}^{\text{in}} = \mathbf{v} + \boldsymbol{\omega} \times \mathbf{r}. \tag{4.13}$$

We can differentiate (4.13) to obtain $\mathbf{a}^{\text{in}}$, the acceleration in the inertial frame:

$$\begin{aligned}\ddot{\mathbf{r}} &= \frac{d}{dt}\left(\dot{x}_i \mathbf{e}_i + \boldsymbol{\omega} \times \mathbf{r}\right) \\ &= \ddot{x}_i \mathbf{e}_i + \dot{x}_i \dot{\mathbf{e}}_i + \dot{\boldsymbol{\omega}} \times \mathbf{r} + \boldsymbol{\omega} \times \dot{\mathbf{r}} \\ &= \ddot{x}_i \mathbf{e}_i + \dot{x}_i \boldsymbol{\omega} \times \mathbf{e}_i + \dot{\boldsymbol{\omega}} \times \mathbf{r} + \boldsymbol{\omega} \times (\mathbf{v} + \boldsymbol{\omega} \times \mathbf{r}) \\ &= \mathbf{a} + 2\boldsymbol{\omega} \times \mathbf{v} + \dot{\boldsymbol{\omega}} \times \mathbf{r} + \boldsymbol{\omega} \times (\boldsymbol{\omega} \times \mathbf{r}).\end{aligned}$$

The first of these terms is the acceleration $\mathbf{a}$ relative to the rotating frame.

In the inertial frame,

$$\ddot{\mathbf{r}} = \ddot{x}_k^{in} \mathbf{i}_k = \mathbf{a}^{in}$$

so

$$\mathbf{a}^{in} = \mathbf{a} + 2\boldsymbol{\omega} \times \mathbf{v} + \dot{\boldsymbol{\omega}} \times \mathbf{r} + \boldsymbol{\omega} \times (\boldsymbol{\omega} \times \mathbf{r}). \tag{4.14}$$

Finally, using Newton's second law, we have

$$\mathbf{a} = \mathbf{F}/m - 2\,\boldsymbol{\omega} \times \mathbf{v} - \dot{\boldsymbol{\omega}} \times \mathbf{r} - \boldsymbol{\omega} \times (\boldsymbol{\omega} \times \mathbf{r}) \tag{4.15}$$

i.e.

$$m\mathbf{a} = \mathbf{F} - 2m\,\boldsymbol{\omega} \times \mathbf{v} - m\dot{\boldsymbol{\omega}} \times \mathbf{r} - m\boldsymbol{\omega} \times (\boldsymbol{\omega} \times \mathbf{r}) \tag{4.16}$$

It should be emphasised that here $\mathbf{v}$ and $\mathbf{a}$ are the velocity and the acceleration *relative to the rotating frame*. The other terms on the right hand side of (4.16) are called *fictitious forces*: they have no physical origin but, relative to the rotating frame, the particle moves as if under the action of such forces. In particular, the $-2m\,\boldsymbol{\omega}\times\mathbf{v}$ term is called the *coriolis force*[10] and the $-m\boldsymbol{\omega}\times(\boldsymbol{\omega}\times\mathbf{r})$ term is called the *centrifugal force*.

[8] In General Relativity, gravitational forces do depend on axes; in fact, gravitational forces are a facet of the choice of axes and this is the basic idea behind the theory.

[9] If the rotating axes are related to an inertial set $\mathbf{i}_k$ via $\mathbf{e}_i = R_{ik}\mathbf{i}_k$, then

$$\mathbf{F} = F_i \mathbf{e}_i = F_k^{\text{in}} \mathbf{i}_k$$

then

$$F_i R_{ik} = F_k^{\text{in}}$$

which is the tensor transformation law for vectors.

[10] Gaspard-Gustave de Coriolis was a French mathematician, mechanical engineer and scientist. As well as studying motion in rotating frames, he was the first to use the term 'work' in the sense of force times distance moved by force.

You quite often find people asking the question ‘Does centrifugal force really exist?’, to which the answer is the obvious ‘No, it is fictitious.’ A slightly better answer is that in an inertial] frame, there is never any need to talk about centrifugal force. It arises, only in rotating frames, as an artifact to make it appear that Newton’s second law holds in this non-inertial frame: it is the force that would be required to make N2 hold.

The fictitious forces can be understood as follows. Suppose that there is no physical force acting. Then the particle moves in a straight line with constant speed in the inertial frame. In the rotating frame, the particle does not move in a straight line, so it seems to be moving under the influence of forces (the fictitious forces in equation (4.16)) which cause it to deviate from a straight line path. Alternatively, if the particle is moving with constant velocity, zero velocity say, with respect to the rotating axes, then it is accelerating in the inertial frame which requires a physical force to act. Since the left hand side of equation (4.16) is zero in this situation, the physical force must exactly balance the fictitious force.[11]

To give another example, suppose you are in a car going round a corner. In the inertial (non-car) frame, you are accelerating inwards and the force to do this is provided by the car door pushing against you (without which you would be flung out of the car). In this frame, nothing need or should be said about fictitious forces. In the rotating frame of the car, you are not moving, but you still feel the physical force of reaction from the car door. Of course, Newton’s second law does not apply in this non-inertial frame; nevertheless, you can still use ‘force = mass times acceleration’ if you introduce a fictitious force which will exactly balance the force from the car door; this is the outward acting centrifugal force.

Similarly if the car door opened there would no longer be a constraint to prevent you from moving in a straight line in the inertial frame. In the frame of the car, you are not moving in a straight line: you are accelerating outwards and this can be accounted for (if again you want to pretend that Newton’s law holds) by the action of a centrifugal force.

We now consider each of the additional terms on the right side of equation (4.15).

4.4 Variable angular velocity: $-\dot{\boldsymbol{\omega}} \times \mathbf{r}$

The origin of this term lies in the acceleration of the axes. It is analogous to the force you feel in an accelerating lift. We won’t need to worry about this term, because the only $\boldsymbol{\omega}$ we will be using is that of the Earth, which is more or less constant and we will take it to be constant. But if you are interested, read on.

The Earth’s axis precesses once round a fixed axis every 25,800 years.[12] This motion is due to the action of the Moon and the Sun on the Earth. It is only present because the Earth is not a perfect sphere: there is flattening at the poles and bulging at the equator caused by its spin about its own axis.

It is instructive to see exactly under what conditions $\dot{\boldsymbol{\omega}}$ can be neglected. For simplicity, we compare with the centrifugal term. We can neglect precession provided

$$|\dot{\boldsymbol{\omega}} \times \mathbf{r}| \ll |\boldsymbol{\omega} \times (\boldsymbol{\omega} \times \mathbf{r})\,|.$$

Roughly, then, we need

$$|\dot{\boldsymbol{\omega}}| \ll |\boldsymbol{\omega}|^2. \tag{4.17}$$

Taking

$$|\boldsymbol{\omega}| = 2\pi \text{ radians per day} = 365 \times 2\pi \text{ radians per year}$$

and

$$|\dot{\boldsymbol{\omega}}| = 2\pi/25,800 \text{ radians per year,}$$

the condition (4.17) becomes

$$\frac{2\pi}{25,800} \ll (365 \times 2\pi)^2,$$

[11]For example, for a pendulum in deep space seen from the rotating frame in which the bob is stationary and the pivot moves, the tension (acting inwards towards the pivot) in the rod exactly balances the fictitious centrifugal force acting outwards.

[12]This precession was apparently discovered by Hipparchus in about 150 B.C., which, given the extremely long period, is astonishing. According to Ptolemy, Hipparchus compared his measurements of the longitudes of certain bright stars with measurements made by other Greek astronomers a century previously and concluded that they had moved by about a degree (i.e. a *tiny* amount). On this basis, he predicted a period of about 36,000 years.

This precession is what is rather mysteriously referred to in Rudyard Kipling’s Just So story, the Elephant’s Child: ‘One fine morning in the middle of the Precession of the Equinoxes this ’satiable Elephant’s Child asked a new fine question that he had never asked before.’ Some of you will know the story, but if you don’t just ignore this comment.

which most certainly holds.

The Earth's axis also *nutates* — i.e. it *nods* — on a much shorter timescale, with period 18.6 years (which is the same period as the precession of the Moon). The amplitude of this effect is only about 20 arc seconds, so

$$|\dot{\boldsymbol{\omega}}| \approx \frac{2\pi \times 20}{360 \times 60^2 \times 18.6} \text{ radians per year}$$

and again the effect is negligible.

Finally, the Earth's angular velocity changes because the Earth's rotation is slowing down: it loses energy through tidal friction. The rate of slowing is very variable, but averages roughly 1.4 milliseconds per day per century, i.e. the length of the day increases by 1.4 milliseconds per century. This implies that

$$\frac{|\dot{\boldsymbol{\omega}}|}{|\boldsymbol{\omega}|} \approx \frac{1.4 \times 10^{-3}}{24 \times 60 \times 60 \times 100}$$

which shows that compared with the centrifugal effect this slowing down is very small indeed.[13]

4.5 Coriolis term: $-2\boldsymbol{\omega} \times \mathbf{v}$

The Coriolis force is the hardest of the three fictitious forces to visualise. It acts in the plane perpendicular to $\boldsymbol{\omega}$, so we can confine our attention to motion on a rotating disc.

Consider the case of a free particle (no physical force acting: the disc is smooth and there are no gravitational or other forces). In the inertial frame the trajectory of a particle is a straight line. The key is to realise that as the particle moves along the straight line, its distance from the axis of rotation is inevitably changing. Suppose that this distance is increasing. As the particle moves outwards, the relative speed between it and the underlying rotating disc increases, because the tangential speed ($r\dot{\theta}$) of the disc increases. Thus, relative to the non-inertial frame of the rotating disc, the particle is accelerating — the coriolis acceleration. If the disc is rotating clockwise, the acceleration required is anticlockwise. In this case, $\boldsymbol{\omega}$ points out of the paper, $\mathbf{v}$ is radially outwards so $-2\boldsymbol{\omega} \times \mathbf{v}$ is in the anticlockwise tangential direction, as expected.

Thus one aspect of coriolis force can be seen to arise from movement towards the axis of rotation to point where the speed relative to the inertial frame is smaller, or away from the axis of rotation to a point where the speed relative to the inertial frame is larger.

But why is there a factor or 2? The reason is that the coriolis acceleration has two equal components which can be understood as follows. Suppose that you are on a round-about at the fair, or in a children's playground, which is rotating at constant angular speed $\dot{\theta}$. You stand in the middle and walk radially outward, along a radius the rotating round-about, at speed v and without slipping. You experience a physical force, namely friction acting on your shoes and this force balances the coriolis force so that in the rotating frame of the roundabout the tangential acceleration vanishes. Two effects contribute to the coriolis force.[14]

1. At a distance r from the axis, your tangential velocity relative to the inertial frame is that of the disc, namely $r\dot{\theta}$. When you increase the distance by dr, your tangential speed must increase by $(dr)\dot{\theta}$. This happens in time dt, where $dr = v\,dt$, so that acceleration is

$$\frac{\dot{\theta}dr}{dr/v} = v\dot{\theta} = |\boldsymbol{\omega} \times \mathbf{v}|. \tag{4.18}$$

2. Meanwhile, relative to the inertial frame, the point where your foot *was* has moved a distance $r\dot{\theta}dt$ and the point you will place your foot has moved a distance $(r + dr)\dot{\theta}dt$. To keep on the same radius in the rotating frame you have to have a tangential speed of $r\dot{\theta}$ and a tangential acceleration, additional to that in equation (4.18), of $\dot{r}\dot{\theta} = v\dot{\theta}$.

To see that these effects are different, just consider jumping from r to $r + dr$. You would still feel a tangential jolt when you landed (just as if you had jumped out of the door of a moving car),

[13] However, it does mean that since the year 1820 in which the second was defined to be 1/86,400 of a day, the length of the day has increased by about 2 milliseconds. Thus every 500 days, a difference of about one second accumulates between the Earth day and the 24-hour day. The Earth day is taken to be standard (it is not very easy to adjust!), so a leap second is added to atomic clocks. This last happened in December 2008 and before that in December 2005. The time between these leap seconds was more than 500 days because the Earth sped up mysteriously.

[14] Note to self: could do with a picture.

allowing you to pick up speed, but you would not land on the same radius because the roundabout would have moved on under your feet while you were in the air.

But this is not the whole story. Clearly, there is also a coriolis affect if the velocity of the particle in the rotating frame is tangential not radial: in this case $-2\boldsymbol{\omega}\times\mathbf{v}$ is radially outward if $\mathbf{v}$ is in the direction of motion of the rotation (i.e. a right-hand screw with respect to $\boldsymbol{\omega}$). How does this arise? One way of understanding it is that it corresponds to an additional centrifugal force necessitated by the additional tangential motion.[15]

4.5.1 Example: coriolis force and cyclones

In meteorology, a cyclone is a closed circular area of rotating air. Normally, the air pressure in the centre (the 'eye') is comparatively low, so the air-flow is inward. In the Northern hemisphere, the angular velocity vector $\boldsymbol{\omega}$ has components in the Northerly and radially outward direction, so the coriolis force ($-2\boldsymbol{\omega}\times\mathbf{v}$) due to the Earth's rotation is anticlockwise, and the air spirals anticlockwise towards the centre. In the Southern hemisphere, the components $\boldsymbol{\omega}$ are Northerly and radially inward, so of the spiral is clockwise.

It is easy to check what the direction of the spirals must be using the right hand rule, but how should this be understood? The answer is that coriolis force is all about moving towards or away from the axis of rotation; in the Northern hemisphere, decreasing distance to the axis means moving towards the North pole, which is in the general direction[16] of $\boldsymbol{\omega}$ whereas in the Southern hemisphere, decreasing distance to the axis means moving towards the South pole.

Note that the coriolis force only plays a significant role when the timescale is long, for example in meteorological effects or in long-distance gunnery.[17]

4.5.2 Example: coriolis effect on a falling particle

We consider the effect of the coriolis force on a particle dropped from a fixed point in the rotating frame of the Earth — the top of a tower, say (as in Galileo's experiment). In the rotating frame, the appropriate equation of motion is (4.15), but omitting the centrifugal and varying angular velocity terms:[18]

$$\ddot{\mathbf{r}} = \mathbf{g} - 2\boldsymbol{\omega}\times\dot{\mathbf{r}} \tag{$*$}$$

where $\mathbf{r}$ and its derivatives are all relative to the rotating frame. We can integrate ($*$) directly once:

$$\dot{\mathbf{r}} - \dot{\mathbf{r}}_0 = \mathbf{g}t - 2\boldsymbol{\omega}\times\mathbf{r} + 2\boldsymbol{\omega}\times\mathbf{r}_0, \tag{$**$}$$

where $\mathbf{r}_0$ is the initial position (the point from which the particle is dropped) and $\dot{\mathbf{r}}_0$ is the initial velocity in the rotating frame. We are considering a dropped particle, so we take $\dot{\mathbf{r}}_0 = 0$.

We could at this point simply lurch into components and integrate the system of first order equations using methods from the Differential Equations course but, because we are already ignoring terms of $O(\omega^2)$ by omitting the centrifugal acceleration, we may as well be consistent. Substituting ($**$) into ($*$) gives

$$\ddot{\mathbf{r}} = \mathbf{g} - 2\boldsymbol{\omega}\times(\mathbf{g}t - 2\boldsymbol{\omega}\times(\mathbf{r}-\mathbf{r_0}))$$

[15] Note to self: provide details — not so easy!

[16] To be more precise, in the direction of the component of $\boldsymbol{\omega}$ that is tangent to the Earth's surface

[17] It definitely does not apply on the scale of a bath or in a small bowl of water 20 metres from the equator (or anywhere else): see http://www.youtube.com/watch?v=ZU1EB0z-h7w for an extremely convincing demonstration of how to extract money from gullible tourists (and also, in a video from his television series, from Michael Palin). Of course, the demonstrator stirs the water in the right direction as he puts the twig in — or is it the right direction: you should check!

[18] Both these terms can in fact be neglected unless the tower is enormous. We have already discussed the varying angular velocity of the Earth. Comparing the coriolis and centrifugal effects, and taking the height h of the tower to be 10m, and the radius of the Earth to be $6.4\times10^6 m$ and $g = 10$m/sec/sec gives:

$$\frac{\text{centrifugal effect}}{\text{coriolis effect}} \approx \frac{\omega^2 R}{hg\omega} = \frac{\omega R}{hg} \approx \frac{(2\pi/86400)\times 6.4\times10^6}{10\times10} \approx 4.5$$

which is not small. However, the accelerations are not in the same direction: at the equation, the centrifugal force is radial and the coriolis force for a falling particle is roughly horizontal.

Instead, we compare the centrifugal acceleration with the acceleration due to gravity:

$$\frac{\text{centrifugal effect}}{\text{gravitational effect}} = \frac{\omega^2 R}{g} = \frac{(2\pi/86400)^2\times(6.4\times10^6)}{10} \approx 0.003$$

which is negligible.

and ignoring the last term $-2\boldsymbol{\omega} \times (\mathbf{r} - \mathbf{r_0})$, (which is small compared with $\mathbf{g}t$), we obtain

$$\ddot{\mathbf{r}} = \mathbf{g} - 2\boldsymbol{\omega} \times \mathbf{g}t.$$

The right hand side of this equation can be integrated twice directly:

$$\mathbf{r} = \tfrac{1}{2}\mathbf{g}t^2 - \tfrac{1}{3}\boldsymbol{\omega} \times \mathbf{g}t^3 + \mathbf{r}_0.$$

Now at last we take choose axes. We will assume for simplicity that our tower is at the equator. Our axial directions at the top or bottom of the tower are, as usual:
$\mathbf{e}_3$ radially outwards;
$\mathbf{e}_2$ northwards;
$\mathbf{e}_1$ easterly;
which form a right-handed set.

With respect to these axes,

$$\mathbf{g} = (0, 0, -g), \quad \boldsymbol{\omega} = (0, \omega, 0), \quad \mathbf{r}_0 = (0, 0, R + h), \quad \mathbf{r} = (x, y, z)$$

where R is the radius of the Earth and h is the height of the tower (above the surface of the Earth). Thus

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \tfrac{1}{2}gt^2 \begin{pmatrix} 0 \\ 0 \\ -1 \end{pmatrix} + \tfrac{1}{3}\omega g t^3 \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ R + h \end{pmatrix} \tag{†}$$

To this approximation, surface of the Earth is $z = 0$.[19] Substituting $z = R$ into the third component of the (†) reveals that the descent time is $\sqrt{2h/g}$, as in the non-rotating case. At this time,

$$x = \frac{2\sqrt{2}}{3} \frac{\omega h^{3/2}}{g^{1/2}} \tag{4.19}$$

which is the distance Eastwards from the bottom of the tower at which the particle lands.[20]

This can easily be understood in the inertial (non-rotating) frame. Just before being dropped, the particle is at radius $(R+h)$ and co-rotating, so it has speed $(R+h)\omega$ and angular momentum per unit mass $(R+h)^2\omega$. As it falls, its angular momentum is conserved (the only force is central), so its final speed v in the (Eastward) direction of rotation satisfies $Rv = (R+h)^2\omega$, and $v == (R+h)^2\omega/R$. Since this is larger than the speed $R\omega$ of the foot of the tower, the particle gets ahead of the tower.

The horizontal velocity relative to the tower is approximately $2h\omega$ (ignoring the h^2 term), so the average relative speed over the fall is about $h\omega$. We now see that the displacement (4.19) can be expressed in the form (time of flight) times (average relative velocity) as might be expected.

4.5.3 Foucault pendulum

The Foucault pendulum is just a simple pendulum with its pivot fixed in the rotating frame of the Earth. Ours consists of a weight of mass m suspended by a string of length l.

First we write down the equation of motion (4.15):

$$[m\ddot{\mathbf{r}} + 2m\omega \times \dot{\mathbf{r}} = -T\widehat{\mathbf{r}}_p + m\mathbf{g} \tag{4.20}$$

where $\mathbf{r}$ is the position vector of the mass with respect to an arbitrary (so far) origin, $\widehat{\mathbf{r}}_p$ is the unit vector from the pivot to the mass and T is the tension in the pendulum string. The direction of the force on the mass due to tension is from the mass to the pivot (hence the minus sign in the term $-T\widehat{\mathbf{r}}_p$ in the equation of motion (4.20)), but its magnitude can only be found by solving the equations of motion. We have ignored both the centrifugal ($\boldsymbol{\omega} \times (\boldsymbol{\omega} \times \mathbf{r})$) contribution and the $\dot{\boldsymbol{\omega}} \times \mathbf{r}$ contribution, both being small in this context.

Next we choose our axes. We take the origin to be the centre of the Earth, and the z axis radially outwards. The x-y plane is therefore parallel to the plane that is tangent to the surface of the Earth at the location of the pendulum. We take the x-axis into the paper and the y-axis in the plane of the paper, so that in the tangent plane these point East and North, respectively.[21] In these axes, the coordinates of point of suspension are $(0, 0, R)$ (where R is the radius of the Earth

[19] The $z = 0$ plane is in fact tangent to the surface so for a very high tower we would have to take into account the curvature of the surface of the Earth to find the value of z that corresponded to hitting the ground again.

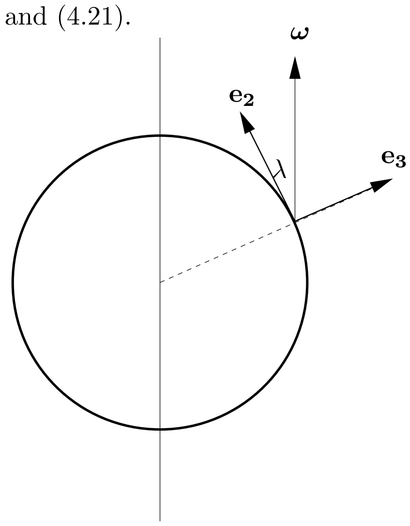
[20] About 55cm for a particle dropped from the Burj Dubai.

[21] Check that they form a right-handed set.

plus a little bit extra), and the coordinates of the mass are $(x, y, R+z)$, where $\widehat{\mathbf{r}}_p = (x, y, z)$ and $z < 0$ because the mass hangs below the pivot. Note the constraint

$$x^2 + y^2 + z^2 = l^2, \tag{4.21}$$

so that $(x/l, y/l, z/l)$ is a unit vector pointing from the point of suspension to the mass. We therefore have four unknowns x, y, z and T, and four equations (4.20) and (4.21).



In these axes, we have

$$\boldsymbol{\omega} = (0, \omega\cos\lambda, \omega\sin\lambda),$$
$$\mathbf{g} = (0, 0, -g)$$

where λ is the latitude. (Remember that $\boldsymbol{\omega}$ points in the direction of the *axis* of rotation and that the usual polar coordinate θ is given by $\theta = \pi/2 - \lambda$). We can write $\mathbf{T}$ in the form

$$\mathbf{T} = -(x, y, z)T/l$$

where T is the magnitude of the tensional force.

$$m\ddot{x} = -Tx/l - 2\omega(\dot{z}\cos\lambda - \dot{y}\sin\lambda) \tag{i}$$
$$m\ddot{y} = -Ty/l + 2\omega\dot{x}\sin\lambda \tag{ii}$$
$$m\ddot{z} = -mg - Tz/l - 2\omega\dot{x}\cos\lambda \tag{iii}$$

These three equations, together with the constraint (4.21) provide four equations for the four unknowns x, y, z and T.[22]

If we assume, as is the case in practice,[23] that $x/l \ll 1$ and $y/l \ll 1$, then, from the constraint (4.21), we have

$$\frac{z}{l} = -\left(1 - \frac{x^2+y^2}{l^2}\right)^{\frac{1}{2}} = -\left(1 - \frac{x^2+y^2}{2l^2} + \cdots\right) \qquad \text{(Taylor series)}$$

so z/l differs from -1 by second order quantities (x/l is first order); $\dot{z}$ and $\ddot{z}$ are second order quantities and, as we have already assumed in ignoring the centrifugal forces, ω is first order. Thus, ignoring second order quantities, equation (iii) above becomes

$$T = mg.$$

Substituting this into equations (i) and (ii), and again ignoring second order quantities,[24] gives

$$\ddot{x} = -\frac{g}{l}x + 2\omega\dot{y}\sin\lambda, \tag{$*$}$$
$$\ddot{y} = -\frac{g}{l}y - 2\omega\dot{x}\sin\lambda. \tag{$**$}$$

Now let $\zeta = x + iy$. Taking $(*) + i(**)$ gives

$$\ddot{\zeta} = -\frac{g}{l}\zeta - 2i\omega\dot{\zeta}\sin\lambda.$$

This is the equation of a damped harmonic oscillator, but in the complex variable ζ. Setting $\zeta = e^{\alpha t}$ gives the auxiliary equation

$$\alpha^2 + 2i\omega\alpha\sin\lambda + \frac{g}{l} = 0,$$

[22]If some of the signs in the above equations seem odd, it is because we have chosen axes such that $z < 0$ (the pendulum is hanging down and z is measured upwards).

[23]Foucault's original pendulum, which hangs in the Pantheon in Paris, is 67 metres long and the amplitude of its swing is a few metres. In 1851, it provided the first dynamical evidence that the Earth rotates. There is a Foucault pendulum in the Science Museum in London, and another in the United Nations headquarters in Geneva.

[24]The following calculations are a bit involved; you should aim to understand the method and not attempt to learn the details.

i.e.,

$$\alpha = -i\omega\sin\lambda \pm \sqrt{-\omega^2\sin^2\lambda - \frac{g}{l}} \approx -i\omega\sin\lambda \pm i\sqrt{g/l}.$$

The general solution is therefore

$$\zeta = e^{-i\omega t\sin\lambda}\left(A\cos(\sqrt{g/l}\,t) + B\sin(\sqrt{g/l}\,t)\right).$$

This is not quite as straightforward as it looks, because A and B are in general both complex. Without the ω term, the bob moves on an ellipse, as can be seen by taking real and imaginary parts and eliminating the parameter t. However, we can easily see the effect of the ω term, by writing the equation in the form

$$\zeta e^{i\omega t\sin\lambda} = A\cos(\sqrt{g/l}\,t) + B\sin(\sqrt{g/l}\,t)$$

which shows that the with-ω solution is that same as the without-ω solution except that the argand diagram for ζ, which is just the x-y plane, is rotating in the negative sense, i.e. *clockwise* viewed from above,[25] at a constant angular frequency of $\omega\sin\lambda$. Thus the axes of the ellipses rotate on account of the rotation of the Earth. This is easily pictured if the pendulum is suspended at the North pole, when, regarded from an inertial frame, the Earth just rotates under the pendulum; or if it is suspended at the equator when the rotation of the Earth has no coriolis effect. At intermediate positions, the situation is more difficult to picture.

This rotation is relatively slow: its period is $2\pi/(\omega\sin\lambda)$, and $\omega = 2\pi/(1\,\text{day})$, so at $\lambda = 48^\circ\,52'$N (Paris) the period is around 32 hours.

An aside on the Foucault pendulum

The above results can be obtained, without solving any equations, and without approximation. I will argue that the effect is entirely geometric: instead of thinking of the spinning Earth, we can consider a pendulum moving along a line of latitude (or any other path) on a fixed non-rotating sphere. The plane of the pendulum rotates if the path is not a great circle[26] The total phase shift is proportional to the amount the path differs from a great circle, which can be calculated by approximating the path by a sequence of very small great circles and measuring the angles where these great circles meet. The sum of all these angles is (modulo 2π) the phase change of the plane of the pendulum.

Of course, this is no help unless there is a good way of calculating the sum of all these very small angles; and there is. The Gauss Bonnet theorem says that it is proportional to the area of the surface of the sphere enclosed by the path: a very simple calculation for a line of latitude, which gives $2\pi(1-\sin\lambda)$.

4.6 Centrifugal term: $-\boldsymbol{\omega}\times(\boldsymbol{\omega}\times\mathbf{r})$

This is just the familiar outwards force your feel when you go round a corner in a car, dressed up in unfamiliar vector notation. In the diagram, $\boldsymbol{\omega}\times\mathbf{r}$ points into the paper[27] and therefore $-\boldsymbol{\omega}\times(\boldsymbol{\omega}\times\mathbf{r})$ points outwards from the axis. Its magnitude is $\omega^2 r\sin\phi$, where ϕ is the angle between $\mathbf{r}$ and $\boldsymbol{\omega}$; we can write this as the expected expression $\omega^2 d$ where d is the distance from the axis.

The centrifugal force is conservative (omitting the mass):

$$\boldsymbol{\nabla}\times[\boldsymbol{\omega}\times(\boldsymbol{\omega}\times\mathbf{r})] \equiv \boldsymbol{\nabla}\times[(\boldsymbol{\omega}\cdot\mathbf{r})\,\boldsymbol{\omega} - (\boldsymbol{\omega}.\boldsymbol{\omega})\,\mathbf{r}] = (\boldsymbol{\nabla}(\boldsymbol{\omega}\cdot\mathbf{r}))\times\boldsymbol{\omega} - (\boldsymbol{\omega}\cdot\boldsymbol{\omega})\boldsymbol{\nabla}\times\mathbf{r}$$

and both of these terms vanish.[28]

The centrifugal force can therefore be written in terms of the gradient of a scalar using (again omitting mass)[29]:

$$-\boldsymbol{\omega}\times(\boldsymbol{\omega}\times\mathbf{r}) \equiv -\tfrac{1}{2}\boldsymbol{\nabla}\ ((\boldsymbol{\omega}\times\mathbf{r})\cdot(\boldsymbol{\omega}\times\mathbf{r}))$$

[25] It would be anticlockwise in the southern hemisphere.

[26] Lines of longitude are great circles, but the only line of latitude that is a great circle is the equation.

[27] Note to self: need a diagram

[28] Recall that $\boldsymbol{\nabla}(\boldsymbol{\omega}.\mathbf{r}) = \boldsymbol{\omega}$.

[29] This identity is easily checked by expanding:

$$(\boldsymbol{\omega}\times\mathbf{r})\cdot(\boldsymbol{\omega}\times\mathbf{r} = \omega^2 r^2 - (\boldsymbol{\omega}\cdot\mathbf{r})^2.$$

then differentiating.

A particle rotating with the Earth will therefore experience a force $\mathbf{F}(\mathbf{r})$ due to the sum of the gravitational and centrifugal potentials:

$$\mathbf{F}(\mathbf{r}) = -\boldsymbol{\nabla}\Phi(\mathbf{r}); \qquad \text{where} \qquad \Phi(\mathbf{r}) = -\frac{GM}{r} + \tfrac{1}{2}(\boldsymbol{\omega}\cdot\mathbf{r})^2$$

This force is called the *apparent gravitational force*. It given the direction in which a plumb line will hang at the point $\mathbf{r}$, which is in general not parallel with the radial vector. The surface of the Earth is approximately an equipotential of $\Phi(\mathbf{r})$; if it were perfectly calm, the sea would be exactly an equipotential of the apparent gravitational potential, though the GM/r part would need some modification to take into account the fact that the Earth is not in fact spherically symmetrical.[30] The shape of the Earth is in fact roughly an oblate spheroid, i.e., a sphere flattened at the poles and bulging at the equator, which is the shape one would expect a spinning fluid to assume.[31]

The magnitude of the apparent gravitational force determines the period of a pendulum: the Frenchman Jean Richter (d. 1696) found out that his pendulum clock was two minutes slow at the Equator compared to Paris time. The effect of the centrifugal force is to reduce gravity: $g' < g$, where g' is the magnitude of the apparent gravitational acceleration. The period of a simple pendulum of length l is $2\pi\sqrt{l/g'}$, so the period is greater at the equator. This means that the rate of ticking is reduced and the clock does indeed run slow.

4.6.1 Example: apparent gravity

We will find the angle between the radial vector from the centre of the Earth and a particle at rest suspended by a string from a fixed point at latitude λ.

In the rotating frame of the Earth, the equation of motion of the motionless particle is

$$\mathbf{0} = \mathbf{T} + m\mathbf{g} - m\boldsymbol{\omega} \times (\boldsymbol{\omega} \times \mathbf{r}),$$

where $\mathbf{T}$ is force due to the tension in the string, which is of course a vector pointing in the direction of the string.

We choose the usual axes:

$\mathbf{e_1}$: East
$\mathbf{e_2}$: North
$\mathbf{e_3}$: radially outwards.

In these axes,

$$\mathbf{g} = (0, 0, -g), \quad \mathbf{r} = (0, 0, R), \quad \boldsymbol{\omega} = (0, \omega\cos\lambda, \omega\sin\lambda),$$

and

$$\boldsymbol{\omega} \times (\boldsymbol{\omega} \times \mathbf{r}) = R\omega^2\cos\lambda(0, \sin\lambda, -\cos\lambda)$$

as can be verified by calculating the cross products.

The angle between the string and $\mathbf{e_3}$ can be found from the ratio of the components of $\mathbf{T}$:

$$\text{angle between apparent gravity and radial direction} = \tan^{-1}\frac{\omega^2 R\cos\lambda\sin\lambda}{g - \omega^2 R\cos^2\lambda}.$$

This is zero at the poles ($\lambda = \pm\pi/2$) and at the equator ($\lambda = 0$).

This is a small angle. It is certainly less than $\omega^2 R/2g$, which comes from taking $\sin\lambda\cos\lambda = \frac{1}{2}$ (its largest value) and ignoring the additional term in the denominator. Putting in approximate values gives

$$\frac{1}{2}\frac{\omega^2 R}{g} \approx \frac{1}{2}\left(\frac{2\pi}{60\times60\times24}\right)^2 \times \left(6.4\times10^6\right) \times \frac{1}{10} \approx 0.0016$$

in radians.

[30] This is a complicated business: the gravitational potential depends on the shape of the Earth, and the shape of the Earth depends on the gravitational potential.

[31] In the early eighteenth century, there was a dispute between the scientific academies of London and Paris about whether the Earth was shaped like a lemon or like an orange. To settle the dispute Pierre Louis Moreau de Maupertuis, a member of the French Academy succeeded in persuading King Louis XV to send expeditions both to the Equator in Peru and to the arctic circle in Northern Sweden to measure the length of a degree of latitude.

4.7 Fictitious forces: a summary

This is just to confirm what we already know about the origin and direction of the two main fictitious forces.

First, the centrifugal force. Imagine you are sitting on a roundabout in a children's playground. The roundabout rotates anticlockwise, so $\boldsymbol{\omega}$ points upwards. You are sitting still in the rotating frame of the roundabout, so in the inertial frame of the playground you are accelerating away from straight line motion towards the centre of the roundabout. The inward force that provides this acceleration is the friction on your clothes. In the rotating frame you are not moving, so there must be a force to balance the frictional force and this is the outward centrifugal force.

Second, the coriolis force. Imagine you on the same roundabout walking slowly, or shuffling radially (in the rotating frame) outwards. Because the tangential speed of the roundabout increases as you move outwards, you are accelerating, relative to the inertial frame, in the direction of motion of the roundabout, i.e. to your left. Relative to the rotating frame, you are not accelerating so there must be a force to your right, and this is the coriolis force.

Chapter 5

Systems of particles

So far, we have discussed only the application of Newton's laws to particles, or to bodies in situations where they can be treated as particles — planets in orbit round the Sun, for example.

In this chapter, we apply Newton's laws to systems of interacting particles, which could be as simple as two particles moving in each other's gravitational field (for example, a planet moving round the Sun that it is large enough for it not be a good approximation to regard the Sun as fixed) or as complicated as a solid body comprising many billions of atoms or a multi-component system such as a rocket.

Luckily, we do not have to analyse the motion of the individual particles: there are useful general results that give nearly all the information we need. Examples are conservation of momentum, suitably defined, and '$F = Ma$', where M is the total mass, a the acceleration of the centre of mass and F the sum (suitably defined) of the external forces.

The situation is analogous to the theory of perfect gases. We could in principle, but not in practice, calculate the individual motions of the 10^{23} atoms in a mole of the gas, but we don't have to, because almost everything we need can be obtained from relations between thermodynamics variables pressure, volume, temperature and entropy.

5.1 Equations of motion

We consider a system of n particles. The ith particle has mass m_i, position vector $\mathbf{r}_i$ [1] relative to some inertial frame and momentum $\mathbf{p}_i$, where[2]

$$\mathbf{p}_i = m_i \dot{\mathbf{r}}_i.$$

Two sorts of force act on the ith particle: an external force due to some force field independent of our system (gravity, for example) and $n-1$ internal forces (e.g. gravitational or electromagnetic) due to the other particles in the system. The internal forces hold the system together, possibly but not necessarily, rigidly. We denote the external force on the ith particle by $\mathbf{F}_i^e$, and the force on the ith particle due to the jth particle by $\mathbf{F}_{ij}$. Note that

$$\mathbf{F}_{ij} = -\mathbf{F}_{ji} \tag{5.1}$$

by Newton's third law.

Newton's second law holds for each individual particle:

$$\frac{d\mathbf{p}_i}{dt} = \mathbf{F}_i^e + \sum_{j \neq i} \mathbf{F}_{ij}. \tag{5.2}$$

and these are the equations of motion of the system. However, they are not much use in this form, at least if n is large, and in the following sections we explore ways of obtaining the essence of the motion without having to integrate n second-order differential equations.

[1] Careful! The subscript labels particles not components of vectors; you really need to underline your vectors so as not to get in a muddle.

[2] If we were considering relativistic dynamics rather than Newtonian dynamics (see section 6.7), then we should use

$$\mathbf{p}_i = m_i \gamma_i \dot{\mathbf{r}}_i,$$

where m_i is rest mass and $\gamma_i = \left(1 - \dot{\mathbf{r}}_i \cdot \dot{\mathbf{r}}_i / c^2\right)^{-\frac{1}{2}}$.

5.1.1 Momentum of the system

We start by defining the total momentum of the system, which will help us understand how that system as a whole responds to external forces.

So far, we have not encountered the concept of momentum except for a single particle. Does it make sense to ask what is the total momentum of two particles? If so, how should it be defined? Since we are free to define the momentum of the system as we wish, we will obviously choose a definition that results in nice equations; maybe something like Newton's second law.

We define the *total momentum* $\mathbf{P}$ of the system of particles, in the obvious way,[3] as the sum of the individual momenta:

$$\mathbf{P} = \sum_{i=1}^{n} \mathbf{p}_i \tag{5.3}$$

We wish to investigate the way the total momentum varies (aiming for something like Newton's second law), so it is a good plan to differentiate it:

$$\begin{aligned}
\frac{d\mathbf{P}}{dt} &= \frac{d}{dt}\sum_{i=1}^{n} \mathbf{p}_i && \text{(by definition of } \mathbf{P}\text{)}\\
&= \sum_{i=1}^{n} \frac{d\mathbf{p}_i}{dt}\\
&= \sum_{i=1}^{n} \left(\mathbf{F}_i^e + \sum_{j\neq i} \mathbf{F}_{ij} \right) && \text{(using Newton's second law (5.2))}\\
&= \mathbf{F}^e + \sum_{i=1}^{n}\sum_{j\neq i} \mathbf{F}_{ij} && \text{(we define } \mathbf{F}^e \text{ as sum of external forces)}\\
&= \mathbf{F}^e + \sum_{i=1}^{n}\sum_{j\neq i} \tfrac{1}{2}\left(\mathbf{F}_{ij} - \mathbf{F}_{ji}\right) && \text{(by Newton's third law (5.1))}\\
&= \mathbf{F}^e && \text{(because the double sum is symmetric in } i \text{ and } j\text{)}
\end{aligned}$$

Thus

$$\frac{d\mathbf{P}}{dt} = \mathbf{F}^e \tag{5.4}$$

This is a pleasing result, exactly analogous to Newton's second law for a single particle, provided the total force is suitably defined (as the sum of the external forces). In particular, if $\mathbf{F}^e = \mathbf{0}$,

$$\frac{d\mathbf{P}}{dt} = \mathbf{0}$$

so if the external forces on the individual particles sum to zero, the total momentum is conserved.

If the external force is gravitational, so that

$$\mathbf{F}_i^e = m_i \mathbf{g}$$

and

$$\mathbf{F}^e = \sum_{i=1}^{n} m_i \mathbf{g} = M\mathbf{g},$$

where M is the total mass of the system, then

$$\frac{d\mathbf{P}}{dt} = M\mathbf{g}\,. \tag{5.5}$$

In this case, the rate of change of total momentum is governed by the total mass rather than by the individual masses.

[3] This might seem a bit random. How do we know what the definition of total momentum should be; or indeed if any such definition would prove useful? In chapter 2, I mentioned (in a footnote) that the underlying symmetry of the theory gives rise, by Noether's theorem to conserved quantities. The same is true for a system of particles. The translation symmetry $\mathbf{r}_i \to \mathbf{r_i} + \mathbf{a}$ (same $\mathbf{a}$ for each i) implies that there is a conserved momentum corresponding to the sum of the individual momenta.

We can go further. We define the *centre of mass*, $\mathbf{R}$, of the system as a weighted average:

$$M\mathbf{R} = \sum_{i=1}^{n} m_i \mathbf{r}_i. \tag{5.6}$$

The origin of the name 'centre of mass' comes from the action of a uniform gravitational field $\mathbf{g}$ on the system. The total moment of the force about the origin (which is an arbitrary fixed point) is

$$\sum_{i=0}^{n} (\mathbf{r}_i \times (m_i \mathbf{g})) = \mathbf{R} \times (M\mathbf{g})$$

so the moment of the gravitational force can be thought of as acting at the point $\mathbf{R}$; one could place a pivot at this point and the system would balance (since the moment of the force about the pivot would be zero). system

Differentiating equation (5.6) gives

$$M\frac{d\mathbf{R}}{dt} = \frac{d}{dt}\sum_{i=1}^{n} m_i \mathbf{r}_i, = \sum_{i=1}^{n} m_i \dot{\mathbf{r}}_i = \mathbf{P} \tag{5.7}$$

and

$$M\frac{d^2\mathbf{R}}{dt^2} = \mathbf{F}^e. \tag{5.8}$$

Thus the acceleration of the centre of mass is determined directly by the total external force and the centre of mass of the system moves exactly as a single particle of mass M in a force field $\mathbf{F}^e$.

If the total external force is zero, the centre of mass moves with constant speed. If the external force is gravitational, then

$$M\frac{d^2\mathbf{R}}{dt^2} = \sum_{i=1}^{n} (m_i \mathbf{g}) = M\mathbf{g}$$

5.1.2 Angular momentum

Having obtained a generalisation of Newton's second law to systems of particles, we can now apply the same technique to angular momentum, hoping to generalise the single particle 'rate of change of angular momentum equals torque' result.

To save writing, we consider angular momentum about the origin; having obtained the result, the transformation $\mathbf{r} \to \mathbf{r} - \mathbf{a}$ will yield the corresponding result for the angular momentum about any fixed point $\mathbf{a}$.

We define the *total angular momentum* of the system $\mathbf{H}$ about a fixed origin to be the sum of the angular momenta of the individual particles:

$$\mathbf{H} = \sum_{i=1}^{n} \mathbf{r}_i \times \mathbf{p}_i. \tag{5.9}$$

Now we differentiate:

$$\begin{aligned}
\frac{d\mathbf{H}}{dt} &= \sum_{i=1}^{n} \frac{d\mathbf{r}_i}{dt} \times \mathbf{p}_i + \sum_{i=1}^{n} \mathbf{r}_i \times \frac{d\mathbf{p}_i}{dt} && (5.10)\\
&= \sum_{i=1}^{n} \dot{\mathbf{r}}_i \times (m_i \dot{\mathbf{r}}_i) + \sum_{i=1}^{n} \mathbf{r}_i \times \left(\mathbf{F}_i^e + \sum_{j\neq i} \mathbf{F}_{ij}\right) && \text{(using Newton's second law (5.2))}\\
&= \mathbf{0} + \mathbf{G} + \sum_{i=1}^{n}\sum_{j\neq i} \mathbf{r}_i \times \mathbf{F}_{ij} && (5.11)
\end{aligned}$$

where $\mathbf{G}$ is the *total external torque* defined by $\mathbf{G} = \sum\limits_{i=1}^{n} \mathbf{r}_i \times \mathbf{F}_i^e$.

If, as is the case for a system of gravitating masses, the internal forces are central, so that

$$\mathbf{F}_{ij} \parallel (\mathbf{r}_i - \mathbf{r}_j) \tag{5.12}$$

we can simplify further. For convenience, we define

$$\mathbf{F}_{11} = \mathbf{F}_{22} = \cdots = \mathbf{F}_{nn} = \mathbf{0}.$$

Then

$$\begin{aligned}
\sum_{i=1}^{n}\sum_{j\neq i}\mathbf{r}_i \times \mathbf{F}_{ij} &= \sum_{i=1}^{n}\sum_{j=1}^{n}\mathbf{r}_i \times \mathbf{F}_{ij} \\
&= \sum_{j=1}^{n}\sum_{i=1}^{n}\mathbf{r}_j \times \mathbf{F}_{ji} && \text{(relabelling the suffices)} \\
&= -\sum_{j=1}^{n}\sum_{i=1}^{n}\mathbf{r}_j \times \mathbf{F}_{ij} && (\mathbf{F}_{ij} \text{ is antisymmetric}) \\
&= -\sum_{i=1}^{n}\sum_{j=1}^{n}\mathbf{r}_j \times \mathbf{F}_{ij} && \text{(exchanging order of sumations)} \\
&= \tfrac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}(\mathbf{r}_i - \mathbf{r}_j) \times \mathbf{F}_{ij} && (\mathbf{F}_{ij} \text{ is antisymmetric}) \\
&= \mathbf{0} && (\text{since } \mathbf{F}_{ij} \text{ is central (5.12)})
\end{aligned}$$

To obtain the penultimate equation, we added the right hand sides of the first line and the fourth line.

Thus for central internal forces, we have the result

$$\frac{d\mathbf{H}}{dt} = \mathbf{G}, \tag{5.13}$$

just as in the case of a single particle; in particular, if $\mathbf{G} = \mathbf{0}$, the total angular momentum of the system is conserved.

If the external force is a uniform gravitational field, we obtain another very important result:

$$\begin{aligned}
\mathbf{G} &= \sum_{i=1}^{n}\mathbf{r}_i \times \mathbf{F}^e \\
&= \sum_{i=1}^{n}\mathbf{r}_i \times (m_i\mathbf{g}) \\
&= \left(\sum_{i=1}^{n} m_i\mathbf{r}_i\right) \times \mathbf{g} \\
&= M\mathbf{R} \times \mathbf{g} \\
&= \mathbf{R} \times (M\mathbf{g}).
\end{aligned} \tag{5.14}$$

Thus the total external torque acts as if all the mass of the system were concentrated at the centre of mass.

5.1.3 Energy

How should we define the kinetic energy of the system? As before, we assume that it is additive and accordingly define the *total kinetic energy* of the system, T, by

$$T = \sum_{i=0}^{n}\tfrac{1}{2}m_i\dot{\mathbf{r}}_i \cdot \dot{\mathbf{r}}_i. \tag{5.15}$$

Rather less clear is how to define the potential energy of the system. We assume that the internal force between any two particles is central (acts along the line $\mathbf{r}_i - \mathbf{r}_j$) and depends only on the distance between the two particles, as in the gravitational case. In this case, the force can be derived from a potential, and we assume that the potential function is the same for all pairs of particles. We denote the potential at $\mathbf{r}_i$ due to a particle at $\mathbf{r}_j$ is $\phi(r_{ij})$, so that the force experienced by a particle at $\mathbf{r}_i$ due to the particle at $\mathbf{r}_j$ is

$$\mathbf{F}_{ij} = -\boldsymbol{\nabla}_i\phi(r_{ij}) \tag{5.16}$$

Here, the subscript i on the $\boldsymbol{\nabla}$ means that the derivative is with respect to the components of $\mathbf{r}_i$ (it does not denote the ith component, and in what follows summation convention does not apply) and

$$r_{ij} = |\mathbf{r}_i - \mathbf{r}_j)| = r_{ji}.$$

Note that

$$\boldsymbol{\nabla}_i\phi(r_{ij}) = \frac{d\phi(r_{ij})}{dr_{ij}}\boldsymbol{\nabla}_i r_{ij} = -\boldsymbol{\nabla}_j\phi(r_{ij}).$$

The second equality follows from the chain rule. where ...[4]

5.1.4 The centre of mass frame

Sometimes, it is helpful to use the centre of mass as the origin of coordinates. This can lead to simplifications in the equations and a greater level of understanding, but we have to be careful: the centre of mass may be accelerating (in fact, it must if $\mathbf{F}^e \neq 0$), so the new axes may not be inertial.

The main results of this section show that the motion of the system can be very conveniently decomposed as the motion of the particles with respect to the centre of mass and the independent motion of the centre of mass. The motion of the centre of mass can be regarded as that of a single particle of mass M, the total mass of the system.

We denote the position of the ith particle with respect to the centre of mass $\mathbf{R}$ by $\mathbf{y}_i$, so that

$$\mathbf{r}_i = \mathbf{R} + \mathbf{y}_i. \tag{5.17}$$

Thus

$$\sum_{i=1}^n m_i\mathbf{y}_i = \sum_{i=1}^n m_i\mathbf{R} - \sum_{i=1}^n m_i\mathbf{r_i} = M\mathbf{R} - M\mathbf{R} = \mathbf{0} \tag{5.18}$$

and similarly

$$\sum_{i=1}^n m_i\dot{\mathbf{y}}_i = \mathbf{0} \tag{5.19}$$

This last equation says that the momentum in the centre of mass frame, $\mathbf{P}_M$ is zero:

$$\mathbf{P}_M \equiv \sum_{i=1}^n m_i\dot{\mathbf{y}}_i = \mathbf{0} \tag{5.20}$$

and, for this reason, the centre of mass frame is sometimes called the *centre of momentum frame*. Note that equation (5.20) does not imply that the total external force on the system vanishes in the centre of mass frame; it just emphasises the point that Newton's second law, in the form (5.4) holds generally only in inertial frames.

We obtain yet another very useful result by considering the angular momentum with respect to the centre of mass $\mathbf{H}_M$. We can relate $\mathbf{H}_M$ to $\mathbf{H}$ as follows:

$$\begin{aligned}
\mathbf{H}_M &\equiv \sum_{i=1}^n m_i\mathbf{y}_i \times \dot{\mathbf{y}}_i \\
&= \sum_{i=1}^n m_i(\mathbf{r}_i - \mathbf{R}) \times (\dot{\mathbf{r}}_i - \dot{\mathbf{R}}) \\
&= \sum_{i=1}^n m_i\mathbf{r}_i \times \dot{\mathbf{r}}_i - \sum_{i=1}^n m_i\mathbf{R} \times \dot{\mathbf{r}}_i - \sum_{i=1}^n m_i\mathbf{r}_i \times \dot{\mathbf{R}} + \sum_{i=1}^n m_i\mathbf{R} \times \dot{\mathbf{R}} \\
&= \mathbf{H} - \mathbf{R} \times \sum_{i=1}^n m_i\dot{\mathbf{r}}_i - \left(\sum_{i=1}^n m_i\mathbf{r}_i\right) \times \dot{\mathbf{R}} + M\mathbf{R} \times \dot{\mathbf{R}} \\
&= \mathbf{H} - \mathbf{R} \times (M\dot{\mathbf{R}}) - M\mathbf{R} \times \dot{\mathbf{R}} + M\mathbf{R} \times \dot{\mathbf{R}} \qquad \text{(using 5.6 twice)} \\
&= \mathbf{H} - M\mathbf{R} \times \dot{\mathbf{R}}.
\end{aligned}$$

In words, this says that the angular momentum $\mathbf{H}$ about the origin is equal to the angular momentum $\mathbf{H}_M$ about the centre of mass plus the angular momentum about the origin of a particle of total mass M situated at the centre of mass.

[4] Note to self: this is a bit complicated. Is it really necessary? It's OK just to do it for the two-particle case in the next section.

Now we consider the rate of change of angular momentum about the centre of mass. At first sight, it seems that we can simply change the origin in equation (5.13) to $\mathbf{R}$, obtaining

$$\frac{d\mathbf{H}_M}{dt} = \mathbf{G}_M, \tag{5.21}$$

where

$$\mathbf{H}_M = \sum_{i=1}^{n} \mathbf{y}_i \times (m_i \dot{\mathbf{y}}_i)$$

and

$$\mathbf{G}_M = \sum_{i=1}^{n} \mathbf{y}_i \times \mathbf{F}_i^e.$$

But one cannot simply change the origin to $\mathbf{R}$ because the derivation of equation (5.13) required the use of Newton's second law, which in turn assumed that the frame was inertial, whereas the centre of mass accelerates if the total external force is non-zero (see equation (5.8)). Nevertheless, very pleasingly, it turns out that the result (5.21) is correct, as can easily be seen. We have

$$\begin{aligned}
\frac{d\mathbf{H}_M}{dt} &= \sum_{i=1}^{n} m_i \mathbf{y}_i \times \ddot{\mathbf{y}}_i \\
&= \sum_{i=1}^{n} m_i \mathbf{y}_i \times \left(\ddot{\mathbf{r}}_i - \ddot{\mathbf{R}}\right) \\
&= \sum_{i=1}^{n} m_i \mathbf{y}_i \times \ddot{\mathbf{r}}_i - \left(\sum_{i=1}^{n} m_i \mathbf{y}_i\right) \times \ddot{\mathbf{R}} \\
&= \sum_{i=1}^{n} \mathbf{y}_i \times (m_i \ddot{\mathbf{r}}_i) && \text{(using (5.18))} \\
&= \sum_{i=1}^{n} \mathbf{y}_i \times \left(\mathbf{F}_i^e + \sum_{j \neq i} \mathbf{F}_{ij}\right). && \text{(using (5.2))}
\end{aligned}$$

In the case when the internal forces are central, so that $\mathbf{F}_{ij} \parallel (\mathbf{r}_j - \mathbf{r}_i) = (\mathbf{y}_j - \mathbf{y}_i)$, we can repeat exactly the calculations that lead to equation (5.13) to obtain the required result

$$\frac{d\mathbf{H}_M}{dt} = \sum_{i=1}^{n} \mathbf{y}_i \times \mathbf{F}_i^e \equiv \mathbf{G}_M. \tag{5.22}$$

In the external force is a uniform gravitational field,

$$\mathbf{G}_M = \sum_{i=1}^{n} \mathbf{y}_i \times \mathbf{F}_i^e = \sum_{i=1}^{n} \mathbf{y}_i \times (m_i \mathbf{g}) = \left(\sum_{i=1}^{n} m_i \mathbf{y}_i\right) \times \mathbf{g} = 0$$

by (5.18), so the system rotates freely about its centre of mass.

The kinetic energy of the system comes out nicely in the centre of mass frame. Using (5.17), we have

$$\begin{aligned}
T &= \sum_{i=0}^{n} \tfrac{1}{2} m_i \left(\dot{\mathbf{y}}_i + \dot{\mathbf{R}}\right) \cdot \left(\mathbf{y}_i + \dot{\mathbf{R}}\right) \\
&= \sum_{i=0}^{n} \tfrac{1}{2} m_i \dot{\mathbf{y}}_i \cdot \dot{\mathbf{y}}_i + \tfrac{1}{2} \sum_{i=1}^{n} m_i \dot{\mathbf{R}} \cdot \dot{\mathbf{R}} + \dot{\mathbf{R}} \cdot \sum_{i=1}^{n} m_i \dot{\mathbf{y}}_i \\
&= \sum_{i=0}^{n} \tfrac{1}{2} m_i \dot{\mathbf{y}}_i \cdot \dot{\mathbf{y}}_i + \tfrac{1}{2} M \dot{\mathbf{R}} \cdot \dot{\mathbf{R}}
\end{aligned} \tag{5.23}$$

where, for the last equality, we have used (5.19).

In words, equation (5.23) says that the total kinetic energy can be thought of as the total kinetic energy in the centre of mass frame plus the kinetic energy of the total mass positioned at the centre of mass. This result holds in all circumstances: we have not assumed that the centre of mass frame is inertial.

5.1.5 Example: drum majorette's baton

A drum majorette's baton is a short heavy stick which the drum majorette throws spinning into the air and, if all goes well, catches again.

We model the baton as a light rod of length ℓ with masses m_1 and m_2 attached firmly to the ends. What happens when the stick is thrown up into the air?[5]

Let $\mathbf{y}_1$ and $\mathbf{y}_2$ be the position vectors of the two masses with respect to the centre of mass. Then

$$m_1\mathbf{y}_1 + m_2\mathbf{y}_2 = 0$$

(see equation (5.6)).

Setting $|\mathbf{y}_i| = y_i$, we have $y_1 + y_2 = \ell$ and (from the above equation)

$$m_1 y_1 = m_2 y_2 \tag{5.24}$$

The external force on the system is the uniform gravitational field $\mathbf{g}$. The internal force is the stress or tension in the light rod. This force is central: it acts in the direction of the vector joining the two particles.

Let $\mathbf{R}$ be the position of the centre of mass. From equation (5.7) we know that

$$M\ddot{\mathbf{R}} = \mathbf{F}^e = m_1\mathbf{g} + m_2\mathbf{g} = M\mathbf{g}$$

so the centre of mass moves exactly as if it were a single particle of mass M in a gravitational field.

The angular momentum $\mathbf{H}_M$ about the centre of mass satisfies equation (5.21)

$$\dot{\mathbf{H}}_M = \mathbf{G}_M.$$

The gravitational torque $\mathbf{G}_M$ about the centre of mass is

$$\mathbf{y}_1 \times (m_1\mathbf{g}) + \mathbf{y}_2 \times (m_2\mathbf{g}) = (m_1\mathbf{y}_1 + m_2\mathbf{y}_2) \times \mathbf{g} = \mathbf{0} \tag{5.25}$$

by definition of the centre of mass (5.6). Thus the angular momentum about the centre of mass of the baton is constant.

Since the rod is rigid, the two masses are rotating about the centre of mass with the same angular velocity $\boldsymbol{\omega}$. The velocity of the mass m_i is therefore $\boldsymbol{\omega} \times \mathbf{y}_i$ and

$$\mathbf{H}_M = m_1\mathbf{y}_1 \times (\boldsymbol{\omega} \times \mathbf{y}_1) + m_2\mathbf{y}_2 \times (\boldsymbol{\omega} \times \mathbf{y}_2). \tag{5.26}$$

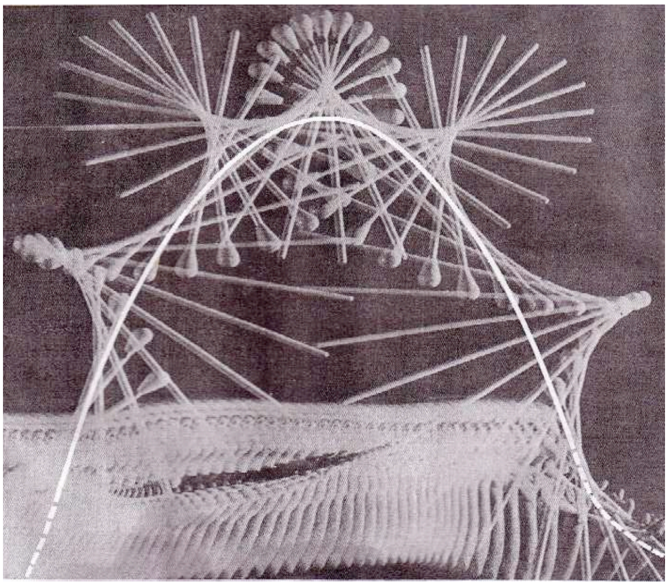
The axis of rotation is perpendicular to the rod; since the rod is thin and the masses are particles they cannot rotate about an axis parallel to the rod. Expanding the vector products in equation (5.26) and using $\boldsymbol{\omega} \cdot \mathbf{y}_i = 0$ shows that

$$\mathbf{H}_M = (m_1 y_1^2 + m_2 y_2^2)\boldsymbol{\omega}.$$

The centre of mass is fixed in the rod, so y_1^2 and y_2^2 are constant. The external torque about the centre of mass being zero therefore implies that $\boldsymbol{\omega}$ is constant, so $\dot{\theta}$ is constant in the motion, where θ is the angle the baton makes with the vertical, and $|\dot{\theta}| = |\boldsymbol{\omega}|$.

The time lapse photograph below shows this nicely: the centre of mass moves on a parabola and the angle of the rod changes by the same amount between each exposure.

[5] Note to self: could do with a picture.



5.2 The two-body problem

One thing we can be absolutely sure about is that the Sun is not, as was assumed in chapter 3, fixed in space, so the orbital calculations that turned out so well, giving conic sections, are at best approximations: accurate approximations in the case of a small planet such as Mercury, but perhaps not very accurate for giants such as Saturn and Jupiter. Clearly, we must investigate the two-body problem urgently.

Remarkably, the two-body problem turns out to be no more complicated, and indeed equivalent to, the single-body problem. This is in complete contrast to the three-body problem which very intractable,[6] having in some circumstances chaotic solutions.

5.2.1 Equations of motion

For the general two-body problem, when the particles are not a fixed distance apart, let $\mathbf{r_1}$ and $\mathbf{r_2}$ be the positions of particles of masses m_1 and m_2, respectively. The centre of mass is at $\mathbf{R}$, where

$$M\mathbf{R} = m_1\mathbf{r_1} + m_2\mathbf{r_2}$$

and $M = m_1 + m_2$. We can use $\mathbf{R}$ as one of our variables. It works nicely, but only in this two-particle case, to choose the relative position $\mathbf{r}$, defined by

$$\mathbf{r} = \mathbf{r_1} - \mathbf{r_2}$$

as a second variable, these two variables replacing $\mathbf{r_1}$ and $\mathbf{r_2}$.[7]

We can of course express $\mathbf{r}_1$ and $\mathbf{r}_2$ in terms of $\mathbf{R}$ and $\mathbf{r}$:

$$\mathbf{r}_1 = \mathbf{R} + \frac{m_2}{M}\mathbf{r}, \qquad \mathbf{r}_2 = \mathbf{R} - \frac{m_1}{M}\mathbf{r}. \tag{5.27}$$

We assume that there are no external forces (as is the case for the Sun-Earth system), so that

$$\ddot{\mathbf{R}} = 0.$$

[6]Though not insoluble: in the late 1800s King Oscar II of Sweden established a prize for anyone who could find the solution to the problem. The problem was solved by Karl Sundman, who showed that the solution could be expressed as a power series in $t^{\frac{1}{3}}$; and the prize was awarded to Poincare (to be fair, the prize was intended for the n-body problem and Sundman solved only the case $n = 3$. Poincare's treatise laid the foundations for chaos theory.

[7]We could instead use the centre of mass position vectors $\mathbf{y}_1$ and $\mathbf{y}_2$. In terms of $\mathbf{r}$, these are given by $M\mathbf{y}_1 = m_2\mathbf{r}$ and $M\mathbf{y}_2 = -m_1\mathbf{r}$. The reason for not doing so is that, in the two-body problem, we only need one more variable besides $\mathbf{R}$ (note that $\mathbf{y}_1$ and $\mathbf{y}_2$ are not independent: they satisfy $m_1\mathbf{y}_1 + m_2\mathbf{y}_2 = \mathbf{0}$) and as we shall see it works well to use $\mathbf{r}$.

The motion of the relative position vector is governed by

$$\begin{aligned}
\ddot{\mathbf{r}} &\equiv \ddot{\mathbf{r}}_1 - \ddot{\mathbf{r}}_2 && \text{(by definition)}\\
&= \frac{1}{m_1}\mathbf{F}_{12} - \frac{1}{m_2}\mathbf{F}_{21} && \text{(applying Newton's second law)}\\
&= \frac{m_1+m_2}{m_1 m_2}\mathbf{F}_{12} && (5.28)
\end{aligned}$$

which we can write as

$$\mu\ddot{\mathbf{r}} = \mathbf{F}_{12} \tag{5.29}$$

where μ, the *reduced mass*, is defined by

$$\mu = \frac{m_1 m_2}{m_1 + m_2}. \tag{5.30}$$

In the gravitational case, the equation of motion (5.29) becomes

$$\mu\ddot{\mathbf{r}} = -\frac{G m_1 m_2}{r^3}\mathbf{r} \quad \text{i.e.} \quad \ddot{\mathbf{r}} = -\frac{GM}{r^3}\mathbf{r} \tag{5.31}$$

which is exactly the same as the equation of motion for a single particle of mass M. As we showed in section 3.5, the motion is planar and can be written in the form

$$r = \frac{\ell}{e\cos\theta \pm 1}.$$

From (5.27), we see that

$$\mathbf{r_1} - \mathbf{R} = \frac{m_2}{M}\mathbf{r}$$

so the motion of each particle relative to the centre of mass is directly determined by the motion of $\mathbf{r}$. That means the motion of the individual particles consists of a constant drift due to the motion of the centre of mass, plus elliptical, hyperbolic or parabolic motion with the centre of mass as the focus.

If $\mathbf{F}_{12}$ is a central force, i.e. it is directed along $\mathbf{r}$ and depends in magnitude only on $|\mathbf{r}|$, we can write it in the form[8]

$$\mathbf{F}(\mathbf{r}) = -\boldsymbol{\nabla}\phi(r)$$

for some potential function $\phi(r)$ (see section 2.1). In this case, we can define the total energy E of system by

$$E = T + \phi(r)$$

where the total kinetic energy T is given by

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

where the reduced mass μ is given by definition (5.30).

We can now easily show that the total energy E is a constant of the motion:

$$\begin{aligned}
\frac{dE}{dt} &= M\dot{\mathbf{R}}\cdot\ddot{\mathbf{R}} + \mu\dot{\mathbf{r}}\cdot\ddot{\mathbf{r}} + \frac{d\phi}{dt}\\
&= \dot{\mathbf{r}}\cdot\mathbf{F} + \boldsymbol{\nabla}\phi\cdot\dot{\mathbf{r}} && (\ddot{\mathbf{R}} = \mathbf{0} \text{ (no external forces) and (5.29))}\\
&= 0 && \text{(definition of } \phi)
\end{aligned}$$

[8]For all the previous work, we could just as well have used variables $\mathbf{R}$ and (say) $\mathbf{y}_1$ (the position vector of one of the particles with respect to the centre of mass). For the potential, it is much more convenient to use $\mathbf{R}$ and $\mathbf{r}$ as above.

5.2.2 Tides

This is another extended footnote: interesting, I hope, but not part of the course.

We consider the Earth-Moon system as consisting of a pair of gravitationally interacting point masses.

Recall that for a two-body system, we write

$$\mathbf{r} = \mathbf{r}_1 - \mathbf{r}_2$$

so that

$$\ddot{\mathbf{r}} = \ddot{\mathbf{r}}_1 - \ddot{\mathbf{r}}_2 = \frac{1}{m_1}\mathbf{F}_{12} - \frac{1}{m_2}\mathbf{F}_{21} = \frac{m_2 + m_1}{m_1 m_2}\mathbf{F}_{12}$$

where $\mathbf{F}_{12}$ is the force on particle 1 due to particle 2. In the gravitational case, this becomes

$$\ddot{\mathbf{r}} = \left(\frac{m_2 + m_1}{m_1 m_2}\right)\left(\frac{-Gm_1 m_2}{r^3}\mathbf{r}\right) = -\frac{GM}{r^3}\mathbf{r}.$$

where $M = m_1 + m_2$. This is exactly the same equation as for a particle moving in the gravitational field of a fixed mass M, so we can wheel out all our standard orbit calculations.

As it happens, the Earth-Moon distance is roughly constant, so we can treat the orbit as a circle:

$$r\dot{\theta}^2 = \frac{GM}{r^2}$$

where $M = M_e + M_m$ is the combined mass of the Earth and Moon. The period of the orbit, $2\pi/\dot{\theta}$ for a circular orbit, is therefore (cf Kepler's third law)

$$2\pi\sqrt{\frac{r^3}{GM}}. \qquad (*)$$

We can evaluate GM using $GM/R_e^2 = g$, where R_e is the radius of the Earth and g is the acceleration due to gravity. Taking $R_e \approx 6.4 \times 10^6$ metres, $g \approx 10\text{ms}^{-2}$, the Earth-Moon distance $r \approx 4 \times 10^8$m and $2\pi \approx 6$ gives a period of 2.4×10^6 seconds, which is surprisingly close (given our rather cavalier approximations) to 28 days.

Now

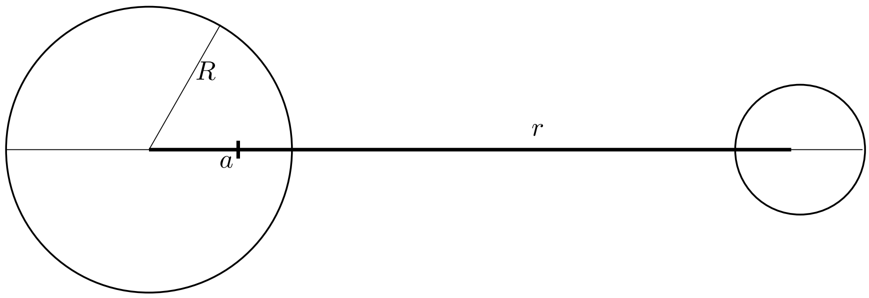
$$\mathbf{r}_1 = \mathbf{R} + \frac{m_1}{m_1 + m_2}\mathbf{r}$$

where $\mathbf{R}$ is the position vector of the centre of mass, which we assume to be fixed in space (though in fact it rotates round the Sun), and there is a corresponding result for $\mathbf{r}_2$. This means that both the Earth and the Moon orbit their common centre of mass. Since the mass of the moon is about 1% of the mass of the Earth, the centre of the Earth is about 4000 kilometres from the common centre of gravity which means that the joint centre of mass lies inside the Earth.

From this picture, we can understand the reason that there are two tides a day. At the centre of the Earth, in the rotating frame of the Earth-Moon, the centrifugal force and the gravitational attraction of the moon balance. At the point on the surface of the Earth nearest the Moon, the Moon's pull exceeds the centrifugal force[9]. At the point on the Earth's surface furthest from the Moon, the centrifugal force exceeds the Moon's gravitational pull. Thus a particle of seawater experiences an excess force towards the Moon on one side of the Earth and away from the Moon on the other side of the Earth. The sea will therefore bulge on both sides.

We can do the calculations: we will prove that the force acting on a particle of seawater at the point on the Earth's surface nearest the moon is the same as at the point on the Earth's surface furthest from the moon, which shows not only that there are high tides at the same time on opposite sides of the Earth, but also that the heights of the high tides are the same.

[9]It is best in this explanation to think of the Earth and Moon rotating round a common centre of mass somewhere in space between the two bodies. In fact, as noted above, the centre of mass lies within the Earth. Thus at the point on the Earth's surface nearest to the Moon, the centrifugal force and the Moon's pull are in the same direction.



We consider a particle of unit mass that is a distance x from the centre of the Earth (where $x > R$) In the diagram, r is the distance between the centres of the Earth and Moon and a is the distance between the centre of mass of the two-body system (about which the system is rotating) and the centre of the Earth. In the rotating frame, there are three forces on the particle: the gravitational force due to the moon; the gravitational force due to the Earth and the centrifugal force. The total force towards the centre of the Earth is therefore

$$F_1(x) = \frac{GM_e}{x^2} - \frac{GM_m}{(r-x)^2} - \omega^2(x-a). \tag{$\dagger$}$$

We have

$$a = \frac{rM_m}{M},$$

by definition of the centre of mass of the two bodies. Furthermore, from the equation of motion for a circular orbit (cf ($*$)), we have

$$\omega^2 r = \frac{GM}{r^2} \tag{$**$}$$

The corresponding result to ($\dagger$) for a particle a distance x from the centre of the Earth on the side away from the Moon is

$$F_2(x) = \frac{GM_e}{x^2} + \frac{GM_m}{(r+x)^2} - \omega^2(x+a). \tag{$\ddagger$}$$

To prove our result, we have to show that $F_1(R) = F_2(R)$, i.e. that (omitting some terms that agree)

$$-\frac{GM_m}{(r-R)^2} - \omega^2(R-a) = \frac{GM_m}{(r+R)^2} - \omega^2(R+a)$$

or (substituting for ω^2 from ($**$) and cancelling G)

$$-\frac{M_m}{(r-R)^2} + \frac{Ma}{r^3} = \frac{M_m}{(r+R)^2} - \frac{Ma}{r^3}$$

or (substituting for a and cancelling M_m)

$$-\frac{1}{(r-R)^2} + \frac{1}{r^2} = \frac{1}{(r+R)^2} - \frac{1}{r^2}.$$

Now we expand both sides in a Taylor series in R, taking just the first few terms since $R/r \ll 1$:

$$-\frac{1}{r^2} - 2\frac{R}{r^3} - 3\frac{R^2}{r^4} + \cdots + \frac{1}{r^2} = \frac{1}{r^2} - 2\frac{R}{r^3} + 3\frac{R^2}{r^4} + \cdots - \frac{1}{r^2}$$

This shows that the two forces are indeed equal provided we ignore quadratic and higher terms in the small quantity R/r.

Thus there are two tides a day except when local conditions are exceptional, such as in the Solent where there are four tides and in Karumba, Australia where there is only one.

5.3 Variable mass problems

It was shown in section 5.1 that Newton's second law can be applied not just to a single particle, but to a system of particles. One application is to a system consisting of a rocket and its fuel. We assume that in the process of burning the fuel, the assumptions behind (5.2) apply. Another application is an avalanche, or a rolling snowball, where a mass of snow is picked up during the motion.

5.4 Rockets

5.4.1 The rocket equation

We consider a rocket which has mass $m(t)$ at time t, where $m(t)$ includes both the fixed mass of the rocket without fuel and the mass of the fuel on board at time t. For simplicity, we assume that the rocket is moving along the positive x-axis. Its velocity $v(t)$ and we take $v(t) > 0$ (this is helpful but not necessary). It emits exhaust backwards at a velocity $-u$ (where $u > 0$) relative to the rocket. The rocket is subject to an external force which could, for example, be gravity or friction.

We use the equation

$$\frac{d\mathbf{P}}{dt} = \mathbf{F}^e$$

where $\mathbf{P}$ is the total momentum of the system and $\mathbf{F}^e$ is the total external force (see section 5.1.1). This holds in any inertial frame. It only assumes that the internal forces (in this case, the explosive burning of the rocket fuel) obey Newton's third law.

We therefore calculate the rate of change of momentum between times t and $t+\delta t$; we do not need to worry about all the exhaust gasses that were emitted before this time.

The momentum of the system at time t is

$$m(t)v(t)$$

and the momentum at time $t+\delta t$, when the mass of the rocket has decreased to $m(t+\delta t)$ and a mass $m(t) - m(t+\delta t)$ of fuel has been converted to exhaust gases and ejected at speed $v - u$, is

$$m(t+\delta t)v(t+\delta t) + [m(t) - m(t+\delta t)](v-u).$$

The rate of change of momentum is therefore given by

$$\begin{aligned}
\frac{dP}{dt} &= \lim_{\delta t\to 0} \frac{m(t+\delta t)v(t+\delta t) + [m(t) - m(t+\delta t)](v-u) - m(t)v(t)}{\delta t} \\
&= \lim_{\delta t\to 0} \frac{m(t+\delta t)v(t+\delta t) - m(t)v(t)}{\delta t} + \lim_{\delta t\to 0} \frac{[m(t) - m(t+\delta t)](v-u)}{\delta t} \\
&= \frac{d(mv)}{dt} - \frac{dm}{dt}(v-u) \\
&= m\frac{dv}{dt} + u\frac{dm}{dt}
\end{aligned}$$

so

$$m\frac{dv}{dt} + u\frac{dm}{dt} = F \tag{5.32}$$

which is the *rocket equation*. Here, F is the external force on the rocket; the external force on the ejected fuel will tend to zero when we take the limit $\delta t \to 0$ since then $\delta m \to 0$.

If $F = 0$, we can write the rocket equation as

$$\frac{dv}{dt} + u\frac{1}{m}\frac{dm}{dt} = 0$$

which can be integrated in the case when u does not vary with time:

$$v(t) - v(0) = u \ln\left(\frac{m(0)}{m(t)}\right). \tag{5.33}$$

This is called the Tsiolkovski equation.[10]

In the case

$$\frac{dm}{dt} = -\alpha$$

[10] The equation was apparently (though this seems very surprising) first written down by Konstantin Tsiolkovski as recently as 1903. Tsiolkovski was also responsible for the idea of the space elevator, which consists of a cable (carbon nanotubes) attached to a geostationary satellite. The cable extends higher than the satellite and has a mass on the end; this provides tension, since on this portion of the cable the centrifugal force outwards exceeds the gravitational force inwards. On the lower portion of the cable, the reverse is true, and the elevator has to have an engine to power it up the cable. As the elevator goes up the cable, its tangential velocity increases; total angular momentum is conserved, so the climber's increased angular momentum is compensated by a decrease of angular velocity of the Earth. There will a horizontal coriolis force, dragging the cable, due to the vertical velocity of the elevator in the rotating frame. I bet it never happens.

where α is a positive constant, we can write

$$m(t) = m(0) - \alpha t$$

and substituting this into the Tsiolkovski equation (5.33) gives an equation that can be easily integrated to give $x(t)$.

The rocket equation (5.32) can easily be understood (in fact, can be much more easily derived) working in the instantaneous rest frame of the rocket at fixed time t. The *instantaneous* rest frame moves at a constant velocity which is the velocity of the rocket at one instant only. It is an inertial frame. In contrast, the centre of mass frame of the rocket is thei generally non-inertial frame that moves with the centre of mass of the rocket; in this frame the momentum of the rocket is always zero.

We work in the frame that has the same velocity as the rocket at time t. In this frame, the momentum at time t is zero:

$$P(t) = 0.$$

At time $t+\delta t$, when the speed of the rocket has increased by δv, the mass of the rocket has increased by δm (actually, a decrease of $-\delta m$ where $\delta m < 0$) and a mass $-\delta m$ of fuel has been ejected, so

$$P(t + \delta t) = (m + \delta m)\delta v + (-\delta m)(-u).$$

Ignoring second order quantities, this leads immediately to

$$F = \lim_{\delta t \to 0} \left(m\frac{\delta v}{\delta t} + u\frac{\delta m}{\delta t} \right).$$

Of course, this is the same calculation as the previous calculation, except that we have replaced any velocity V with $V - v$.

The rocket equation can be obtained even more directly as follows. Since it doesn't matter what happens to the exhaust gases after they are ejected, we assume that they move unimpeded, i.e. they experience no force and move with constant velocity. The total momentum of the exhaust gases at time t is then

$$\int (v - u)dm$$

with appropriate limits. Thus

$$P(t) = mv + \int (v - u)dm$$

or

$$P(t) = mv + \int_0^t (v - u)\frac{dm}{dt}dt.$$

Differentiating this with respect to t and using Newton's second law in the form (5.4) gives the rocket equation without further work.

5.4.2 Example: rocket with linear drag

A rocket burns fuel at a constant mass rate α and expels it at constant relative speed u. It experiences linear air resistance. The initial mass of the rocket is m_0, of which a fraction $1 - \beta$ $(0 < \beta < 1)$ is fuel, and it is initially at rest in deep space. What is the speed of the rocket when all the fuel has been burnt?

Let v be the velocity and m the mass at time t. We do not need to find the time, so we will eliminate t from the rocket equation using the chain rule:

$$\frac{d}{dt} = \frac{dm}{dt}\frac{d}{dm} = -\alpha\frac{d}{dm}.$$

The rocket equation (5.32) with resistive force kv is

$$m\dot{v} + u\dot{m} = -kv$$

where k is the coefficient of air resistance; so

$$
\begin{aligned}
-\alpha m\frac{dv}{dm} - \alpha u &= -kv \\
\implies \qquad \frac{dv}{dm} &= \frac{kv-\alpha u}{\alpha m} \qquad (5.34)\\
\implies \qquad \int \frac{dv}{kv-\alpha u} &= \int \frac{dm}{\alpha m} \\
\implies k^{-1}\ln[(\alpha u - kv)/\alpha u] &= \alpha^{-1}\ln(m/m_0)\,,
\end{aligned}
$$

where we have used the initial condition $v = 0$ when $m = m_0$ in the last equation. Note that we expect the velocity to increase when the mass decreases, so equation (5.34) shows that $kv - \alpha u > 0$.

Tidying up a bit:

$$
\begin{aligned}
\frac{1}{k}\ln\left(1-\frac{kv}{\alpha u}\right) &= \frac{1}{\alpha}\ln\frac{m}{m_0} \\
\implies \qquad 1-\frac{kv}{\alpha u} &= \left(\frac{m}{m_0}\right)^{k/\alpha} \\
\implies \qquad v &= \frac{\alpha u}{k}\left\{1-\left(\frac{m}{m_0}\right)^{k/\alpha}\right\}.
\end{aligned}
$$

When all the fuel is burnt, $m = \beta m_0$ so $v_{\text{final}} = (\alpha u/k)(1-\beta^{k/\alpha})$.

We see that there is a theoretical maximum speed of $\alpha u/k$, achievable only if the rocket is entirely made up of fuel.

5.4.3 Example: avalanches

The model is a compact mass of snow sliding down a slope, picking up the snow immediately in front of it as it goes. It is like rolling a snow ball to make a snowman, except that (i) avalanches slide rather than roll and (ii) avalanches take up the width of the slope (more cylindrical then spherical).[11]

Let m be the mass of snow in the avalanche at time t, and let v be the speed at which it is sliding down the slope. We can use the rocket equation (5.32) directly, with $u = v$ since the snow on the slope is at rest:

$$m\frac{dv}{dt} + v\frac{dm}{dt} = mg\sin\alpha,$$

where α is the inclination of the slope. Note that this is exactly the equation we would have obtained using Newton's second law directly:

$$\frac{d(mv)}{dt} = \text{force}.$$

If we assume that the depth of snow is a constant h on the slope and that the avalanche picks it all up, then

$$m = \rho h x,$$

where ρ is the density of the snow and x is the distance moved by the avalanche at time t. Thus

$$\frac{d(\rho h x v)}{dt} = \rho h x g \sin\alpha.$$

Cancelling the constant factors and using the chain rule gives

$$v\frac{d(xv)}{dx} = xg\sin\alpha$$

which can be easily solved. For example, we could write it as

$$(xv)\frac{d(xv)}{dx} = x^2 g\sin\alpha$$

[11]Note to self: can the two-snowplough problem be adapted? In that problem, the speed of the ploughs was inversely proportional to the depth of snow, but if the ploughs worked at constant power, the result might be similar.

giving

$$\tfrac{1}{2}(xv)^2 = \tfrac{1}{3}x^3 g \sin\alpha$$

i.e.

$$v^2 = \tfrac{2}{3}xg\sin\alpha \tag{5.35}$$

and hence, if we want it, x as a function of time. We can obtain a more interesting result by differentiating equation (5.35) respect to time t:

$$2v\frac{dv}{dt} = \tfrac{2}{3}xvg\sin\alpha$$

which shows (after cancelling v) that the acceleration of an avalanche in this very crude model, is $\frac{1}{3}$ of the acceleration of a skier heading down the same slope (ignoring resistance). That suggests that skiers should be able to out-run avalanches, especially if they have a head start.[12]

5.5 Moment of inertia

5.5.1 Definition

For a rotating system, with all particles rotating about the same axis with the same angular velocity (such as a rigid body), it is convenient to use angular acceleration rather than linear acceleration in the equations of motion. These equations, using angular acceleration, can be made to resemble the equations using linear acceleration by replacing mass by *moment of inertia*, defined as follows.

We start by considering a single particle rotating with angular velocity $\boldsymbol{\omega}$ about an axis through the origin. The kinetic energy T is given by

$$\begin{aligned} T = \tfrac{1}{2}m\dot{\mathbf{r}}\cdot\dot{\mathbf{r}} &= \tfrac{1}{2}m(\boldsymbol{\omega}\times\mathbf{r})\cdot(\boldsymbol{\omega}\times\mathbf{r}) \\ &= \tfrac{1}{2}m\omega^2(\mathbf{k}\times\mathbf{r})\cdot(\mathbf{k}\times\mathbf{r}) \qquad \text{(setting } \boldsymbol{\omega}=\omega\mathbf{k}\text{, where } \mathbf{k}\cdot\mathbf{k}=1) \\ &= \tfrac{1}{2}ma^2\omega^2 \end{aligned}$$

where, in spherical polar coordinates,

$$a = |\mathbf{k}\times\mathbf{r}| = r\sin\theta$$

so a is the distance from the particle from the axis of rotation. By analogy with

$$T = \tfrac{1}{2}mv^2$$

we write

$$T = \tfrac{1}{2}I\omega^2 \tag{5.36}$$

where I, defined by,

$$I = ma^2 \equiv m(\mathbf{k}\times\mathbf{r})\cdot(\mathbf{k}\times\mathbf{r}) \tag{5.37}$$

is the *moment of inertia* of single particle of mass m about an axis through the origin which is a distance a from the particle.

For a rigid body consisting of n particles (all necessarily rotating about the same axis at the same angular speed), we have for the total kinetic energy

$$T = \sum_{i=1}^{n} \tfrac{1}{2}m_i(a_i\omega)^2 = \tfrac{1}{2}\omega^2\sum_{i=1}^{n} m_i a_i^2,$$

where m_i is the mass of the ith particle and a_i is the distance from the ith particle from the axis. Accordingly, we define the moment of inertia of the body by

$$I = \sum_{i=1}^{n} m_i a_i^2 \equiv \sum m_i(\mathbf{k}\times\mathbf{r}_i)\cdot(\mathbf{k}\times\mathbf{r}_i) \tag{5.38}$$

so that, again,

$$T = \tfrac{1}{2}I\omega^2.$$

[12]Don't rely on it: I doubt if a real avalanche bears any resemblance to this model.

For a solid body[13], we replace the summation by an integral, and the individual masses by $\rho(\mathbf{r})dV$:

$$I = \int_{\text{body}} (r\sin\theta)^2 \rho dV$$

where r and θ are spherical polar coordinates and the polar direction of the coordinates is chosen to be along the axis of rotation, so that the distance of the element of mass at $\mathbf{r}$ from the axis is $r\sin\theta$.

5.5.2 Examples

(i) Uniform rod, axis perpendicular to rod.

Let ℓ be the length of the rod and ρ be the line density (mass per unit length). Then the moment of inertial about an axis perpendicular to the rod passing through one end is

$$I = \int_0^\ell \rho x^2 dx = \tfrac{1}{3}\rho\ell^3 = \tfrac{1}{3}M\ell^2,$$

where M is the mass of the rod.

(ii) Circular hoop, axis through centre of hoop perpendicular to the plane of the hoop

We consider a system of particles of total mass M all situated on a circular hoop. The moment of inertia of this system about an axis through the centre of the hoop and perpendicular to the plane of the hoop is just (since the moment of the inertia of the system is the sum of the moments of inertia of the individual particles)

$$I = Ma^2,$$

where a is the radius of the hoop.

(iii) Uniform circular disc, axis through centre perpendicular to disc

We consider a uniform disc of mass M with radius a and surface density (mass per unit area) ρ. In this case we have to do an integral. We can be high-tech and do the surface integral (using plane polar coordinates)

$$\int_0^a \int_0^{2\pi} \rho r^2 r d\theta dr = \tfrac{1}{4}\rho(2\pi a^4) = \tfrac{1}{2}Ma^2;$$

this method would work also for a non-uniform disc, for which ρ depended on r and θ, though the answer would be different. Or we can be low-tech, and break down the disc into hoops add up the individual moments of inertia of the hoops.

(iv) Uniform disc, axis through point on circumference perpendicular to disc

We consider the same disc as above, with the axis of rotation passing through the point $r = a$, $\theta = 0$; call this point $\mathbf{a}$. This time, the distance of the point $\mathbf{r}$ with coordinates (r, θ) from the axis $\mathbf{a}$ is given by

$$\begin{aligned}\text{distance from axis} &= \sqrt{(\mathbf{r}-\mathbf{a})\cdot(\mathbf{r}-\mathbf{a})}\\ &= \sqrt{r^2 + a^2 - 2\mathbf{r}\cdot\mathbf{a}}\\ &= \sqrt{r^2 + a^2 - 2ar\cos\theta}.\end{aligned}$$

so

$$I = \int_0^a \int_0^{2\pi} \rho(r^2 + a^2 - 2ar\cos\theta) r d\theta dr = \tfrac{3}{2}Ma^2.$$

(v) Uniform disc, axis through centre in plane of disc

We choose plane polar coordinates so that $\theta = 0$ corresponds to the direction of the axis of rotation. The point with coordinates (r, θ) is a distance $|r\sin\theta|$ from the axis, so

$$I = \int_0^a \int_0^{2\pi} \rho(r\sin\theta)^2 r d\theta dr = \tfrac{1}{4}Ma^2.$$

(vi) Uniform sphere, axis through centre

[13] By a solid body, we mean a body made of a continuous medium rather than individual particles

We choose spherical polar coordinates with the axis ($\theta = 0$) pointing in the direction of the axis of rotation. Then

$$I = \int_0^a \int_0^\pi \int_0^{2\pi} \rho (r\sin\theta)^2 r^2 \sin\theta d\phi d\theta dr = \tfrac{8}{15}\pi\rho a^5 = \tfrac{2}{5}Ma^2.$$

5.5.3 The parallel axis theorem

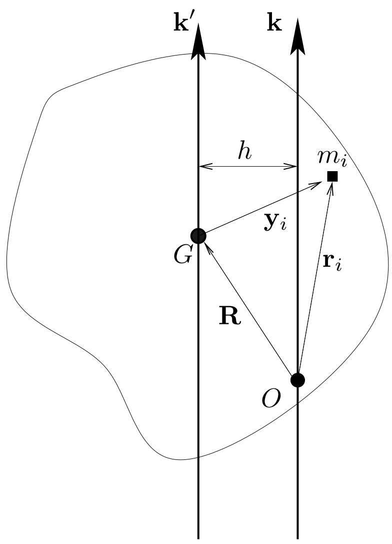
Let I be the moment of inertia of a system of particles about an given axis $\mathbf{k}$ which passes through the origin O and let I' be the moment of inertia about an axis $\mathbf{k}'$ which passes through the centre of mass G of the system. Using the notation of section 5.1.4 we have

$$I = \sum_{i=1}^n m_i(\mathbf{k}\times\mathbf{r}_i)\cdot(\mathbf{k}\times\mathbf{r}_i) \quad \text{and} \quad I' = \sum_{i=1}^n m_i(\mathbf{k}\times\mathbf{y}_i)\cdot(\mathbf{k}\times\mathbf{y}_i)$$

where $\mathbf{r}_i = \mathbf{y}_i + \mathbf{R}$. Then

$$\begin{aligned}
I &= \sum_{i=1}^n m_i(\mathbf{k}\times\mathbf{r}_i)\cdot(\mathbf{k}\times\mathbf{r}_i) \\
&= \sum_{i=1}^n m_i(\mathbf{k}\times(\mathbf{y}_i+\mathbf{R}))\cdot(\mathbf{k}\times(\mathbf{y}_i+\mathbf{R})) \\
&= \sum_{i=1}^n m_i(\mathbf{k}\times\mathbf{y}_i)\cdot(\mathbf{k}\times\mathbf{y}_i) + \sum_{i=1}^n 2m_i(\mathbf{k}\times\mathbf{R})\cdot(\mathbf{k}\times\mathbf{y}_i) + \sum_{i=1}^n m_i(\mathbf{k}\times\mathbf{R})\cdot(\mathbf{k}\times\mathbf{R}) \\
&= I' + 2(\mathbf{k}\times\mathbf{R})\cdot\Big(\mathbf{k}\times\sum_{i=1}^n 2m_i\mathbf{y}_i\Big) + M(\mathbf{k}\times\mathbf{R})\cdot(\mathbf{k}\times\mathbf{R}). \\
&= I' + Mh^2, && \text{(using (5.18)}
\end{aligned}$$

where h is the distance between the two axes. Note that for any vector $\mathbf{X}$, $\mathbf{k}'\times\mathbf{X} = \mathbf{k}\times\mathbf{X}$ since $\mathbf{k}$ and $\mathbf{k}'$ are parallel.



Thus

$$I = I' + Mh^2 \tag{5.39}$$

which is the *parallel axis theorem.* This is an important result, and worth restating. It says that if I' is the moment of inertia of a body of mass M about an axis $\mathbf{k}'$ through the centre of mass, and I is the moment of inertia of the body about an axis $\mathbf{k}$, where $\mathbf{k}$ and $\mathbf{k}'$ are parallel and a distance h apart, then equation (5.39) holds. Note that this does not apply to any two axes: $\mathbf{k}'$ must pass through the centre of mass.

You might like to check that the result derived in example (iv) of the previous section follows immediately from example (iii) by the parallel axis theorem (the axis in example (iv) is the parallel to the axis in example (iii) but displaced by a distance a.)

5.5.4 Angular momentum

For a single particle rotating with angular speed ω about an axis $\mathbf{k}$ passing through the origin, the angular momentum $\mathbf{H}$ about the origin is given by

$$\begin{aligned}\mathbf{H} &= \mathbf{r}\times(m\dot{\mathbf{r}})\\ &= m\mathbf{r}\times(\boldsymbol{\omega}\times\mathbf{r}) \qquad (5.40)\\ &= m\omega\mathbf{r}\times(\mathbf{k}\times\mathbf{r}). \qquad (5.41)\end{aligned}$$

The component of $\mathbf{H}$ in the direction of the axis of rotation is given by

$$\begin{aligned}\mathbf{H}\cdot\mathbf{k} &= m\omega\mathbf{r}\times(\mathbf{k}\times\mathbf{r})\cdot\mathbf{k}\\ &= m\omega(\mathbf{r}\times\mathbf{k})\cdot(\mathbf{r}\times\mathbf{k}) \qquad \text{(using the cyclic property of scalar triple products)}\\ &= I\omega \equiv ma^2\omega \qquad (5.42)\end{aligned}$$

where $a = r\sin\theta$, which is the distance of the particle from the axis, in agreement with the formulae used in section 3 for the angular momentum of a particle moving in a plane ('$mh = mr^2\dot{\theta}$').

If the axis of rotation $\mathbf{k}$ is fixed, we can write

$$\boldsymbol{\omega} = \omega\mathbf{k} \qquad \text{and} \qquad \dot{\boldsymbol{\omega}} = \dot{\omega}\mathbf{k}$$

in which case we can differentiate equation (5.42) to obtain

$$\mathbf{G}\cdot\mathbf{k} = I\dot{\omega}$$

where I is the moment of inertia of about the axis $\mathbf{k}$.

This equation, derived for a single particle, applies also to a rigid system of particles rotating about a fixed axis since we can sum the moments of inertia of the individual particles on the left hand side and sum the torques on the individual particles on the right hand side (provided the internal forces are central — see section 5.1.2).

5.6 The inertia tensor

In the above calculations of moments of inertia of a disc, we considered two cases when the axis is through the centre of the disc: axis perpendicular to the plane of the disc and axis in the plane of the disc. It would be useful to be able to express the results of these two calculations as a single entity. And it would be even more useful if this entity could be used, without further calculation, to provide also the moment of inertia about any other axis. By good fortune, such an entity exists: the *inertia tensor*.[14]

We could define the inertia tensor directly, but the same indirect approach as we used before, via the kinetic energy, is helpful. We consider a particle of mass m rotating with angular speed ω about a fixed vector $\mathbf{k}$ that passes through the origin. The angular velocity $\boldsymbol{\omega}$ is given by

$$\boldsymbol{\omega} = \omega\mathbf{k}.$$

The velocity of the particle is $\boldsymbol{\omega}\times\mathbf{r}$ and the kinetic energy is given by

$$\begin{aligned}T &= \tfrac{1}{2}m(\boldsymbol{\omega}\times\mathbf{r})\cdot(\boldsymbol{\omega}\times\mathbf{r})\\ &= \tfrac{1}{2}m\omega^2(\mathbf{k}\times\mathbf{r})\cdot(\mathbf{k}\times\mathbf{r})\\ &= \tfrac{1}{2}m\omega^2\big((\mathbf{k}\cdot\mathbf{k})(\mathbf{r}\cdot\mathbf{r}) - (\mathbf{k}\cdot\mathbf{r})^2\big)\\ &= \tfrac{1}{2}m\omega^2\left((\mathbf{r}\cdot\mathbf{r})\delta_{ij} - x_ix_j\right)k_ik_j \qquad (5.43)\end{aligned}$$

This last expression (5.43) has three parts to it: the angular speed, the axis of rotation and

$$\tfrac{1}{2}m\left((\mathbf{r}\cdot\mathbf{r})\delta_{ij} - x_ix_j\right) \equiv I_{ij}. \qquad (5.44)$$

These three parts are independent: we can choose, at will, an angular speed, and axis of rotation and the matrix represented by the components (5.44), plug them into (5.43) and out will come the kinetic energy.

[14]If you don't know yet what a tensor is, don't worry: you will soon and anyway for present purposes you can just replace the word 'tensor' with the word 'matrix' and no harm will be done.

The part of the kinetic energy that refers only to the system of particles (and not to the angular velocity) is the matrix (5.44), which is called the *inertia tensor* of the system about the origin $\mathbf{r} = \mathbf{0}$. It tells us how the body responds to rotation about an arbitrary axis. That I_{ij} are the components of a (symmetric) tensor can be seen by applying the quotient rule to the definition

$$T = \tfrac{1}{2} m\omega^2 I_{ij} k_i k_j$$

since $T/m\omega^2$ is a scalar for any vector $\mathbf{k}$.

If the expression (5.44) looks complicated, one has only to write out the individual components to see what it means. For example,

$$I_{11} = m\left((x^2 + y^2 + z^2)\delta_{11} - x^2\right) = m(y^2 + z^2) \tag{5.45}$$

and

$$I_{12} = -mxy.$$

Note that

$$I_{11} = I_{ik} n_i n_k$$

where $\mathbf{n}$ is the vector $(1, 0, 0)$. Also, we see from equation (5.45) that I_{11} is the moment of inertia of the particle about $\mathbf{n}$ (because $(y^2 + z^2)^{\frac{1}{2}}$ is the distance from the particle to the $\mathbf{n}$ axis). Thus the quantity

$$I_{ik} n_i n_k$$

is the moment of inertia of the particle about an axis in the direction of $\mathbf{n}$. This holds independently of the coordinate axes used[15] and so is true for any vector $\mathbf{n}$.

For a rigid system of particles, the moment of inertia is just the sum of the moments of inertia of the individual particles. For a rigid body, the sums become integrals, so for example

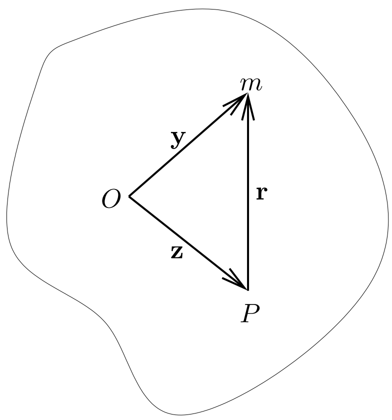
$$I_{11} = \int_{\text{body}} (y^2 + z^2)\rho dV.$$

5.6.1 The parallel axis theorem again

Let I_{ik} be the moment of inertia of a system of particles about the centre of mass and let I'_{ik} be the moment of inertia about an arbitrary point P. Let $\mathbf{z}$ be the position vector of P with respect to the centre of mass.

The moment of inertia about P of an individual particle of mass m with position vectors $\mathbf{y}$ with respect to the centre of mass and $\mathbf{r}$ with respect to P is

$$\begin{aligned} m\big(\mathbf{r}\cdot\mathbf{r}\delta_{ik} - x_i x_k\big) =& m\big((\mathbf{y}-\mathbf{z})\cdot(\mathbf{y}-\mathbf{z})\delta_{ik} - (y_i - z_i)(y_k - z_k)\big) \\ =& m\big(\mathbf{y}\cdot\mathbf{y}\delta_{ik} - y_i y_k\big) + m(\mathbf{z}\cdot\mathbf{z}\delta_{ik} - z_i z_k\big) - m\big(2\mathbf{y}\cdot\mathbf{z}\delta_{ik} + (y_i z_k + z_i y_k)\big). \end{aligned} \tag{5.46}$$



If we now sum (5.46) over all the masses in the system, the terms linear in $\mathbf{y}$ drop out, because by definition

$$\sum_{\text{masses}} m\mathbf{y} = 0 \quad \text{i.e.} \quad \sum_{\text{masses}} m y_i = 0 \quad (i = 1, 2, 3)$$

---

[15] because I_{ik} are the components of a tensor.

Thus

$$I'_{ik} = I_{ik} + M(\mathbf{z}\cdot\mathbf{z}\delta_{ik} - z_i z_k)$$

where M is the total mass of the system. This equation shows how to the inertia tensor about the centre of mass of a body to the inertia tensor about another point fixed in the body.

Note that if we choose axes such that $\mathbf{z} = (h, 0, 0)$ or $(0, h, 0)$, we find that

$$I'_{33} = I_{33} + Mh^2$$

which is the *parallel axis theorem*. This is an important result, and worth restating. It says that if I is the moment of inertia of a body of mass M about an axis $\mathbf{k}$ through the centre of mass, and I' is the moment of inertia of the body about an axis $\mathbf{k}'$, where $\mathbf{k}$ and $\mathbf{k}'$ are parallel and a distance h apart, then

$$I' = I + Mh^2.$$

Note that this does not apply to any two axes: $\mathbf{k}$ must pass through the centre of mass.

5.7 Motion of a rigid body

5.7.1 Velocity

We consider a body consisting of n particles at fixed distances from each other. The mass of the ith particle is m_i and it has position vector $\mathbf{r}_i$ with respect to a given origin fixed in space. The centre of mass of the particles is at $\mathbf{R}$ and the total mass is M where

$$M\mathbf{R} = \sum_{i=0}^{n} m_i \mathbf{r}_i.$$

Let $\mathbf{y}_i$ be the position vector of the ith particle with respect to the centre of mass, so that

$$\mathbf{r}_i = \mathbf{R} + \mathbf{y}_i \tag{5.47}$$

and

$$\dot{\mathbf{r}}_i = \dot{\mathbf{R}} + \dot{\mathbf{y}}_i \tag{5.48}$$

Since the system is rigid, the distance from the centre of mass to each particle is fixed, which means that the only motion possible relative to the centre of mass is a rotation. By Euler's theorem[16], there is at each time an angular velocity vector $\boldsymbol{\omega}$ (the same $\boldsymbol{\omega}$ for each $\mathbf{y}_i$) such that

$$\dot{\mathbf{y}}_i = \boldsymbol{\omega} \times \mathbf{y}_i,$$

so that (5.48) becomes

$$\dot{\mathbf{r}}_i = \dot{\mathbf{R}} + \boldsymbol{\omega} \times \mathbf{y}_i \tag{5.49}$$

or, using (5.47),

$$\dot{\mathbf{r}}_i = \dot{\mathbf{R}} + \boldsymbol{\omega} \times (\mathbf{r}_i - \mathbf{R}). \tag{5.50}$$

We have not used the fact that $\mathbf{y}_i$ is the position vector of the ith particle with respect specifically to the centre of mass so we could equally well have written

$$\dot{\mathbf{r}}_i = \dot{\mathbf{Q}} + \boldsymbol{\omega}' \times (\mathbf{r}_i - \mathbf{Q}) \tag{5.51}$$

where $\mathbf{Q}$ is the position vector of a point Q fixed in the body, and $\boldsymbol{\omega}'$ is the appropriate angular velocity about Q. The question is: how are $\boldsymbol{\omega}'$ and $\boldsymbol{\omega}$ related? And the answer is at first sight surprising.

If we take $\mathbf{r}_i = \mathbf{Q}$ in equation (5.50) we obtain

$$\dot{\mathbf{Q}} = \dot{\mathbf{R}} + \boldsymbol{\omega} \times (\mathbf{Q} - \mathbf{R}). \tag{5.52}$$

Subtracting equation (5.52) from equation (5.50) gives

$$\dot{\mathbf{r}}_i = \dot{\mathbf{Q}} + \boldsymbol{\omega} \times (\mathbf{r}_i - \mathbf{Q}) \tag{5.53}$$

and comparing this with equation equation (5.51), which holds for any arbitrary position vector $\mathbf{r}_i$, we infer that

$$\boldsymbol{\omega}' = \boldsymbol{\omega} \tag{5.54}$$

so, rather remarkably, the angular velocity about any point of a rigid body is the same.

A simple example shows that this not only makes sense, but is also a powerful tool for calculating the speeds of points (or particles) on rigid bodies.

[16]See section 4.2 for an off-syllabus discussion of Euler's theorem

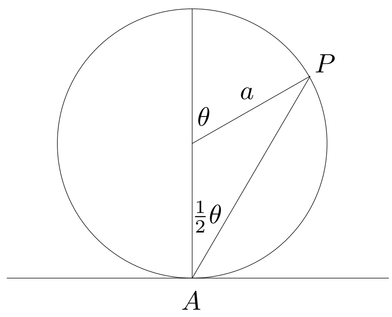
5.7.2 Example: angular velocity of rolling hoop

The figure shows a hoop of radius a rolling, without slipping, on a table. The point A is the point on the hoop which is the instantaneous point of contact with the table. This point is instantaneously at rest, because of the no-slip condition.

The angular speed of the point P with respect to the centre of the hoop is $\dot\theta$ so that the speed of P with respect to the centre is $a\dot\theta$. The speed of the centre itself is $a\dot\theta$. When $\theta = 0$ the velocity of the with respect to the centre and the velocity of the centre are parallel so the velocity of the P when $\theta = 0$ is $2a\dot\theta$.

It looks as if the angular speed about the fixed point A on the table might be $\frac{1}{2}\dot\theta$, but a moment's reflection shows that this must be wrong: when $\theta = 0$ this would give the speed of P as $2a(\frac{1}{2}\dot\theta)$ instead of $2a\dot\theta$. However, if the angular speed of P about A were $\dot\theta$ (as we know it must be from the calculation (5.54)), we obtain the correct speed.

The reason that the angular velocity of P about A is not $\frac{1}{2}\dot\theta$ is that $\frac{1}{2}\theta$ is the angle between AP and the diameter, and the diameter itself is rotating.



Now we see the power of the result (5.54). To calculate the speed of P, to find the kinetic energy of a particle at P for example, all we need is

$$\text{speed} \quad = AP \times \dot\theta = (2a\cos\tfrac{1}{2}\theta)\dot\theta.$$

To work out this result using Cartesian axes, or using (5.49) would have taken very considerably longer.[17]

5.7.3 Kinetic Energy

For a simple motion, such as a ball rolling down an inclined plane, it is often easiest to use the constancy of the total energy of the body to find the motion. We can calculate the kinetic energy T_i of the ith particle using the expression (5.23), which becomes

$$\begin{aligned} T_i &= \tfrac{1}{2}m_i\dot{\mathbf{R}}\cdot\dot{\mathbf{R}} + \tfrac{1}{2}m_i(\boldsymbol{\omega}\times\mathbf{y}_i)\cdot(\boldsymbol{\omega}\times\mathbf{y}_i) \\ &= \tfrac{1}{2}m_i\dot{\mathbf{R}}\cdot\dot{\mathbf{R}} + \tfrac{1}{2}\omega^2 \end{aligned}$$

Summing over all particles gives

$$T = \tfrac{1}{2}M\dot{\mathbf{R}}\cdot\dot{\mathbf{R}} + \tfrac{1}{2}I\omega^2, \tag{5.55}$$

where I is the moment of inertia of the whole system about the axis through it centre of mass.

5.7.4 Angular momentum

For a single particle rotating with angular speed ω about an axis $\mathbf{k}$ passing through the origin, the angular momentum $\mathbf{H}$ about the origin is given by

$$\begin{aligned} \mathbf{H} &= \mathbf{r}\times(m\dot{\mathbf{r}}) \\ &= m\mathbf{r}\times(\boldsymbol{\omega}\times\mathbf{r}) \qquad (5.56) \\ &= m\omega\mathbf{r}\times(\mathbf{k}\times\mathbf{r}). \qquad (5.57) \end{aligned}$$

[17]In cartesians, $\mathbf{r} = (a\theta + a\sin\theta, a + a\sin\theta)$, which describes a cycloid. Thus $\mathbf{v} = (1+\cos\theta, -\sin\theta)a\dot\theta)$, etc.

The component of $\mathbf{H}$ in the direction of the axis of rotation is given by

$$\begin{aligned}
\mathbf{H}\cdot\mathbf{k} &= m\omega\mathbf{r}\times(\mathbf{k}\times\mathbf{r})\cdot\mathbf{k} \\
&= m\omega(\mathbf{r}\times\mathbf{k})\cdot(\mathbf{r}\times\mathbf{k}) \qquad \text{(using the cyclic property of scalar triple products)} \\
&= ma^2\omega \qquad (5.58)
\end{aligned}$$

where $a = r\sin\theta$, which is the distance of the particle from the axis, in agreement with the formulae used in section 3 for the angular momentum of a particle moving in a plane ('$mh = mr^2\dot{\theta}$'). Thus

$$\mathbf{H}\cdot\mathbf{k} = I\omega. \qquad (5.59)$$

We can generalise this result using the inertia tensor. In suffix notation, equation (5.56) becomes

$$\begin{aligned}
H_i &= m\left(\mathbf{r}\times(\boldsymbol{\omega}\times\mathbf{r})\right)_i \\
&= m\left((\mathbf{r}\cdot\mathbf{r})\boldsymbol{\omega} - (\mathbf{r}\cdot\boldsymbol{\omega})\mathbf{r}\right)_i \\
&= m\left((\mathbf{r}\cdot\mathbf{r})\delta_{ij} - x_i x_j\right)\omega_j \\
&= I_{ij}\omega_j \qquad (5.60)
\end{aligned}$$

This is the generalisation of $H = ma^2\omega$, which we can retrieve by dotting both sides of equation (5.60) with k_i:

$$H_i k_i = I_{ij}\omega_j k_i = \omega I_{ij} k_i k_j = \omega I, \qquad (5.61)$$

where I here is the angular momentum about the axis $\mathbf{k}$.

If we apply $\dot{\mathbf{H}} = \mathbf{G}$ to (5.60), we obtain

$$G_i = \frac{d(I_{ij}\omega_j)}{dt}. \qquad (5.62)$$

This is as far as we can go, in general: differentiating the inertia tensor takes us into rather dangerous territory.[18]

However, if the axis of rotation $\mathbf{k}$ is fixed, we can write

$$\boldsymbol{\omega} = \omega\mathbf{k} \qquad \text{and} \qquad \dot{\boldsymbol{\omega}} = \dot{\omega}\mathbf{k}$$

in which case we can differentiate equation (5.61) to obtain

$$\mathbf{G}\cdot\mathbf{k} = I\dot{\omega}$$

where I is the moment of inertia of about the axis $\mathbf{k}$.

This equation, derived for a single particle, applies also to a rigid system of particles rotating about a fixed axis since we can sum the moments of inertia of the individual particles on the left hand side and sum the torques on the individual particles on the right hand side (provided the internal forces are central — see section 5.1.2).

5.7.5 Uniform gravitation forces

We consider the effect of a uniform gravitational force acting on a rigid body. The (external) force on each particle of the body is given by

$$\mathbf{F}_i^e = m_i\mathbf{g}$$

so that the total external force $\mathbf{F}^e$ has a simple form:

$$\mathbf{F}^e \equiv \sum_{i=1}^{n}\mathbf{F}_i^e = M\mathbf{g}.$$

Thus the position vector $\mathbf{R}$ of the centre of mass satisfies

$$M\ddot{\mathbf{R}} = M\mathbf{g}$$

[18] With respect to axes fixed in the body, which is how the inertia tensor would normally by calculated, the components of the inertia tensor (I_{ij}) would be constant (because the particles of the body are fixed relative to one another). However, these axes would move with the body, so the derivative would have to take into account rotation and acceleration of the body: the corresponding equations are named after Euler.

so the centre of mass of the body moves along the trajectory of a single particle in the gravitational field.

The total torque about an arbitrary fixed point, which we take to be the origin $\mathbf{r} = \mathbf{0}$, is given by

$$\mathbf{G} = \sum_{i=1}^{n} \mathbf{r}_i \times (m_i \mathbf{g}) = M\mathbf{R} \times \mathbf{g} \tag{5.63}$$

so again the effect of the gravitational field on the body is the same as that a single particle of mass M situated at its centre of mass:

$$\frac{d\mathbf{H}}{dt} = M\mathbf{R} \times \mathbf{g},$$

where $\mathbf{R}$ is the position vector of the centre of mass with respect to the origin. In particular, if we choose the origin to be at the centre of mass, so that $\mathbf{R} = \mathbf{0}$, we see that the total torque about the centre of mass is zero[19]:

$$\frac{d\mathbf{H}_M}{dt} = \mathbf{0}. \tag{5.64}$$

If the body is rotating about a fixed axis $\mathbf{k}$, we have

$$I\dot{\omega} = M(\mathbf{R} \times \mathbf{g}) \cdot \mathbf{k}. \tag{5.65}$$

We define the gravitational potential energy of the body to be the sum of the potential energies of the individual particles:

$$\phi = -\sum_{i=1}^{n} m_i \mathbf{g} \cdot \mathbf{r}_i = -M\mathbf{g} \cdot \mathbf{R} \tag{5.66}$$

(Recall that in the usual axes, $\mathbf{g} = (0, 0, -g)$, so the minus sign in this expression cancels with the minus sign in the above equation to give the usual '$+mgz$' expression for potential energy.)

From equation (5.66) we see that the gravitational potential of a rigid body is the same as that of a single particle of mass M at the centre of mass.

Now we define the total energy E of the body by

$$E = T + \phi$$

where T is the sum of the kinetic energies of the particles and ϕ is the sum of the gravitational potential energies of the particles. We can easily differentiate this to show that it is constant using equations (5.64), (5.55) and (5.66).

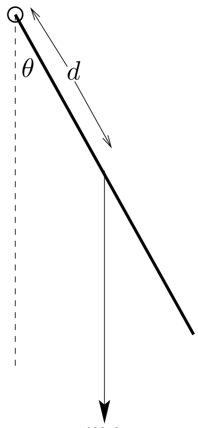
5.7.6 Example: motion of a swinging rod

A pendulum consists of a thin rod of mass m suspended from one end in such a way that it can swing in one vertical plane. Let I be the angular momentum of the rod about a horizontal axis perpendicular to the rod passing through its pivoted end. Let d be the distance between the centre of mass and the pivoted end of the rod.

[19]Recall that

$$\frac{d\mathbf{H}_M}{dt} = \mathbf{G}_M,$$

(see equation (5.21)), even though the centre of mass is not fixed and might even be accelerating.



The kinetic energy of the swinging rod is

$$\tfrac{1}{2}I\dot\theta^2.$$

We could have obtained this a different way, using the expression (5.23) for the kinetic energy of a body:

$$\begin{aligned} T &= \text{`KE of centre of mass'} + \text{`KE relative to centre of mass'} \\ &= \tfrac{1}{2}m(d\dot\theta)^2 + \tfrac{1}{2}I_M\dot\theta^2, \end{aligned}$$

where I_M is the moment of inertia of the rod about the centre of mass. These two expressions agree provided

$$I = I_M + md^2$$

which holds by virtue of the parallel axis theorem.

The potential energy of the rod relative the point of suspension is

$$-mgd\cos\theta$$

because, by (5.66), for a rigid body in a gravitational field, the potential energy is that of a particle of mass m situated at the centre of mass.

Thus the total energy E is given by

$$E = \tfrac{1}{2}I\dot\theta^2 - mgd\cos\theta \qquad (\dagger)$$

from which we obtain, by differentiating with respect to time t and cancelling an overall factor of $\dot\theta$, the equation of motion

$$I\ddot\theta = -mgd\sin\theta.$$

This is equivalent to a simple pendulum of length I/md and the period of small oscillations is $2\pi\sqrt{I/mgd}$.

We could equally well have obtained the equation of motion by taking moments about the point of suspension, using torque = moment of inertia × angular acceleration:

$$-mg\times d\sin\theta = I\ddot\theta\,,$$

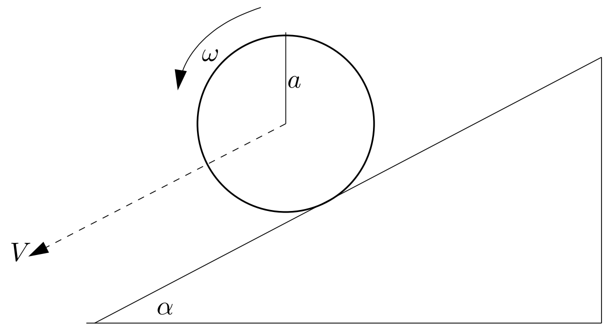
since, by (5.65) the total torque is the same as for a particle of mass m situated at the centre of mass.

If we needed to integrate this, it would probably be best to start instead with the energy conservation equation (†), which is already a first integral of the equation of motion. A further integration gives an elliptic integral.

5.7.7 Example: rolling disc

A disc of mass M and radius a rolls without slipping down a line of greatest slope of an inclined plane of angle α. The plane of the disc is vertical. The moment of inertial of the disc about and axis through its centre perpendicular to the plane of the disc is I.

The motion of the disc consists of the linear motion of the centre of mass, which moves with speed V down the plane, and rotation about the centre of mass with angular speed ω, as shown. The angular velocity vector sticks out of the paper (use the righ-handed corkscrew rule).



The point on the circumference of the disc that is instantaneously in contact with the plane is instantaneously at rest, because of the no-slip condition. This means that V and ω are related by

$$V - a\omega = 0\,.$$

This comes from $\mathbf{V} + \boldsymbol{\omega} \times \mathbf{y} = \mathbf{0}$, were $\mathbf{y}$ is the position vector of the instantaneous point of contact with respect to the centre of the disc. Taking instead the instantaneous point of contact as the origin, this equation says that the velocity the centre of mass is due to the rotation with angular velocity of $\boldsymbol{\omega}$ about the point of contact.

Using conservation of energy

The kinetic energy (using the result that the total KE is 'KE of centre of mass' plus 'KE relative to centre of mass') of the disc is

$$\tfrac{1}{2}MV^2 + \tfrac{1}{2}I\omega^2 = \tfrac{1}{2}MV^2 + \tfrac{1}{2}\;I(V/a)^2 = \tfrac{1}{2}(I/a^2 + M)V^2.$$

Let x be the distance down the plane that the disc has rolled at time t, so that $\dot{x} = V$. Then conserving energy[20] gives

$$\tfrac{1}{2}(I/a^2 + M)\dot{x}^2 - Mgx\sin\alpha = \text{constant}.$$

We have used the result (5.66) that the gravitational potential energy of the body (in a uniform gravitational field) is the same as that of a single particle of mass M at the centre of mass. Curiously, the quickest way to integrate this is to differentiate it and cancel a factor of $\dot{x}$, leaving a linear equation:

$$(I/a^2 + M)\ddot{x} = Mg\sin\alpha$$

which can then be integrated twice. We see that the acceleration of a rolling disc is less, by a factor of $1 + I/Ma^2$, than that of the same disc sliding without rolling down the same plane.

Using forces

The external forces on the disc are shown in the diagram on the next page.

Again regarding the disc as a system of particles, we have two general results derived from Newton's second law:

$$M\ddot{\mathbf{R}} = \mathbf{F}^e = \sum_{i=1}^{n} m_i\mathbf{g} + \mathbf{N} + \mathbf{F} = M\mathbf{g} + \mathbf{N} + \mathbf{F} \tag{5.67}$$

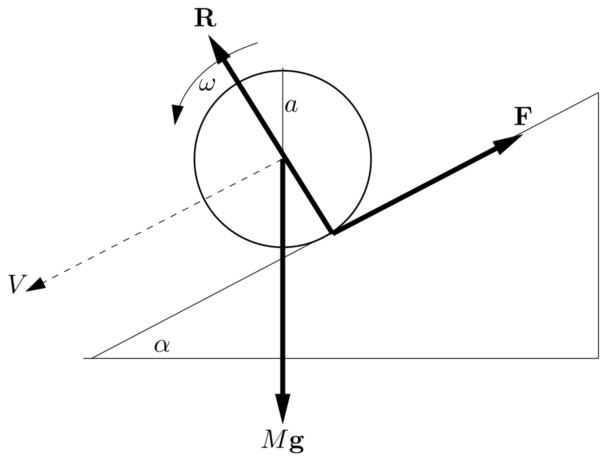
where $\mathbf{R}$ is the position of the centre of mass, $\mathbf{F}^e$ is the sum of the external forces namely gravity, and the frictional $\mathbf{F}$ and the normal reaction $\mathbf{N}$ which act $\mathbf{y}_P$, the point of contact between the disc and the plane;

$$\frac{d\mathbf{H}_M}{dt} = \mathbf{G}_M = \sum_{i=0}^{n} \mathbf{y}_i \times (m_i\mathbf{g}) + \mathbf{y}_P \times \mathbf{F} + \mathbf{y}_P \times \mathbf{N} \tag{5.68}$$

$$= \mathbf{y}_P \times \mathbf{F} \tag{5.69}$$

where $\mathbf{H}_M$ is the total angular momentum about the centre of mass and $\mathbf{G}_M$ is the total external torque about the centre of mass (i.e. the total moment of the external forces). The first term on the right of equation (5.68) vanishes because $\sum m_i\mathbf{y}_i = 0$ and last term vanishes because the force $\mathbf{N}$ is parallel to $\mathbf{y}_P$. Note that

[20]The minus sign in the following equation arises because x is distance *down* the plane.



The component of equation (5.67) parallel to the plane, and the component of (5.69) in the direction $\mathbf{k}$ parallel to the axis of rotation give, respectively,

$$m\dot{V} = Mg\sin\alpha$$
$$I\dot{\omega} = aF,$$

using $\mathbf{H}_M \cdot \mathbf{k} = I\omega$. The second of the above equations holds because the direction of the axis of rotation is constant (even though the axis itself is translating). Eliminating F from these equations, and using $\omega = V/a$ gives

$$(m + I/a^2)\dot{V} = Mg\sin\alpha \tag{5.70}$$

which is the same equation as motion as that derived using conservation of energy.

We could have obtained this same result more directly using again

$$\dot{\mathbf{H}} = \mathbf{G}$$

where now the angular momentum and the torque are about the point of contact between the disc and the plane. Again $\mathbf{H} \cdot \mathbf{k} = I'\omega$, but I' is now the moment of inertial of the disc about an axis parallel to $\mathbf{k}$ passing through the point of contact, which by the parallel axis theorem is given by

$$I' = I + ma^2.$$

This gives the same equation as (5.70), since the shortest distance between the line of action of the force of gravity acting through the centre of the disc and the point of contact is $a\sin\alpha$.

Note that the ω in this calculation is the same as the ω that led directly to (5.70), because angular velocity is independent of position.

Chapter 6

Special Relativity

6.1 Basic concepts

6.1.1 Comparison with Newtonian dynamics

Three basic assumptions of Newtonian dynamics are:

1. There is a special class of reference frames, called inertial frames: an inertial frame is one in which Newton's first law holds.
2. Any two inertial frames are related by a Galilean transformation.

$2'$ Time is absolute.

3. Newton's second law (with Galilean invariant forces, including the universal law of gravitation) holds in any inertial frame. This is sometimes called the *principle of Galilean relativity.*

The corresponding assumptions from Special Relativity are:

1. Same as the Newtonian assumption.
2. Any two inertial frames are related by a Lorentz transformation[1].

$2'$ The speed of light is absolute.

3. Newton's second law, with momentum suitably defined, (with Lorentz invariant forces, including the Lorentz force[2]) holds all inertial frames. This is sometimes called the *principle of special relativity.*[3]

6.1.2 Consequences for Special Relativity

Some consequent differences between Newtonian Dynamics and Special Relativity are as follows.

- Newtonian absolute time is replaced by absolute speed of light: in Special Relativity, the speed of light in vacuo is c (about 3×10^8 metres per second) for all observers.
- In Special Relativity, the notion of simultaneity is frame-dependent. This is a simple consequence of the fact that time is no longer absolute: each observer has his or own time and so observers will not agree on which events are simultaneous.
- In Special Relativity, lengths are contracted and time is dilated[4] in frames moving relative to the observer. This does not follow obviously from the assumptions of the previous section; it will be derived once we have an algebraic form for Lorentz transformations.

[1]See section 6.3 for the definition of a Lorentz transformation.

[2]Note to self: did I ever mention that the Lorentz force is not invariant under Galilean transformations? (It is approximately invariant for small speeds.) If not, it was a bit naughty to stick the Lorentz force on the right hand side of N2.

[3]The 'special' refers to the fact that only inertial frames are considered; the principle of general relativity is (roughly) that the laws of physics are the same in all frames. It was (I think) Max Planck who used the name 'special relativity'; Einstein himself, in his 1905 paper, did not call his theory anything particular: his paper was entitled 'On the electodynamics of moving bodies'.

[4]clocks tick slower

- There is a different velocity addition law in Special Relativity which means that the speed of light cannot be attained by massive particles. Clearly, if the speed of light is to the same in all frames, we cannot use the Galilean formula $c' = c + v$ for the velocity c' of light in a crame that is moving at velocity v with respect to the original frame.
- New definitions for energy and momentum; for example '$E = mc^2$' (though we will never use the definition of energy in this form).

6.1.3 The need for Special Relativity

The invariance of the speed of light that led Einstein to propose a theory invariant under Lorentz transformations instead of Galilean transformations is both an experimental and a theoretical result.

The most significant experiment was performed by Michelson and Morley (actually a series of experiments beginning in 1887), in which they tried to detect the motion of the Earth through the ether[5]. However, they failed to detect any difference in light speed in orthogonal directions even when they took the apparatus up a high mountain (in case the ether was stuck to the Earth at low altitudes). Fitzgerald and Lorentz independently proposed that the effect of the ether on a moving body was to exert a force that compressed the body by a factor $(1 - v^2/c^2)^{\frac{1}{2}}$. It was later shown that the idea of the contraction being caused by a physical force was untenable.

On the theoretical side, Maxwell's equations for electromagnetic fields showed that the fields propagated at a speed almost exactly equal to the measured speed of light. In 1903, Lorentz showed that Maxwell's equations are invariant under what are now called *Lorentz transformations*[6], implying that the speed at which the waves travel would be the same in all frames related by Lorentz transformations.

In 1905, from a very different starting point, Einstein showed that if the speed of light is the same in all frames, then the frames are related by Lorentz transformations.

6.2 Space-time diagrams

As in Newtonian dynamics, it is often convenient to exhibit information by means of a diagram. The usual convention in relativity is to have the time axis vertical and the space axis or axes horizontal; this is the other way round from the usual convention in A-level mechanics (for example). Thus a particle at rest moves along a vertical line. In particular, an observer at rest at the spatial origin $x = 0$ moves up the time axis. In this case, we say that the axes correspond to the *observer's frame* or the *observer's frame of reference*; the observer's frame is the set of axes for which the observer moves along the trajectory $x = 0$.

The trajectory of a particle or observer in space-time is called a *world line.* The gradient of the world line of a physical particle must be steeper than that of a light ray because it cannot move faster than the speed of light.[7] In two dimensions (one time, one space), the pair of light trajectories, one moving to the left and one to the right, through a point P is called the *light cone.* In three dimensions (one time, two space), the light cone through P would look like a double cone, one cone representing light spreading out from P and the other representing light converging on P. In four dimensions (one time, three space) the light cone would consist of one set of expanding spherical surfaces corresponding to a flash of light spreading out from P, and one set of contracting spherical surfaces representing light arriving at P.[8]

The units on the axes of a space-time diagram are chosen to have the same dimension by using ct rather than t on the time axis, so that one unit up the ct axis represent a time taken by light to travel one unit of distance. This means that the light cone, represented by dashed lines, is at 45° to the axes.

Each point in a space-time diagram represents an *event.*[9] If the gradient of the straight line joining two points in a space-time diagram is greater than one in magnitude (i.e. it is steeper than

[5]It was thought at that time that light must propagate through some all pervading medium, which was called the ether. If this were the case, and the Earth was moving through the ether, light should travel at different speeds in orthogonal directions: compare with a swimmer in a river swimming either with the current or across the current.

[6]Lorentz discovered these transformations after playing around with the effect of different transformations on the Maxwell equations which were known not to be invariant under Galilean transformations.

[7]If we assume that there is an inertial frame in which the observer is at rest, and that inertial frames are related by Lorentz transformations, then there is no inertial frame in which an observer moves as fast as the speed of light. This will become clearer in the next section.

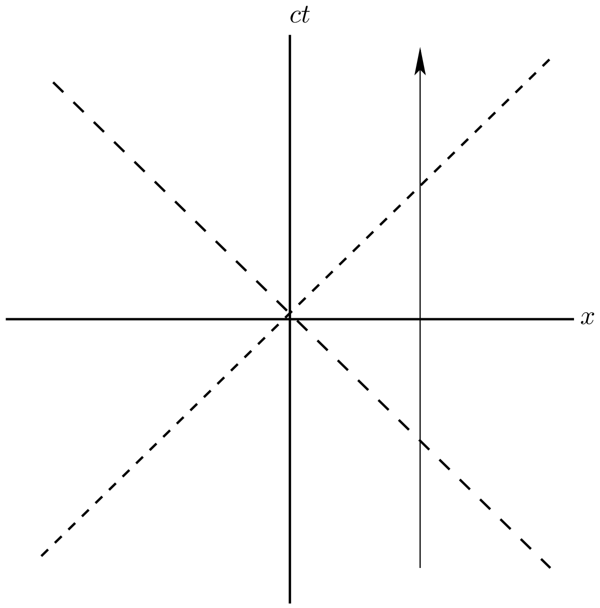
[8]These light cones are difficult to picture, being three-dimensional surfaces embedded in four-dimensional space-time; suppressing one space dimension results in the easily visualised double cones.

[9]We will not always make a distinction between the event and the point representing it.

a light ray), then the corresponding events are said to be *causally related*; this means that one event could influence the other by sending a signal which did not have to travel faster than the speed of light. If event E_1 occurs within the light cone of event E_2, then E_1 and E_2 are causally related.

If events represented by points A and B in a space-time diagram are causally related, the straight line joining A and B in the space-time diagram is said to be *time-like*; if A lies on the light cone through B, the line is said to be *null* or *light-like*; otherwise, A and B are connected by a *space-like* line.

The *rest frame* of an observer is a frame in which the observer's x coordinate is constant. The world line is represented on a space-time diagram by a vertical line; normally one refers to the rest frame as being the frame in which the observer is at $x = 0$, so his or her world line is the ct axis. The *steeper* the gradient of the world line of an observer, the *slower* the observer is moving in that frame.



A space-time diagram. The x-axis is horizontal and the ct axis is vertical. The vertical line with an arrow is the world line of an observer at rest in these axes.

6.3 Lorentz transformations (one space dimension case)

6.3.1 Definition

In two dimensions (one space, one time) we define a Lorentz transformation relating a frame S with coordinates (ct, x) and a frame S' with coordinates (ct', x') moving with velocity v relative to S by

$$x' = \gamma(x - vt) \tag{6.1}$$
$$t' = \gamma(t - vx/c^2) \tag{6.2}$$

where

$$\gamma = \left(1 - \frac{v^2}{c^2}\right)^{-\frac{1}{2}}. \tag{6.3}$$

Note that $\gamma \geq 1$ (and is real, since we assume that $|v/c| \leq 1$). If $v/c \ll 1$ we retrieve the Newtonian limit which is the Galilean transformation $x' = x - vt$, $t' = t$.

Since the transformations are linear, they also relate displacements between two space-time events, which may be infinitesimally close:

$$\Delta x' = \gamma(\Delta x - v\Delta t) \tag{6.4}$$
$$\Delta t' = \gamma(\Delta t - v\Delta x/c^2) \tag{6.5}$$

or

$$dx' = \gamma(dx - vdt) \tag{6.6}$$
$$dt' = \gamma(dt - vdx/c^2) \tag{6.7}$$

If we take as an axiom in Special Relativity that inertial frames are related by Lorentz transformations instead of by the Newtonian Galilean transformations, we find that the speed of light is the same in all inertial frames: if $x = ct$ then

$$\begin{aligned} x' &= \gamma(x - vt) && \text{(by the definition of the Lorentz transformation (6.1))} \\ &= \gamma(ct - vt) && \text{(since } x = ct\text{)} \\ &= \gamma(ct - vx/c) && \text{(since } t = x/c\text{)} \\ &= ct' && \text{(by the definition of the Lorentz transformation (6.2))} \end{aligned}$$

Alternatively, we could define Special Relativity by the requirement that the transformations between inertial frames are such that:

- Newton's first law holds in all inertial frames, so that a particle moving with constant velocity in one inertial frame also moves with constant velocity in any other inertial frame (or, to put it geometrically, straight lines in space-time diagrams map to straight lines);
- the speed of light is the same in all inertial frames (or, to put it geometrically, the straight lines in space-time diagrams with gradient $\pm c$ map to straight lines with the same gradient.

We can then show that inertial frames are related by Lorentz transformations. The proof is not very interesting.[10]

It is helpful to picture the effect of a Lorentz transformation in a space-time diagram. The important thing to notice is that in the space-time diagram corresponding to the frame S, the axes of a frame S' moving with speed v relative to S are inclined symmetrically to the light cone. We can see this from the transformations (6.1) and (6.2). The time axis for S' has equation $x' = 0$ which corresponds to

$$x = vt = \frac{v}{c}(ct)$$

in S (not surprisingly: the frame is moving with speed v relative to S). The space axis for S' has equation $t' = 0$ which corresponds to

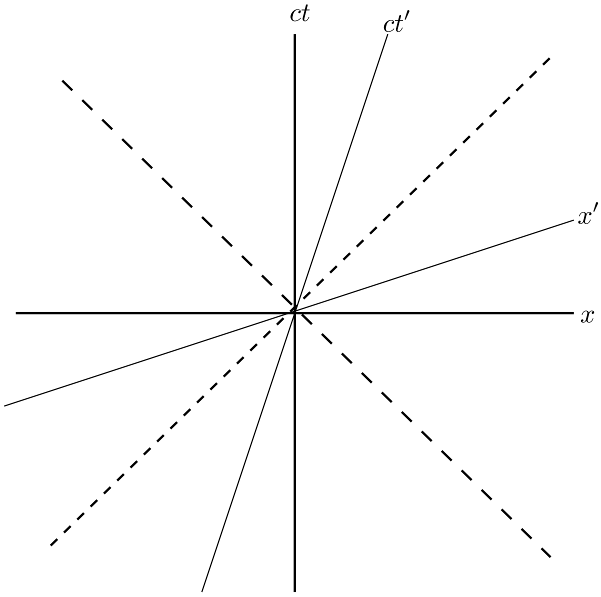
$$t = vx/c^2, \quad \text{i.e.} \quad x = \frac{c}{v}(ct)$$

in S. Since the gradients of these lines are reciprocal, the two lines are equally inclined to the t-axis and the x-axis respectively, as shown below.

[10] If the transformation from frame S to frame S' is $x' = f(x,t), t' = g(x,t)$ a particle moving with speed v in S moves with speed v' in S', where

$$v' = \frac{f_x v + f_t}{g_x v + g_t},$$

where the subscripts denote partial differentiation. Since v' is constant whenever v is constant, and remembering that f and g do not depend on v, one can argue or prove that g and g are linear in both x and t. Finally, using that conditions that $v' = c$ when $v = c$, and $v' = -c$ when $v = -c$ nails the Lorentz transformation (up to an overall multiple which corresponds to a change of units of both x and t. The argument is a bit finicky, because it is too glib to say that 'f and g do not depend on v'. It is true, of course; but the *arguments* of these functions can (and will), so there is still quite a lot of work to do.)



6.3.2 Matrix representation

Sometimes it is helpful to write the Lorentz transformation in matrix form:

$$\begin{pmatrix} ct' \\ x' \end{pmatrix} = \begin{pmatrix} \gamma & -v\gamma/c \\ -v\gamma/c & \gamma \end{pmatrix} \begin{pmatrix} ct \\ x \end{pmatrix}. \tag{6.8}$$

The matrix representation of the two-dimensional Lorentz transformation works out neatly if the transformation is parameterised by a hyperbolic angle β, instead of by v/c, where β is defined by

$$\tanh\beta = \frac{v}{c}. \tag{6.9}$$

In terms of this new parameter,

$$\gamma = \frac{1}{\sqrt{1-\tanh^2\beta}} = \cosh\beta$$

so we can write the matrix (6.8) as

$$\begin{pmatrix} \cosh\beta & -\sinh\beta \\ -\sinh\beta & \cosh\beta \end{pmatrix} \equiv L(\beta). \tag{6.10}$$

Note the similarity to a rotation matrix. In fact (and this is a spoiler for section 6.6), the exact correspondence is that the Lorentz transformation preserves the matrix

$$\eta = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{6.11}$$

in the sense

$$L^T \eta L = \eta \tag{6.12}$$

as can easily be verified directly, and a rotation matrix preserves the unit matrix I:

$$R^T I R = I.$$

We see immediately from this matrix form (6.10) that Lorentz transformations form a group, the multiplication law being

$$L(\beta_1)L(\beta_2) = L(\beta_1+\beta_2).$$

In particular, $(L(\beta))^{-1} = L(-\beta)$ which is what one would expect on physical grounds.[11]

[11]You ought to be able to return from the moving frame S' to your original frame S, which is moving with velocity $-v$ relative to your transformation, by applying the inverse transformation.

6.3.3 Simultaneity

In Newtonian dynamics, where time is absolute, all observers agree on the set of events that are simultaneous with a given event. Clearly, they will not agree in Special Relativity because the transformation law (6.2) means that time is different in different frames.[12] To an observer stationary in S, all events with

$$t = \text{constant}$$

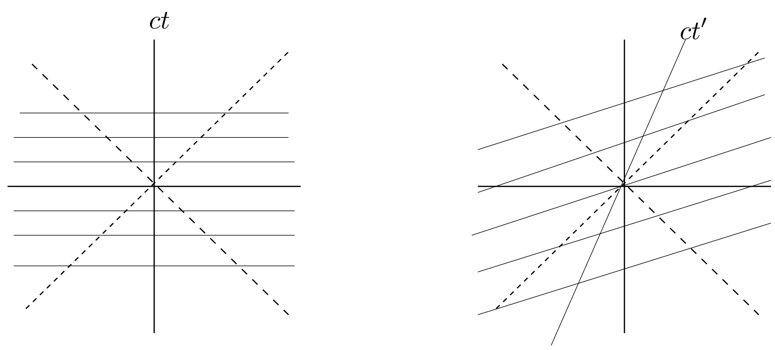
are simultaneous. Similarly, to an observer stationary in S', the all events with

$$t' = \text{constant}$$

but this corresponds to

$$t - vx/c^2 = \text{constant}$$

in S.



Lines of simultaneity in S (horizontal, in the diagram on the left) and in S′ (slanted, in the diagram on the right).

Note that, given an event E outside the light cone of the event $(0,0)$, there exists a Lorentz transformation to a new frame in which these two events are simultaneous, but if the event E lies inside the light cone of the event $(0,0)$, there is no Lorentz transformation to a new frame in which these two events are simultaneous. It therefore makes sense to say that events within the light cone of $(0,0)$ with $t > 0$ (the *future* light cone) are to the future of $(0,0)$ and events within the light cone with $t < 0$ (the *past* light cone) are to the past of $(0,0)$; this statement is invariant under Lorentz transformations (there is no frame, related to the original frame by a Lorentz transformation, in which the statement is false). It follows that the event at the origin can be influenced by events within or on the past light cone but not by other events (assuming of course that no signals can travel fast than the speed of light).

6.4 Time dilation and length contraction

These two phenomena are closely related: the same observations can often be explained in one frame in terms of time dilation and in another frame in terms of length contraction. It is helpful, before defining exactly what these terms mean and doing the mathematics, to give an example.

6.4.1 Meson decay

A muon is a charged particle that decays into an electron or positron, a neutrino and an anti-neutrino:

$$\mu^+ \to e^+ + n_1 + n_2 \quad \text{or} \quad \mu^- \to e^- + n_1 + n_2$$

Muons occur in cosmic rays travelling through the atmosphere at speeds very close to that of light.[13] In 1941, Rossi and Hall measured the flux of muons in a laboratory at 6300 feet above sea level (top of Mount Washington) and in a laboratory at 2000 feet above sea level (bottom of Mount Washington).

At the top they measured 550 muons per hour. At the foot (simultaneously) they measured 422 muons per hour. The half-life of the muon is 1.56 microseconds.

[12] They don't have to argue over it: all they have to do is respect the other observer's point of view and then there is no inconsistency, though you wouldn't think so from the spats that occur in some web sites.

[13] They are created by collisions between protons in cosmic rays with atoms of air in the upper atmosphere.

From this information, one can calculate how long the muons spend travelling between the two laboratories and hence the speed of the muons *and the speed turns out to be much faster than the speed of light.* The muons travel a distance D at speed v taking time $T = D/v$; during this time the number $N(T)$ of muons remaining is given by

$$N(T) = N(0)\left(\tfrac{1}{2}\right)^{(T/T_{\text{half}})}.$$

Thus

$$v = \frac{D}{T} = \frac{D \log \frac{1}{2}}{T_{\text{half}} \log\left(N(T)/N(0)\right)} = \frac{4300 \log \frac{1}{2}}{1.56 \log(422/550)} = 7212$$

in units of feet per microsecond. The speed of light in these units is about 1000.

To put it another way, the observed flux of muons at the lower laboratory is far too high for particles covering the distance at less than the speed of light: many more should have decayed in the travel time.

What is the explanation? As we will see, it depends on whether we work in the rest frame of the muon or the rest frame of the laboratory, the two being in relative motion at close to the speed of light.

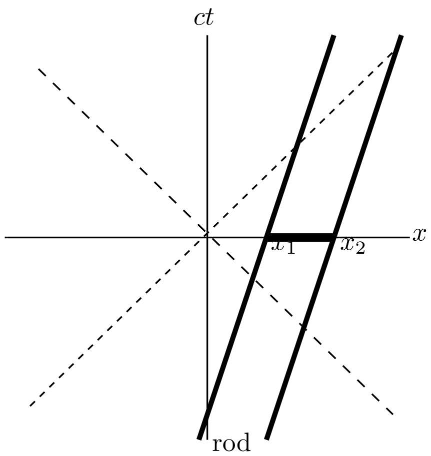
In the rest frame of the laboratory, the explanation is time dilation: time in the moving frame is dilated relative to time in the rest frame, which means that clocks are ticking slower, by a factor of γ, in the moving frame. Suppose the travel time is T seconds as measured (by distance/speed) in the laboratory frame. Then in this interval, only T/γ seconds have elapsed in the moving muon frame, so far fewer muons will decay, corresponding to a half-life of 1.56γ microseconds.

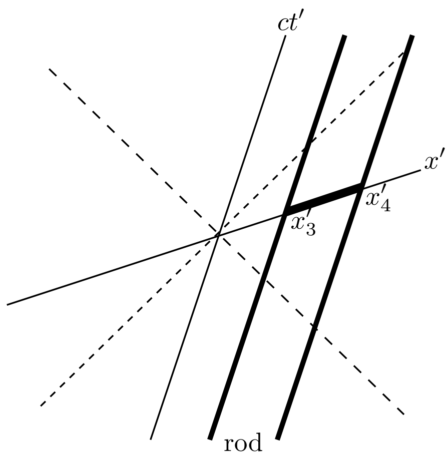
But how can this be explained in the muon frame, where the half-life is 1.56 microseconds? The explanation now is length contraction. In the rest frame of the muon, Mount Washington, which is zooming towards the muon at high speed, is only $6000/\gamma$ feet high, because lengths of moving rulers are contracted. Thus the time taken to cover this contracted distance is short: only T/γ seconds. There is little time for the muons to decay.

6.4.2 Length contraction

We consider a rod of length L' (in its own rest frame). It is moving (along its length, like a javelin) at velocity v relative to the lab frame.

The space-time diagrams below show the world lines of the two ends of the rod, in the lab frame (left hand diagram) and in the rest frame of the rod (right hand diagram).





The event with coordinates $(0, x_1)$ and $(0, x_2)$ in S are simultaneous in S and the events with coordinates $(0, x_3')$ and $(0, x_4')$ in S' are simultaneous in S'.[14]

What does it mean to say that the rod has length L in S? Clearly, you have to contrive to measure the position of each end of the rod at the same time, and then subtract the x coordinates: that will give you the length. Thus the length L in S of the moving rod is given by

$$L = x_2 - x_1$$

[14] You can save a bit of writing by choosing axes such that $x_1 = x_3' = 0$; it is just a translation of the origins of the two frames.

and the length L' of the rod in its rest frame S' is given by

$$L' = x'_3 - x'_4.$$

Our task is to relate L and L'.

This is only a matter of coordinate geometry.[15]

The equation in S' of the world line of the leading end of the rod is

$$x' = x'_4$$

and the equation of the trailing end is

$$x' = x'_3.$$

We can use the Lorentz transformations (6.1) and (6.2) to transform these lines into S:

$$\gamma(x - vt) = x'_4$$

and

$$\gamma(x - vt) = x'_3.$$

These lines meet the x-axis ($t = 0$) at $x = x_2$ and $x = x_1$ respectively, so

$$\gamma x_1 = x'_3 \quad \text{and} \quad \gamma x_2 = x'_4.$$

Putting these together;

$$L = x_2 - x_1 = \frac{x'_3 - x'_4}{\gamma} = \frac{L'}{\gamma}$$

which represents a contraction since $\gamma \geq 1$; the moving rod is shorter in the laboratory frame than if it were at rest in the laboratory frame.[16]

But hold on! If we look at the lengths of the rod, marked by the short heavy lines in the two diagrams, it seems that we have got it the wrong way round: the $(x'_4 - x'_3)$ definitely looks longer than $(x_2 - x_1)$. This is a trap: lengths in space-time diagrams are not like lengths in the more familiar x-y plane and we must rely on our calculations.[17]

6.4.3 Time dilation

We consider a clock moving at velocity v with respect to the laboratory frame. It ticks at times t'_1 and t'_2, measured in its rest frame S', where $t'_2 - t'_1 = \Delta t'$. These events (the ticks) have coordinates $(ct'_1, 0)$ and $(ct'_2, 0)$ in S',

Let the coordinates of the ticks in S be (ct_1, x_1) and (ct_2, x_2). Then using the Lorentz transformation (6.2) with v replaced by $-v$ (because the lab frame moves with velocity $-v$ with respect to the clock's frame)

$$t_1 = \gamma t'_1 \quad \text{and} \quad t_2 = \gamma t'_2. \tag{6.13}$$

and Δt, the time interval between ticks in S, is given by

$$\Delta t = \gamma \Delta t' \geq \Delta t'.$$

In this interval, an identical clock at rest in S would have ticked more than once.

The precise situation described above concerns a lab frame S and two events in another (moving) frame S'. The time interval between the two events is longer in the lab frame than in the moving frame. This is exactly the situation in the meson decay experiment. At the first event, there is a flux of N mesons per second and at the second event there is a flux of $N/2$ mesons per second; the time interval between these events is the half-life. The time between these two events in the lab frame is longer (the half-life is longer in the lab frame than in the rest frame of the

[15] There are lots of ways of getting the required answer. Sometimes, you find at the end that you have a factor $\gamma(1 - v^2/c^2)$, which simplifies to γ^{-1}; this usually indicates that there was a quicker way, for example, transforming from S to S' rather from S' to S.

[16] Note that we say 'is shorter' not 'appears shorter'. It is important not confuse the fact of being shorter with the appearance of being shorter: the appearance might be the result of taking a photograph, for example, which would be affected by the time delay of light travelling to the camera.

[17] To anticipate section 6.5: lengths are shorter the closer the inclination to the 45° of the null cone. This is because instead of the Euclidean norm (Pythagoras), one must use the norm $||(ct, x)|| = (c^2t^2 - x^2)^{\frac{1}{2}}$.

meson), which is why fewer decay in the transit time than would have been expected in Newtonian dynamics. This is *time dilation*.[18]

Of course, one can present this result from the point of view of the observer in S', who sees S as the moving frame. Such an observer would notice that the time interval between two events in his or her rest frame (S') is shorter than in any other frame.

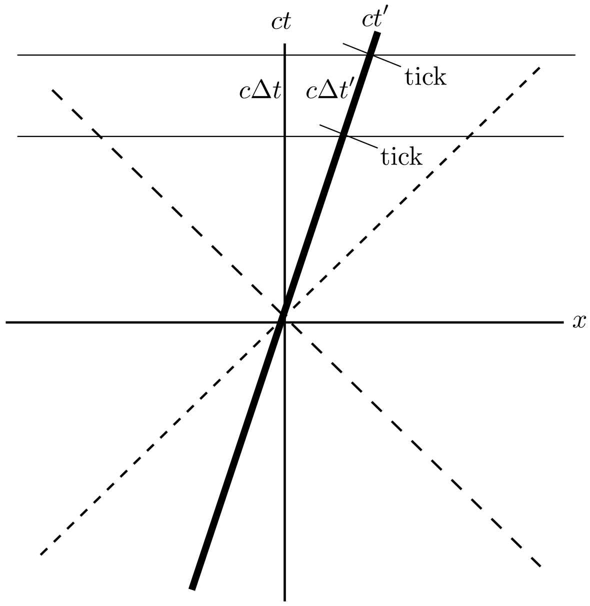
It is instructive to see what would have happened if we had woodenly used the forward transformation (6.2 with v instead of $-v$) to obtain (6.13). We have, for both t'_1 and t'_2,

$$t' = \gamma(t - vx/c^2), \quad 0 = x' = \gamma(x - vt).$$

The second of these equations gives $x = vt$, which we knew anyway, and substituting this into the first equation gives

$$\begin{aligned} t' &= \frac{t - vx/c^2}{\sqrt{1 - v^2/c^2}} && \text{(by definition of } \gamma\text{)} \\ &= \frac{t - v^2t/c^2}{\sqrt{1 - v^2/c^2}} && \text{(using } x = vt\text{)} \\ &= t/\gamma \end{aligned}$$

which is the same result as before.



6.4.4 The ladder-and-barn non-paradox

A builder runs towards a barn of length L carrying a ladder of length $2L$ at a speed[19] such that $\gamma = 2$ so that the length contraction factor is $\frac{1}{2}$.

- In the barn's rest frame, the moving ladder undergoes length contraction and has length L. It can therefore fit snugly in the barn.
- In the builder's rest frame, the barn is rushing towards the ladder and undergoes length contraction to $L/2$. There is no way the ladder can fit in.

How can these two statements be reconciled?

The answer stems, as is often the case with apparent paradoxes in relativity, from loose use of language. In this case, it is the use of the word 'fit'; what does it mean to say the ladder 'fits'' exactly into the barn? Clearly, we mean that the two events:

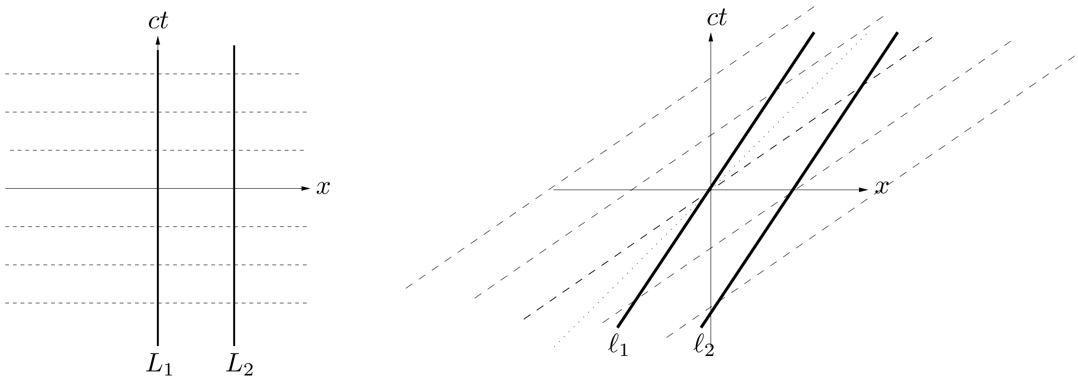
[18] I'm not sure that this is a helpful description: what exactly is dilated?? It is better, as always in Special Relativity, to fix on a precise space-time description of the situation: what events we are considering and in which frame.

[19] If $v = \sqrt{3}c/2$, then $\gamma^{-2} = 1 - 3/4$ and $\gamma = 2$.

(i) front end of ladder hits back of barn; (ii) back end of ladder goes through the door

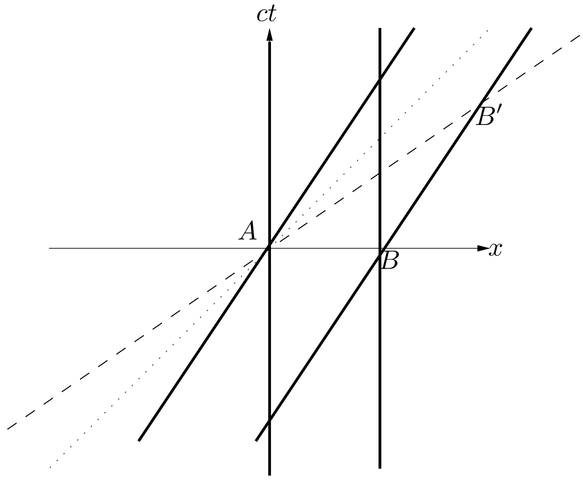
are simultaneous. But observers in different frames do not agree on simultaneity, so 'fit into' is a frame-dependent concept: we should not expect observers in different frames to agree so there is no paradox to account for. The two statements are true and compatible and that is really the end of the story. However, we can investigate further.

The situation can best be understood by means of space-time diagrams.



The lines L_1 ($x = 0$) and L_2 ($x = L$) in the left hand figure are the world lines of the ends of the barn in axes corresponding to the rest-frame of the barn; the barn door is L_1. The lines of simultaneity in the barn frame, $t =$ constant, are shown as broken lines.

The lines ℓ_1 ($x = vt$) and ℓ_2 ($x = vt + L$) in the right hand figure are the world lines of the ends of the ladder, again in the axes corresponding to the rest-frame of the barn. The lines of simultaneity in the ladder frame, $t' =$ constant, i.e. $t - vx/c^2 =$ constant, are shown as broken lines. The light cone is shown as a dotted line.



The previous diagrams are superimposed. The event A is 'back of ladder goes through barn door'. The event B is 'front of ladder hits back of barn'. In the barn frame, these events are simultaneous. In the ladder frame, A is simultaneous with B': by the time A occurs, the front of the ladder has burst through the back of the barn.

Regarded from the point of view of a space-time diagram, the paradox dissolves. One consequence of time not being invariant under Lorentz transformations is that the ladder 'fits in' the barn in one frame but does not 'fit in' in another.

6.4.5 The twins non-paradox

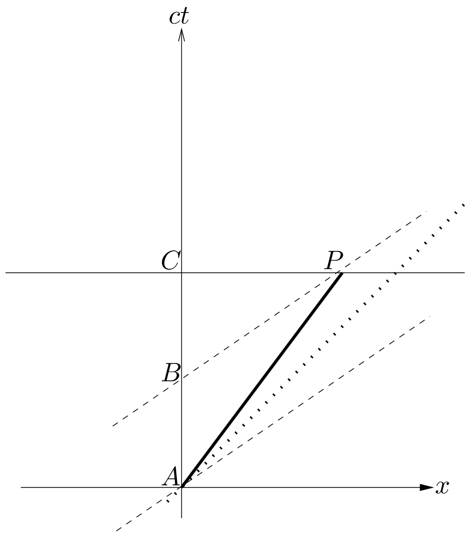
Twins Alice and Bob synchronise watches in an inertial frame and then Bob sets off at speed $\sqrt{3}c/2$, which corresponds to $\gamma = 2$. When Bob has been travelling for a time T according to Alice, he reaches Proxima Centauri[20] and turns round by means of accelerations that are very large in his

[20]The closest star to the Sun: about 4.2 light years away.

frame and goes back to Alice at the same speed. Since Bob is in a moving frame, relative to Alice, his time runs slower by a factor of γ than Alice's, so he will only have aged by $2T \times \frac{1}{2}$ on the two legs of the journey. Thus when they meet up again, Alice has aged by $2T$ but Bob has aged only by T. *This is not the paradox: it is just a fact of life.*[21]

The difficulty some people have with Alice and Bob is the apparent symmetry: surely exactly the same argument could be made, from Bob's point of view, to show that Alice would be the younger when they met again? But the same argument *cannot* be made for Bob because the situation is not symmetric: Alice's frame is inertial, whereas Bob has to accelerate to turn round: while he is accelerating, his frame is not inertial.

BUT, some people might say, suppose we just consider the event of Bob's arrival at Proxima Centauri, so as not to worry about acceleration. Now the situation is symmetric. Surely from Alice's point of view, when Bob arrives he will have aged half as much as Alice, and from Bob's point of view, when he arrives, Alice will have aged half as much as Bob? The answer to this is a simple 'yes'. Surely, they would then say, this doesn't make sense? But it does, as long as you are careful about the word 'when'.



In the above diagram, Alice's world line is the ct (containing points A, B and C) axis and Bob's world line is the line containing A and P. P represents the event 'Bob arrives at Proxima Centauri'.

The line CP is a line of simultaneity in Alice's frame and C is the event 'Alice is at this point in space-time *when* — according to Alice — Bob arrives at Proxima Centauri'; the first use of the word 'when'.

The line BP is a line of simultaneity in Bob's frame and B is the event 'Alice is at this point in space-time *when* — according to Bob — he arrives Proxima Centauri'; the second use of the word 'when'. The two 'whens' don't mean the same thing, since one is a 'when' in Alice's frame the other is a 'when' in Bob's frame.

We can do the calculation. Let us assume for simplicity that Bob sets off the moment he is born. The event C has coordinates $(cT, 0)$ in Alice's frame, and the event P has coordinates (cT, vT). In Bob's frame, the elapsed time T' is given by the Lorentz transformation:

$$T' = \gamma(T - v^2T/c^2) = T/\gamma = \tfrac{1}{2}T.$$

This is just the usual time dilation calculation. Thus Bob and Alice agree that Bob's age at Proxima Centauri is $\frac{1}{2}T$. In Alice's frame, Bob has aged half as much as Alice.

We now work out the coordinates of the event B, sticking with Alice's frame. The line of simultaneity, BP has equation $t' = \frac{1}{2}T$, i.e. (using a Lorentz transformation)

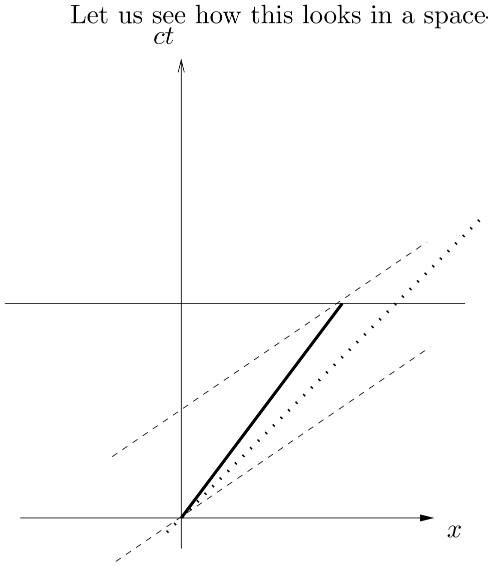
$$\gamma(t + vx/c^2) = \frac{1}{2}T$$

[21]In 1971, Hafele and Keating packed four atomic (caesium) clocks into suitcases and went round the Earth, in different directions, on commercial flights. When they returned, they found that the clocks were slightly behind a clock remaining at the first airport. The result was somewhat inconclusive. The calculations are complicated by the fact that the rate of the clocks is also affected by the gravitational field: clocks run slower in stronger fields, and in fact the two affects balance at $3R/2$ (where R is the radius of the Earth). Thus the heights of the aircraft had to be taken into account as well as their speeds, and it turns out that the two effects are of comparable magnitude, namely of the order of 100 nanoseconds.

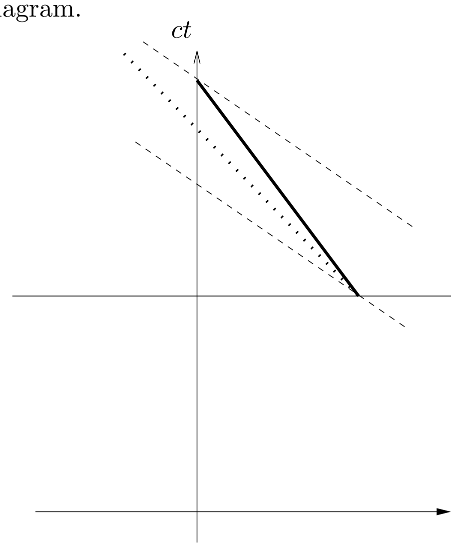
so the point B, for which $x = 0$, has coordinates $(\frac{1}{2}cT/\gamma, 0)$, i.e. $(\frac{1}{4}cT, 0)$. Alice's age when, according to Bob, he arrives at Proxima Centauri is therefore $\frac{1}{4}T$, which is indeed half of Bob's age. So no paradox there either.

BUT, some other people might say, suppose Bob does not turn round but just synchronises his watch at Proxima Centauri with that of another astronaut, Bob$'$, who is going at speed v in the opposite direction (like two trains passing at a station). Each leg of the journey is then symmetric, so why should Alice age faster or slower Bob and Bob$'$ during their legs of the journey? There's no mystery here, either: the situation is indeed symmetric and Alice does indeed age by the same amount as Bob+Bob$'$. But at the synchronisation event, Bob and Bob$'$ do not agree on Alice's age, because in their different frames the synchronisation event is simultaneous with different times in Alice's life.

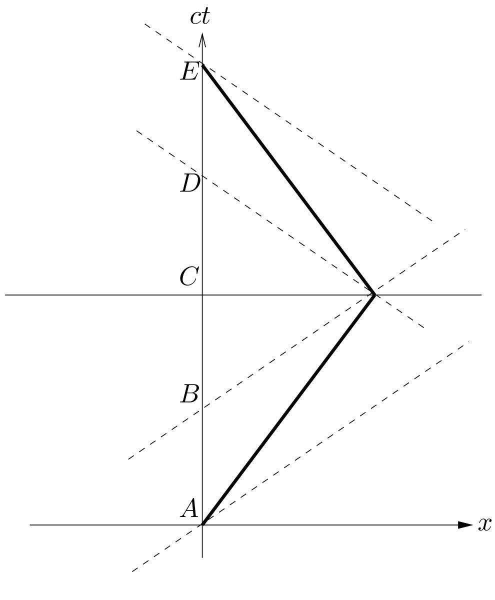
Let us see how this looks in a space-time diagram.



The outward journey. The heavy line is Bob's world line. The dotted line through the origin is the light cone. The dashed lines are the lines of simultaneity in Bob's frame.



The return journey. The heavy line is the world line of Bob$'$. The dotted line through the turn-round event is the light cone. The dashed lines are the lines of simultaneity in the frame of Bob$'$.



The superposition of the previous two pictures.

As before, Bob ages by $\frac{1}{2}T$ on the outward journey to Proxima Centauri. By symmetry Bob$'$ ages by $\frac{1}{2}T$ on the inward journey from Proxima Centauri.

However, according to Bob's idea of time, the clock synchronisation occurs when Alice is at B, and according to Bob$'$'s it occurs when Alice is at D. Thus Bob$'$'s clock will read time T when he meets Alice and Alice's clock will read $2T$. But the time Alice spends between B and D is accounted for by Bob in his journey *after* Proxima Centauri and by Bob$'$ in his journey *before* reaching Proxima Centauri, so the two Bobs would say that, while they were travelling between Earth and Proxima Centauri, Alice travelled from A to B and then from D to E, taking on her clock a total time T — the same as the journey time of the two Bobs.

Finally, we see that if, instead of meeting Bob$'$, Bob turns round at Proxima Centauri, Alice ages rapidly (according to Bob) from B to D while he is changing direction.

6.5 Velocity transformation

The frame S' moves at constant velocity v with respect to a frame S. A particle P moves with constant velocity u in S. What is the velocity u' of P in S'?

In Newtonian physics, the answer is simple: $u' = u - v$. However, this cannot be the right answer in Special Relativity because it would imply that the speed of light would not be invariant: $c' = c - v$, where c' is the speed of light in S'.

The world line of P has equation $x = ut$ in S and $x' = u't'$ in S'. We have:

$$\begin{aligned} u' &= \frac{x'}{t'} \\ &= \frac{\gamma(x - vt)}{\gamma(t - vx/c^2)} && \text{(using Lorentz transformations)} \\ &= \frac{x - vt}{t - vx/c^2} && \text{(cancelling } \gamma) \\ &= \frac{ut - vt}{t - uvt/c^2} && (x = ut, \text{ twice}) \\ &= \frac{u - v}{1 - uv/c^2} && (6.14) \end{aligned}$$

which is the required result. When $uv \ll c^2$, we retrieve the Newtonian result.

If $u = c$ then

$$u' = \frac{c - v}{1 - cv/c^2} = c$$

so the velocity transformation law preserves that speed of light: it is the same to all observers.

Note that this velocity transformation law does not permit a transformation to a speed greater than c. If $u < c$ and $|v| < c$, then

$$\begin{aligned} c - u' &= c - \frac{u - v}{1 - uv/c^2} \\ &= \frac{c(1 - uv/c^2) - (u - v)}{1 - uv/c^2} \\ &= \frac{c(c - u)(c + v)}{c^2 - uv} > 0. \end{aligned}$$

Thus $u' < c$. Similarly, we can show that $-u' < c$ if $-u < c$, so $|u'| < c$ if $|u| < c$ (assuming $|v| < c$).

If we write the velocities u, u' and v in terms of hyperbolic angles:

$$\frac{u}{c} = \tanh\beta, \qquad \frac{u'}{c} = \tanh\beta' \qquad \frac{v}{c} = \tanh\alpha$$

and substitute into the velocity transformation law (6.14) we find

$$\tanh\beta' = \frac{\tanh\beta - \tanh\alpha}{1 - \tanh\alpha\tanh\beta} = \tanh(\beta - \alpha)$$

so an alternative form of the transformation law is

$$\beta' = \beta - \alpha$$

i.e.

$$\tanh^{-1}(u'/c) = \tanh^{-1}(u/c) - \tanh^{-1}(v/c).$$

6.6 Proper time

6.6.1 Definition

The fact that the concept of time is frame dependent can be rather unsettling. It would be good to have some quantity that corresponds to time but does not vary at the whim of the observer. Such a quantity exists and is called *proper time*.

We define the proper time $\Delta\tau$ between two events E_1 and E_2 on the world line of an observer, with coordinates (ct, x) and $(ct + c\Delta t, x + \Delta x)$ respectively, by

$$c^2(\Delta\tau)^2 = c^2(\Delta t)^2 - (\Delta x)^2. \tag{6.15}$$

For this to make sense, we require either $c\Delta t \geq |\Delta x|$ or $c\Delta t \leq -|\Delta x|$, which means that either E_1 is in the past light cone of E_2, or vice versa (which is why we specified that the events are on the world line of an observer). In the former case, we choose $\Delta\tau \geq 0$.

Note that $(\Delta\tau)^2 > 0$ if the vector joining E_1 and E_2 is time-like, and $(\Delta\tau)^2 = 0$ if the vector joining E_1 and E_2 is null. For points joined by a space-like vector, one can define *proper distance*, s, by

$$(\Delta s)^2 = (\Delta x)^2 - c^2(\Delta t)^2\,.$$

In the rest frame of the observer, $\Delta x = 0$, and $\Delta\tau = \Delta t$. Thus proper time measures rest frame time.

We can verify that proper time is Lorentz invariant by brute force. Since Lorentz transformations are linear, we have

$$\begin{aligned}\Delta t' &= \gamma(\Delta t - v\Delta x/c^2)\\ \Delta x' &= \gamma(\Delta x - v\Delta t).\end{aligned} \tag{6.16}$$

Therefore

$$\begin{aligned}c^2(\Delta\tau')^2 &\equiv c^2(\Delta t')^2 - (\Delta x')^2\\ &= c^2\left(\gamma(\Delta t - v\Delta x/c^2)\right)^2 - \left(\gamma(\Delta x - v\Delta t)\right)^2\\ &= \gamma^2\left(c^2(\Delta t)^2 - v^2(\Delta t)^2\right) + \gamma^2\left(v^2(\Delta x)^2/c^2 - (\Delta x)^2\right) && \text{(the cross terms cancel)}\\ &= c^2(\Delta t)^2 - (\Delta x)^2 && (\gamma^2 = (1 - v^2/c^2)^{-1})\\ &= c^2(\Delta\tau)^2 && \text{(as required)}\end{aligned}$$

However, we can most easily verify that this quantity is Lorentz invariant using the matrix form of the Lorentz transformation:

$$\begin{aligned}c^2(\Delta\tau)^2 &\equiv \begin{pmatrix} c\Delta t, & \Delta x\end{pmatrix}\begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix}\begin{pmatrix}c\Delta t\\ \Delta x\end{pmatrix} && (6.17)\\ &= \begin{pmatrix} c\Delta t, & \Delta x\end{pmatrix}\eta\begin{pmatrix}c\Delta t\\ \Delta x\end{pmatrix} && \text{(by definition (6.11))}\\ &= \begin{pmatrix} c\Delta t, & \Delta x\end{pmatrix}L^T\eta L\begin{pmatrix}c\Delta t\\ \Delta x\end{pmatrix} && \text{(see (6.12))}\\ &= \begin{pmatrix} c\Delta t', & \Delta x'\end{pmatrix}\eta\begin{pmatrix}c\Delta t'\\ \Delta x'\end{pmatrix}\\ &= c^2(\Delta t')^2 - (\Delta x')^2\\ &\equiv c^2(\Delta\tau')^2.\end{aligned}$$

The proper time between two infinitesimally separated points (ct, x) and $(ct + cdt, x + dx)$ is given by

$$c^2d\tau^2 = c^2dt^2 - dx^2. \tag{6.18}$$

Again, we note that if these points represent events on the world line of an observer then in the rest frame of the observer

$$d\tau = dt_{\text{rest}}$$

so infinitesimal proper time measures infinitesimal time displacements in the rest frame; ticks of the observer's clock. Comparing with (6.18) we see that in a general frame

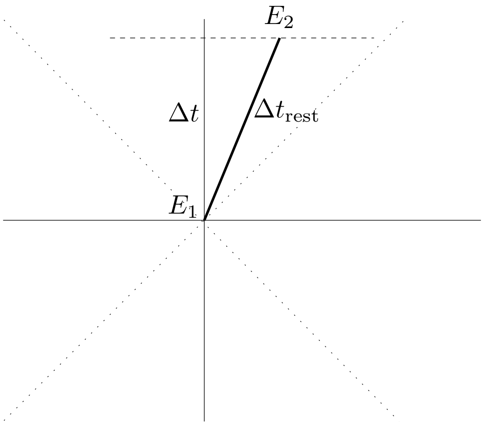
$$\sqrt{dt^2 - dx^2/c^2} = d\tau = dt_{\text{rest}}$$

so

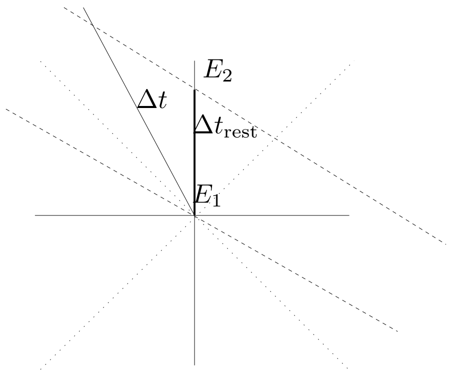
$$dt > dt_{\text{rest}}$$

which is time dilation.

Rather confusingly, this comparison cannot be easily inferred from a space-time diagram: in fact, there is a temptation (which must be resisted) to assume that lengths behave in a Euclidean fashion (so that Pythagoras applies) which would lead to completely incorrect conclusions[22]. This illustrated in the following diagrams. In both diagrams, the null cone is represented by the pair of dotted lines.



E_1 and E_2 are events on the world line of an observer represented in this diagram as moving with respect to a 'general' observer. The horizontal dashed line is simultaneity for the 'general' observer. $\Delta t_{\text{rest}} < \Delta t$, even though it doesn't look as if this is the case.



E_1 and E_2 are events on the world line of an observer represented in this diagram as stationary. The world line of the 'general' observer is slanted. The dashed lines are simultaneity for the 'general' observer. $\Delta t_{\text{rest}} < \Delta t$, which does look to be the case, though the difference in times is not as much as it 'seems'.

Now suppose that these two points represent events on the world line of an observer moving with velocity v with respect to some given frame (the 'lab frame'). The world line can be written in the form

$$x = x(t)$$

in which case

$$v = \frac{dx}{dt}.$$

Then

$$\begin{aligned} d\tau^2 &= dt^2 \left(1 - \frac{1}{c^2}\frac{(dx)^2}{(dt)^2}\right) \\ &= dt^2 \left(1 - \frac{v^2}{c^2}\right) \end{aligned}$$

and

$$\frac{dt}{d\tau} = \gamma. \tag{6.19}$$

Note that here γ is not linked to the velocity between two frames explicitly, though of course it is implicitly related to the velocity between the rest frame of the observer and the lab frame.

[22]You wouldn't dream of using Pythagoras in an A-level type distance-time graph; nevertheless, it is a good instinct to suppose that there is some concept of length that can be applied to a space-time diagram in Special Relativity and this is touched on in the next section.

The total time that elapses on the world-line of an observer moving with (not necesarily constant) velocity in a frame S is given by

$$\int d\tau = \int \gamma^{-1} dt;$$

this is the observer's actual time (clock or biological).

We can use proper time to derive the velocity addition formula (6.14) for an observer moving with non-constant velocity. We parameterise the observer's world line by τ:

$$\begin{aligned} x = x(\tau)\,, \quad t = t(\tau) \quad &\text{in } S \\ x' = x'(\tau)\,, \quad t' = t'(\tau) \quad &\text{in } S' \end{aligned}$$

and

$$u = \frac{dx}{d\tau}\Big/\frac{dt}{d\tau}, \quad u' = \frac{dx'}{d\tau}\Big/\frac{dt'}{d\tau},$$

We can differentiate the Lorentz transformation (6.2) and (6.1) to obtain

$$\begin{aligned} \frac{dx'}{d\tau} &= \gamma\left(\frac{dx}{d\tau} - v\frac{dt}{d\tau}\right) = \gamma(u-v)\frac{dt}{d\tau} \\ \frac{dt'}{d\tau} &= \gamma\left(\frac{dt}{d\tau} - (v/c^2)\frac{dx}{d\tau}\right) = \gamma(1-(uv/c^2))\frac{dt}{d\tau} \end{aligned} \tag{6.20}$$

and dividing these expressions gives

$$u' = \frac{u-v}{1-uv/c^2}$$

as required. [23]

6.6.2 Line elements and metrics

The infinitesimal version of the formula (6.17) for proper time

$$c^2 d\tau^2 \equiv \begin{pmatrix} cdt, & dx \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} cdt \\ dx \end{pmatrix}$$

can be written in suffix notation

$$c^2 d\tau^2 = \eta_{ij} dx^i\, dx^j, \tag{6.21}$$

where the infinitesimal vector[24] dx^i has components (cdt, dx). The matrix η in the expression (6.21) is called the *metric*: it tells us how the invariant distance can be calculated from the coordinates. The whole of the right hand side of equation (6.21) is called the *line element*.

In two dimensional Euclidean space in Cartesian coordinate,

$$ds^2 = dx^2 + dy^2,$$

which corresponds to a metric which is the unit 2×2 matrix. In plane polar coordinates,

$$ds^2 = dr^2 + r^2 d\theta^2$$

so the metric is

$$\begin{pmatrix} 1 & 0 \\ 0 & r^2 \end{pmatrix}.$$

On the surface of a sphere of radius a, infinitesimal distances are given by

$$ds^2 = a^2 d\theta^2 + a^2 \sin^2\theta d\phi^2,$$

corresponding to the metric

$$\begin{pmatrix} a^2 & 0 \\ 0 & a^2\sin^2\theta \end{pmatrix}.$$

This metric describes a curved surface, and the curvature and other properties of the surface can be deduced from the metric.

[23] Note to self: how about adding something about the action of the Lorentz group on space-time (the orbits are hyperbolas given by $(ct)^2 - x^2 = constant$, etc).

[24] The index is conventionally placed upstairs (superscript), rather than downstairs (subscript) as in normal vector calculus. The reason for this is related to the way that this vector transforms under Lorentz transformations. We will not need to worry about this in this course, but it seems best to use the correct convention.

6.7 Constant acceleration in Special Relativity

This section is not examinable and is included for interest

It is often said, erroneously, that Special Relativity cannot deal with acceleration because it deals only with inertial frames, and that therefore acceleration must be the preserve of General Relativity. We must, of course, only allow transformations between inertial frames; the frames must not accelerate, but the observers in the frame can move as the please. Special Relativity can deal with anything kinematic but General Relativity is required when gravitational forces are present.

As an example of non-uniform motion, we consider an observer who is moving with constant acceleration.

The first step is to define what we mean by 'constant acceleration' which is certainly a frame-dependent concept. The most common situation is that of an observer in a rocket experiencing a constant 'G-force' due to the rocket thrust. This corresponds to the acceleration measured in the instantaneous (inertial) rest frame of the rocket being constant (acceleration having the usual definition of dv/dt), so we take this to be our definition.

For reasons that will later become clear, we need to determine the way that acceleration transforms under Lorentz transformations. We can do this in a number of ways. We will here start with the velocity transformation law (6.20) for an observer with world line given in S by $\big(ct(\tau), x(\tau)\big)$ and in S' by $\big(ct'(\tau), x'(\tau)\big)$. Forgetting the acceleration problem for the moment, we assume that these frames have a constant relative velocity v.

The velocities u and u' in the two frames are related by

$$u' = \frac{u-v}{1-uv/c^2} \equiv \frac{(c^2/v)(1-v^2/c^2)}{1-uv/c^2} - \frac{c^2}{v}$$

(the equivalent form is just a bit of algebra to obtain a useful expression). Differentiating this with respect to τ gives

$$\frac{du'}{d\tau} = \frac{1-v^2/c^2}{(1-uv/c^2)^2}\frac{du}{d\tau}. \tag{6.22}$$

The acceleration, a, in S is by definition du/dt and similarly for S' so

$$\begin{aligned}
a' &= \frac{du'}{dt'} \\
&= \frac{du'}{d\tau}\bigg/\frac{dt'}{d\tau} \\
&= \frac{1-v^2/c^2}{(1-uv/c^2)^2}\frac{du}{d\tau}\bigg/\frac{dt'}{d\tau} && \text{(using (6.22)} \\
&= \frac{1-v^2/c^2}{(1-uv/c^2)^2}\frac{du}{d\tau}\bigg/\gamma(1-uv/c^2)\frac{dt}{d\tau} && \text{(using (6.20))} \\
&= \frac{(1-v^2/c^2)^{\frac{3}{2}}}{(1-uv/c^2)^3}a. && (6.23)
\end{aligned}$$

As mentioned above there are other ways of obtaining this result; for example, more elegantly using four-vectors (see section 6.7).

In the situation we have in mind, S' is the instantaneous rest frame of the accelerating observer, so that $u' = 0$ and $u = v$, and the acceleration a' in this frame is constant (i.e. independent of v). Thus (6.23) becomes

$$a = (1-u^2/c^2)^{\frac{3}{2}}a'$$

Now

$$a = \frac{du}{d\tau}\bigg/\frac{dt}{d\tau} \quad \text{and} \quad \frac{dt}{d\tau} = (1-u^2/c^2)^{-\frac{1}{2}}$$

so we can find the parameterised equation of the world line by integrating

$$\frac{du}{d\tau} = a\frac{dt}{d\tau} = (1-u^2/c^2)a'.$$

This gives

$$u = c\tanh(a'\tau/c) \qquad \text{(choosing the origin of } \tau \text{ so that } u=0 \text{ when } \tau=0)$$

and hence

$$\gamma = \cosh(a'\tau/c).$$

Then from $dt/d\tau = \gamma$, we find that

$$t = c/a' \sinh(a'\tau/c) \qquad \text{(choosing the origin of } t \text{ such that } t = 0 \text{ when } \tau = 0\text{)}$$

Finally,

$$\begin{aligned} \frac{dx}{d\tau} &= \frac{dx}{dt}\frac{dt}{d\tau} \\ &= u\gamma \\ &= c\sinh(a'\tau/c) \end{aligned}$$

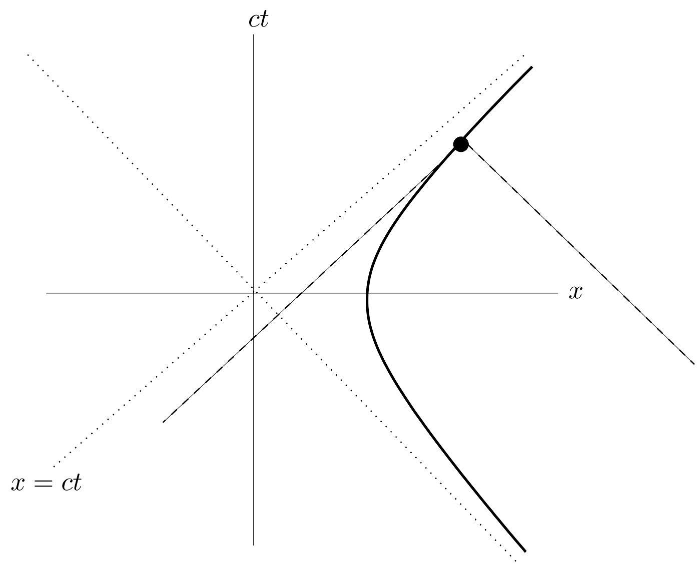
so

$$x = c^2/a' \cosh(a'\tau/c). \qquad \text{(choosing the origin of } x \text{ such that } x = c^2/a' \text{ when } t = 0\text{)}$$

Uniformly accelerated particles therefore move on rectangular hyperbolas of the form

$$x^2 - (ct)^2 = (c^2/a')^2.$$

The diagram shows the trajectory. The dotted lines are the light cones. An event taking place within the dashed lines can influence an accelerated observer at the position shown, but events taking place outside the dashed lines would have to move faster than the speed of light to do so. As $\tau \to \infty$, the whole of the space-time to the left of the dotted line $x = ct$ would be inaccessible to the observer. This line is called the *Rindler event horizon* for the accelerated observer. In some ways, it performs the same function as the event horizon of a black hole. In particular, the observer has to accelerate to avoid falling through it and anything happening on the other side would be hidden to the observer. Of course, the accelerating observer could just stop accelerating whereas the observer in a black hole space-time can do nothing to affect the event horizon.



The space-time diagram for an accelerated observer. The thick hyperbola is the observer's world line. An observer 'below' the dashed lines could in principle send a message to the observer marked as a heavy dot; other observers could not.

6.8 Four-vectors

6.8.1 definitions

In $1 + 3$ dimensions (i.e. one time dimension and three space dimensions), we write the position 4-vector X in the form

$$X = \begin{pmatrix} ct \\ x \\ y \\ z \end{pmatrix}$$

or sometimes

$$X = \begin{pmatrix} ct \\ \mathbf{x} \end{pmatrix}.$$

A (4×4) matrix L representing a transformation from a frame S to a frame S'

$$\begin{pmatrix} ct' \\ x' \\ y' \\ z' \end{pmatrix} \equiv X' = LX \tag{6.24}$$

is a Lorentz transformation if

$$L^T \eta L = \eta \tag{6.25}$$

where η is the diagonal matrix (the Minkowski metric) defined by

$$\eta = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}.$$

This is exactly as in the $(1+1)$ dimensional case (6.11).

Lorentz transformations determine a mapping from an inertial from S to a new frame S' via the mapping (6.24) regarded as a coordinate transformation. We can easily verify that S' is inertial and that the speed of light is preserved.[25]

The Lorentz transformations form a matrix group, as can easily be verified from the definition (6.25). It is called $O(1,3)$. It includes reflections (including $t \to -t$), which one usually wants to exclude. The subgroup of the Lorentz group consisting of matrices that satisfy $\det L = +1$ is called the *proper Lorentz group* and denoted $SO(1,3)$. Those which preserve the orientation of the spatial axes and the direction of time form the *restricted Lorentz group*[26], denoted $SO^+(1,3)$.

It is as first sight rather difficult to understand what sort of matrices might satisfy the definition (6.25) of a Lorentz transformation matrix. However, we can make some progress by counting arguments. A 4×4 matrix has 16 potentially independent elements. In this case they are constrained by the 10 independent equations of (6.25): there are 10 equations, not 16, because both the left hand side and the right hand side are symmetric matrices. Thus just 6 of the 16 elements of L are independent. We express this by saying that the Lorentz matrices form a 6-parameter group.

This group includes rotations of the spatial axes:

$$L = \left(\begin{array}{c|ccc} 1 & 0 & 0 & 0 \\ \hline 0 & & & \\ 0 & & R & \\ 0 & & & \end{array}\right)$$

where R is 3×3 rotation matrix (so that $R^T I R = I$). These rotations form a 3-parameter subgroup, each parameter being (for example) an angle of rotation about one of the three spatial axes.

The group also includes 'boosts' of the form

$$L = \left(\begin{array}{cc|cc} \gamma & -v\gamma/c & 0 & 0 \\ -\gamma v/c & \gamma & & \\ \hline 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{array}\right). \tag{6.26}$$

This represents a boost in the x direction, but it could equally well have been a boost in the y or z directions. Each of these boosts forms a 1-parameter subgroup (the parameter being the velocity along the relevant axis), so that is another three parameters.

Thus the boosts and rotations together require 6 independent matrix elements, which is the total number available. This indicates that any Lorentz matrix is a combination of rotations and boosts.

We will not be interested in rotations. In fact, we will restrict out attention to transformations of the form (6.26) We do not lose much generality by this restriction, because any proper Lorentz

[25] I should do this explicitly later on.

[26] The restricted Lorentz group is isomorphic to the Möbius group, and this fact is the basis of the twistor theory devised by Roger Penrose as a possible route to quantum gravity.

transformation can be expressed as a spatial rotation followed by a transformation of the form (6.26); this means that we could choose axes (by a spatial rotation) such that the required Lorentz transformation is exactly (6.26).

A *scalar* or *invariant* under Lorentz transformations is a quantity that is the same in any two frames related by a Lorentz transformation. One example of a Lorentz scalar is the proper time between two events in space-time; another is the rest mass of a particle, which is the mass measured in the particle's rest frame and therefore is invariant by definition.

A *4-vector* V is a quantity whose components transform according to the rule

$$\begin{pmatrix} V_0' \\ V_1' \\ V_2' \\ V_3' \end{pmatrix} = L \begin{pmatrix} V_0 \\ V_1 \\ V_2 \\ V_3 \end{pmatrix}$$

where L is the matrix of the Lorentz transformation. The position 4-vector transforms as a 4-vector by definition of the transformation (6.24) itself.

Given any two 4-vectors X and Y, we define an inner product called the *scalar product* by

$$X \cdot Y = X^T \eta Y \equiv X_0 Y_0 - X_1 Y_1 - X_2 Y_2 - X_3 Y_3$$

or if

$$X = \begin{pmatrix} X_0 \\ \mathbf{x} \end{pmatrix} \qquad \text{and} \qquad Y = \begin{pmatrix} Y_0 \\ \mathbf{y} \end{pmatrix}$$

then

$$X \cdot Y = X_0 Y_0 - \mathbf{x} \cdot \mathbf{y}.$$

The scalar product between two 4-vectors is Lorentz invariant:

$$\begin{aligned} X' \cdot Y' &= (X')^T \eta Y' && \text{(by definition of the scalar product)} \\ &= (LX)^T \eta L Y && \text{(transforming back to the unprimed frame)} \\ &= X^T L^T \eta L Y && \\ &= X^T \eta Y && \text{(because } L \text{ is a Lorentz transformation)} \\ &= X \cdot Y && \text{(by definition of the scalar product)} \end{aligned}$$

The scalar product of a 4-vector with itself (the 'length squared') can be negative, positive or zero. Since the scalar product is invariant, this classifies the vectors into three classes:

$$\begin{array}{ll} V \cdot V > 0 & V \text{ is time-like} \\ V \cdot V < 0 & V \text{ is space-like} \\ V \cdot V = 0 & V \text{ is null} \end{array}$$

In particular, the quantity $dX \cdot dX$ is invariant, where dX is the infinitesimal 4-vector connecting two neighbouring space-time points. If dX is time-like or null, we define the infinitesimal proper time, $d\tau$, up to a sign, by

$$c^2 d\tau^2 = dX \cdot dX = c^2 dt^2 - dx^2 - dy^2 - dz^2 = c^2 dt^2 - d\mathbf{x} \cdot d\mathbf{x}. \tag{6.27}$$

In the case of a time-like vector, we choose $d\tau > 0$ if $dt > 0$.

6.8.2 4-velocity

Now consider the infinitesimal position vector dX joining two points on the world line $X = X(\tau)$ of an observer. Since $d\tau$ is invariant and dX is a 4-vector, we can form another 4-vector, the *4-velocity* U of the observer, by dividing them:

$$U = \frac{dX}{d\tau} = \begin{pmatrix} c\dfrac{dt}{d\tau} \\ \dfrac{d\mathbf{x}}{d\tau} \end{pmatrix}. \tag{6.28}$$

In the rest frame of the observer, $\mathbf{x} = \mathbf{0}$ and $d\tau = dt$ so

$$U = \begin{pmatrix} c \\ \mathbf{0} \end{pmatrix} \tag{6.29}$$

again in the rest frame of the observer.

Now we have the first demonstration of the power of the 4-vector formalism. The length squared of the velocity 4-vector is

$$\begin{aligned}
U \cdot U &= c^2 \left(\frac{dt}{d\tau}\right)^2 - \frac{d\mathbf{x}}{d\tau} \cdot \frac{d\mathbf{x}}{d\tau} \\
&= c^2 \left(\frac{dt}{d\tau}\right)^2 - \frac{d\mathbf{x}}{dt} \cdot \frac{d\mathbf{x}}{dt} \left(\frac{dt}{d\tau}\right)^2 \\
&= c^2 \left(\frac{dt}{d\tau}\right)^2 \left(1 - \frac{\mathbf{u} \cdot \mathbf{u}}{c^2}\right) \\
&= c^2 \left(\frac{dt}{d\tau}\right)^2 \frac{1}{\gamma^2} \qquad (6.30)
\end{aligned}$$

But in the rest frame,

$$U \cdot U = c^2 \qquad \text{(from (6.29))}$$

and since the scalar product is the same in all frames, we can equate this with (6.30):

$$c^2 \left(\frac{dt}{d\tau}\right)^2 = c^2$$

i.e.

$$\frac{dt}{d\tau} = \gamma \qquad (6.31)$$

(recall that we choose $d\tau$ to have the same sign as dt.)[27]

We can use this result to tidy up the expression (6.28) for the 4-velocity:

$$\begin{aligned}
U &= \begin{pmatrix} c\dfrac{dt}{d\tau} \\ \dfrac{d\mathbf{x}}{d\tau} \end{pmatrix} \\
&= \begin{pmatrix} c\dfrac{dt}{d\tau} \\ \dfrac{d\mathbf{x}}{dt}\dfrac{dt}{d\tau} \end{pmatrix} \\
&= \begin{pmatrix} c\gamma \\ \mathbf{u}\gamma \end{pmatrix} \qquad \left(\text{using } \frac{dt}{d\tau} = \gamma\right) \\
&= \gamma \begin{pmatrix} c \\ \mathbf{u} \end{pmatrix}. \qquad (6.32)
\end{aligned}$$

6.8.3 4-momentum

The rest mass of a particle, defined in the obvious way as the mass of the particle in its rest frame, is Lorentz invariant by its definition.[28] For a particle of rest mass m and 4-velocity U, we define the momentum 4-vector, or the *4-momentum*, P by

$$P = mU = \begin{pmatrix} mc\gamma \\ m\gamma\mathbf{u} \end{pmatrix} = \begin{pmatrix} E/c \\ \mathbf{p} \end{pmatrix}$$

where, by definition the relativistic energy E and the relativistic 3-momentum $\mathbf{p}$ are

$$E = m\gamma c^2 \quad \text{and} \quad \mathbf{p} = m\gamma\mathbf{u}$$

[27] Actually, we could have obtained the same result in a line from the definition of (6.27) proper time by dividing by dt^2 but that shouldn't prevent us from admiring the neatness of the above calculation.

[28] Note to self: probably better to define relativistic three-momentum and energy earlier? Together with proper time?

Note that the mass here is the rest mass: we will not use any other concept for mass.[29]

The justification for calling the quantity $m\gamma c^2$ the relativistic energy is three-fold. First, we will see later that it is conserved in a wide variety of circumstances, as one would hope. Second, in the Newtonian limit $v/c \ll 1$, it approximates the kinetic energy plus a residual energy:

$$\begin{aligned} E &= mc^2(1 - v^2/c^2)^{-\frac{1}{2}} \\ &\approx mc^2(1 + \tfrac{1}{2}v^2/c^2) && \text{(using the binomial expansion)} \\ &= mc^2 + \tfrac{1}{2}mv^2. \end{aligned}$$

Finally, we can justify defining relativistic energy in this way by considering its rate of change in terms of work done by forces; this will be done in the next section.

Similarly, the relativistic 3-momentum approximates the non-relativistic momentum in the Newtonian limit (setting $\gamma \approx 1$).

We can again perform the trick of comparing the length squared of this 4-vector in the general frame and in the rest frame. In the rest frame,

$$P = \begin{pmatrix} mc \\ \mathbf{0} \end{pmatrix}$$

so

$$P \cdot P = m^2c^2.$$

In the general frame

$$P \cdot P = \frac{E^2}{c^2} - \mathbf{p} \cdot \mathbf{p}.$$

Equating these two expressions gives the relationship between relativistic energy and relativistic 3-momentum:

$$E^2 = p^2c^2 + m^2c^4 \tag{6.33}$$

where $p^2 = \mathbf{p} \cdot \mathbf{p}$. We could of course have worked this out more laboriously by eliminating u from

$$E = m\gamma c^2, \quad p^2 = m^2\gamma^2u^2, \quad \gamma^{-2} = 1 - u^2/c^2.$$

6.8.4 Relativistic energy

We seek to justify defining the relativistic energy by $E = m\gamma c^2$, where m is the rest mass, γ is the usual factor associated with the speed v of the particle, $\gamma = (1 - v^2/c^2)^{-\frac{1}{2}}$.

First we consider the Newtonian situation. For a particle of mass m moving in a force field $\mathbf{F}$ in Newtonian dynamics, we have

$$\text{work done } = \text{force} \times \text{distance moved by force} = \int \mathbf{F} \cdot d\mathbf{x} = \int \mathbf{F} \cdot \mathbf{v}\, dt$$

and hence (using Newton's second law)

$$\text{rate of doing work } = \mathbf{F} \cdot \mathbf{v} = \frac{d\mathbf{p}}{dt} \cdot \mathbf{v} = \frac{d}{dt}\left(\tfrac{1}{2}m\mathbf{v} \cdot \mathbf{v}\right) = \frac{dT}{dt},$$

where T is the kinetic energy of the particle. Thus the rate of doing work on the particle is equal to the increase in its kinetic energy.

We now look at the relativistic situation. In a general inertial frame, the 4-momentum P is given by

$$P = (E/c, \mathbf{p})$$

[29] Sometimes, the concept of relativistic mass is introduced. If m_0 is the rest mass of a particle moving with velocity $\mathbf{v}$, then its relativistic mass is $m_0\gamma$. This leads to the famous energy-mass equivalence equation $E = mc^2$. It is a shame not to be able to use this equation, but it does seem terribly profligate to have two names (E and mc^2) for essentially the same quantity: relativistic energy. It avoids confusion to have just one sort of mass, namely rest mass. This was also Einstein's view, stated many times, despite his name being eternally linked to the famous equation. In his 'Annus mirabilis' (1905) when he wrote 5 papers each of which would have made his name (size of molecules, when it wasn't widely accepted that molecules existed; Brownian motion, evidence for the existence of molecules; viscosity formula; on the motion of electromagnetic bodies, introducing Special Relativity; inertia of a body) he related the change of mass to change of energy, but did not explicitly give the energy of the body as mc^2.

and $P \cdot P = m^2c^2$ which is constant. Thus

$$\begin{aligned} 0 &= \frac{d}{dt}(P \cdot P) && \text{(because } P \cdot P = m^2c^2\text{)} \\ &= \frac{d}{dt}\left(E^2/c^2 - \mathbf{p} \cdot \mathbf{p}\right) \\ &= 2\frac{E}{c^2}\frac{dE}{dt} - 2\mathbf{p} \cdot \frac{d\mathbf{p}}{dt} \\ &= 2m\gamma\frac{dE}{dt} - 2m\gamma\mathbf{v} \cdot \frac{d\mathbf{p}}{dt} && \text{(since } E = m\gamma c^2 \text{ and } \mathbf{p} = m\gamma\mathbf{v}\text{)} \\ &= 2m\gamma\left(\frac{dE}{dt} - \mathbf{v} \cdot \frac{d\mathbf{p}}{dt}\right) \end{aligned}$$

so

$$\frac{dE}{dt} = \mathbf{v} \cdot \mathbf{F} \tag{6.34}$$

where $\mathbf{F}$ is the force acting on the particle and causing its momentum to change. The last inequality follows from one of the basic postulates of Special Relativity, namely that Newton's law holds in inertial frames provided the momentum is correctly defined. Equation (6.34) shows that the rate of change of relativistic energy is equal to the rate of doing work, which helps to justify the definition $E = m\gamma c^2$.

6.8.5 Massless particles

The (rest) mass of a particle in Special Relativity is defined invariantly by the length of is 4-momentum vector

$$P \cdot P = m^2c^2$$

or, equivalently, in terms of its relativistic energy and 3-momentum by $E^2 - (\mathbf{p} \cdot \mathbf{p})c^2 = m^2c^4$. Thus a particle can be massless if and only if its 4-momentum is null (i.e. light-like); the momentum 4-vector points along the instantaneous null cone of the particle. [30] In that case,

$$P \cdot P = 0, \qquad \text{i.e.} \qquad P_0^2 = P_1^2 + P_2^2 + P_3^2 = \mathbf{p} \cdot \mathbf{p}.$$

We can therefore write the 4-momentum in the form

$$P = p(1, \mathbf{k}) \equiv (E/c)(1, \mathbf{k})$$

where $\mathbf{k}$ is unit vector and $p^2 = \mathbf{p} \cdot \mathbf{p}$.

The only massless particles we consider here are photons. There are others (or may be).[31] If we use the relation from quantum mechanics between frequency and energy of a photon

$$E = h\nu \equiv hc/\lambda$$

where ν is frequency and λ is wavelength, we can write the 4-momentum as

$$P = h\nu(1, \mathbf{k}).$$

Photons (and other massless particles) move with the speed of light, so that for the world line of a photon,[32]

$$d\tau^2 = c^2dt^2 - d\mathbf{x} \cdot d\mathbf{x} = 0.$$

This means that we cannot construct a 4-velocity vector for a massless particle by means of

$$V = \frac{dX}{d\tau}.$$

[30] Note to self: perhaps expand this section to explain the wave mechanics and justify that P in this case is a 4-vector.

[31] Gluons, which form quarks, are massless. They are predicted by quantum chromodynamics and have been indirectly detected but not observed. Neutrinos were thought to be massless, but are now thought to have mass; a very small mass (luckily, because 50 trillion solar electron neutrinos pass through the human body every second).

[32] The fact that proper time is zero along the world line of a massless particle is taken on Physics Forum and similar web sites to imply a negative answer the question 'do photons experience time?'; but this question seems to me to be akin to the question 'do slugs feel pain?'.

6.8.6 Transformations of 4-vectors

Choosing axes such that the velocity of the inertial frame $\mathcal{S}'$ with respect to the inertial frame $\mathcal{S}$ is along the x axis, the Lorentz transformation matrix can be written in the form (6.26)

$$\begin{pmatrix} \gamma & -\gamma v/c & 0 & 0 \\ -\gamma v/c & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{6.35}$$

where $\gamma = (1 - v^2/c^2)^{-\frac{1}{2}}$.

Let U be the 4-velocity of a particle P. We will calculate the effect of a transformation from a frame $\mathcal{S}$ to a frame $\mathcal{S}'$ when the 3-velocity $\mathbf{v}$ of $\mathcal{S}'$ relative to $\mathcal{S}$ is not aligned with the 3-velocity $\mathbf{u}$ of P. In $\mathcal{S}$ we can write U without loss of generality (i.e. by making a rotation of axes, if necessary, in the y-z plane, so that the z axis is orthogonal to $\mathbf{u}$) as

$$\begin{pmatrix} \gamma_u c \\ \gamma_u u \cos\alpha \\ \gamma_u u \sin\alpha \\ 0 \end{pmatrix}. \tag{6.36}$$

where $u = |\mathbf{u}|$, and we have defined the gamma factor associated with the 4-velocity U in the normal way:

$$\gamma_u = \frac{1}{(1 - u^2/c^2)^{\frac{1}{2}}}.$$

Expressing the components of U in $\mathcal{S}'$ in the same form as (6.36) and using the Lorentz transform (6.35) gives

$$\begin{pmatrix} \gamma_{u'} c \\ \gamma_{u'} u' \cos\alpha' \\ \gamma_{u'} u' \sin\alpha' \\ 0 \end{pmatrix} = \begin{pmatrix} \gamma & -\gamma v/c & 0 & 0 \\ -\gamma v/c & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \gamma_u c \\ \gamma_u u \cos\alpha \\ \gamma_u u \sin\alpha \\ 0 \end{pmatrix}, \tag{6.37}$$

from which formulae that determine u' and α' can be read off. Writing out the three components gives

$$\gamma_{u'} = \gamma\gamma_u\big(1 - (uv/c^2)\cos\alpha\big) \tag{6.38}$$
$$\gamma_{u'} u' \cos\alpha' = \gamma\gamma_u\big(-v + u\cos\alpha\big) \tag{6.39}$$
$$\gamma_{u'} u' \sin\alpha' = \gamma_u u \sin\alpha \tag{6.40}$$

It is all a bit congested, but if we divide (6.39) by (6.38) something nice happens:

$$u' \cos\alpha' = \frac{u\cos\alpha - v}{1 - uv\cos\alpha/c^2}.$$

This shows that the x component of the velocity (which is the component parallel to $\mathbf{v}$) transforms exactly as in the 2-dimensional case (6.14). We could of course obtain this result in identical form to (6.14) by simply setting $\cos\alpha = 1$ in the matrix equation (6.37).

The second component gives

$$\gamma_{u'} u' = \gamma\gamma_u\big(-v + u\big)$$

which doesn't quite look the same as (6.14). However, the two can be reconciled with a few lines of algebra[33]. Alternatively, if we look at the first component of (6.37) we find a useful identity for $\gamma_{u'}$:

$$\gamma_{u'} = \gamma\gamma_u\big(1 - uv/c^2\big)$$

which reconciles our two formulae without any algebra.

[33] By solving for u':

$$\gamma_{u'} u' = \sqrt{\frac{1}{1/u'^2 - 1/c^2}},$$

so square both sides, take reciprocals, add $1/c^2$, take reciprocals and (magic) take square roots.

Dividing equations (6.40) and (6.39) gives a nice result for the angle α' that the 3-velocity vector $\mathbf{u}'$ in S' makes with the x' axis:

$$\tan\alpha' = \frac{u\sin\alpha}{\gamma(u\cos\alpha - v)}.$$

We can do a similar calculation for the momentum 4-vector of a light ray (a photon). In this case, $|\mathbf{p}| = E/c$, so we can write the components of the 4-momentum vector in the form

$$P = \begin{pmatrix} p \\ p\cos\alpha \\ p\sin\alpha \\ 0 \end{pmatrix}$$

where $p = |\mathbf{p}|$.

Using the Lorentz transformation as before gives

$$\begin{pmatrix} p' \\ p'\cos\alpha' \\ p'\sin\alpha' \\ 0 \end{pmatrix} = \begin{pmatrix} \gamma & -\gamma v/c & 0 & 0 \\ -\gamma v/c & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} p \\ p\cos\alpha \\ p\sin\alpha \\ 0 \end{pmatrix},$$

The t' component gives immediately a formula for the momentum (or energy) of the photon[34]:

$$E' = p'c = \gamma pc\left(1 - \frac{v}{c}\cos\alpha\right) = \gamma E\left(1 - \frac{v}{c}\cos\alpha\right).$$

Dividing the y' and x' components gives an expression for α':

$$\tan\alpha' = \frac{\sin\alpha}{\gamma(\cos\alpha - v/c)}$$

or, after some uninteresting algebra, [35]

$$\tan(\alpha'/2) = \frac{\sqrt{1-v/c}}{\sqrt{1+v/c}}\tan(\alpha/2) = e^{\beta}\tan(\alpha/2),$$

where $\tanh\beta = v/c$. This is the stellar aberration formula: it gives the change in angle (which corresponds to a change in observed position) of light from stars due to the motion of the Earth.[36]

6.9 Conservation of 4-momentum

6.9.1 Newton's second law in Special Relativity

The principle of Special Relativity states that the laws of physics hold in all inertial frames. This implies that they must be expressible in terms of invariants and 4-vectors (or 4-tensors). Accordingly, we take Newton's second law for a single relativistic particle to be

$$\frac{dP}{d\tau} = F \tag{6.41}$$

where F is an appropriately defined 4-vector force. We will here always take $F = 0$ so that 4-momentum is conserved. Following the method of Chapter 5, we can extend this to systems of particles, for which the total momentum (i.e. the sum of the momenta of the individual particles) is conserved.

It is beyond the scope of this course to investigate possible right hand sides for the relativistic version of Newton's second law (6.41). The best non-relativistic example was the inverse square gravitational force. Unfortunately, this will not do for relativity: the distance between two particles is invariant under Galilean transformations but is certainly not invariant under Lorentz transformations (because of length contraction). In order to advance to a relativistic theory of gravitation, one has to go to General Relativity, which is exactly that.

[34] This identical formula appeared in Einstein's 1905 paper 'On the Electrodynamics of Moving Bodies'.

[35] Note to self: possible STEP question?

[36] it is a similar effect to the change of angle of rain falling on the windows of a moving train. Generally aberration is the apparent change of an object's direction of movement relative to an observer's frame of reference.

We have mentioned a number of times that Maxwell's equations are invariant under Lorentz transformations so the Lorentz force ought to be a good example. And indeed it is, but it has to be expressed in terms of a second rank field tensor, which again takes us beyond what is possible in this course.[37]

We will study below the use of momentum conservation in elementary particle reactions. There are two strategies. For both, the first thing to do is to write down the 4-momenta of all the particles involved in a suitable frame, which might be the rest frame of one of the particles or might be the centre of momentum frame, which is the frame in which the total (relativistic) 3-momentum is zero. The first strategy is to obtain equations by conserving the time-like and space-like components of the 4-momenta (i.e. by equating total relativistic energy and momentum). For this strategy, it will usually be necessary to eliminate energy or momentum using (6.33)

$$E^2 = p^2c^2 + m^2c^4.$$

The second strategy is to get on one side of the 4-momentum conservation equation one 4-momentum about which one knows nothing and about which one wishes to find nothing and eliminate it using

$$P \cdot P = m^2c^2.$$

6.9.2 Centre of momentum frame

Sometimes is it helpful to work in the *centre of momentum frame* which is the frame in which the total momentum is zero. That such a frame always is exists is not completely obvious. We need some preliminary results.[38]

Lemma If P_1 and P_2 are timelike or null, and future-pointing[39], then $P_1 \cdot P_2 \geq 0$.

Proof If P_1 and P_2 are both null, then they can be written in the form $P_1 = (p_1, \mathbf{p}_1)$ and $P_1 = (p_2, \mathbf{p}_2)$ with $p_i = |\mathbf{p}_i| \geq 0$. Then

$$P_1 \cdot P_2 = p_1p_2 - \mathbf{p}_1 \cdot \mathbf{p}_2 = p_1p_2(1 - \cos\theta) \geq 0,$$

where θ is the angle between the two three-momenta.

If P_1 is timelike, we work in the frame in which $P_1 = (E_1/c, \mathbf{0})$. (We can always think of a timelike 4-vector as the 4-momentum of a particle and this frame is the rest frame of the particle.) Then setting $P_2 = (E_2/c, \mathbf{p}_2)$, we have

$$P_1 \cdot P_2 = E_1E_2/c^2 \geq 0.$$

Lemma If P_1 and P_2 are timelike or null and future pointing, then $P_1 + P_2$ is also timelike or null and future pointing.

Proof To demonstrate this, we need to show that $(P_1 + P_2) \cdot (P_1 + P_2) \geq 0$. But

$$\begin{aligned}(P_1 + P_2) \cdot (P_1 + P_2) &= P_1 \cdot P_1 + P_2 \cdot P_2 + 2P_1 \cdot P_2 \\ &\geq 2P_1 \cdot P_2 && \text{(since } P_1 \text{ and } P_2 \text{ are timelike or null)} \\ &\geq 0. && \text{(by the above lemma)}\end{aligned}$$

[37] Here it is for completeness. We construct the second rank (antisymmetric) 4-tensor from the electric field 3-vector **E** and magnetic field 3-vector **B** as follows:

$$\mathcal{F} = \begin{pmatrix} 0 & -E_x/c & -E_y/c & -E_z/c \\ E_x/c & 0 & -B_z & B_y \\ E_y/c & B_z & 0 & -B_x \\ E_z/c & -B_y & B_x & 0 \end{pmatrix}$$

It can be checked that this transforms as a 4-tensor should under Lorentz transformations (but you have to know how **E** and **B** transform). The Lorentz 4-force on a particle with charge q is then

$$q\mathcal{F}U \equiv q\gamma \begin{matrix} \mathbf{u} \cdot \mathbf{E} \\ \mathbf{E} + \mathbf{u} \times \mathbf{B} \end{matrix},$$

where U is the 4-velocity.

[38] Note to self: the algebra of 4-vectors with the Minkowski inner product could form a separate expanded section. If P_1 and P_2 are spacelike or null, is it the case that $P_1 + P_2$ is spacelike or null?? (There are obvious counterexamples.) This could be tied in with the action of the (proper) Lorentz group on 4-vectors.

[39] This means that the first component of each vector is positive

By the above lemmas, the sum of two, and hence any number of, 4-momenta can always be thought of as a 4-momentum (it is timelike and future pointing). Therefore, we can always choose a frame in which the total momentum is zero by transforming to the rest frame of the total momentum.

Another useful result is the following:

Lemma If P_1 and P_2 are 4-momenta of particles with masses m_1 and m_2 respectively (where either might be massless), then $P_1 \cdot P_2 \geq m_1 m_2 c^2$.

Proof If both particles are massless, the result is trivial because the right hand side is zero. Otherwise, we work in the rest frame of one particle:

$$P_1 \cdot P_2 = (m_1 c, \mathbf{0}) \cdot (E_2/c, \mathbf{p}_2) = m_1 E_2 \geq m_1 m_2 c^2,$$

since $E_2 = \sqrt{(\mathbf{p}_1 \cdot \mathbf{p}_2 c^2 + m_2^2 c^4} \geq m_2 c^2$.

6.9.3 Examples

(i) Decay of a single particle (radioactive decay)

We consider the case of a particle of rest mass m_1 decaying into two particles of rest masses m_2 and m_3. Conservation of 4-momentum gives (in the obvious notation)

$$P_1 = P_2 + P_3\,.$$

Thus

$$E_1 = E_2 + E_3 \qquad \text{and} \qquad \mathbf{p}_1 = \mathbf{p}_2 + \mathbf{p}_3$$

For any particle, the formula $E^2 = p^2c^2 + m^2c^4$ holds so

$$E_1 = \sqrt{p_2^2 c^2 + m_2^2 c^4} + \sqrt{p_3^2 c^2 + m_3^2 c^4} \geq m_2 c^2 + m_3 c^2.$$

In particular, considering this result in the rest frame of the decaying particle (where $p_1 = 0$ and $E_1 = m_1 c^2$) gives $m_1 \geq m_2 + m_3$, so (perhaps unsurprisingly) the particle can only decay if its mass is at least as great as the sum of the rest masses of the product particles.

(ii) Decay of a massless particle

We consider the case of a massless particle decaying into two particles of masses m_2 and m_3 (either or both of which might be zero) . Again we have

$$P_1 = P_2 + P_3.$$

Taking lengths squared of both sides, and remembering that $P \cdot P = 0$ for a massless particle, we have

$$0 = m_2^2 c^2 + m_3^2 c^2 + 2 P_2 \cdot P_3$$

Now $P_2 \cdot P_3 \geq 0$ (see the lemma in the previous section). Since all the terms in this equation are non-negative, they must all vanish. This shows that a massless particle can only decay into massless particles and, if it does decay to two massless particles, their 3-momenta must be parallel.

(iii) Particle creation

If a proton[40] collides at sufficiently high energy with another proton, a proton anti-proton pair can be created, in addition to the original protons:[41]

$$p + p \to p + p + (p + \bar{p}).$$

We define the kinetic energy of a particle to be its energy minus its rest energy, i.e. $m\gamma c^2 - mc^2$; this is of course frame dependent. One might have thought that the minimum kinetic energy necessary to produce the extra two protons would be just $2m_p c^2$, but it turns out that much more is required. In fact, we shall see that the minimum energy required is $6m_p c^2$.

By conservation of 4-momentum we have

$$P = Q_1 + Q_2 + Q_3 + Q_4$$

[40] Protons are denoted by p, anti-protons by $\bar{p}$; protons and anti-protons have the same rest mass m_p.

[41] The first production of anti-protons (on Earth) was achieved by Chamberlain and Segre at the Berkley Bevatron in California; they received the Nobel prize for this in 1959.

where P is the total 4-momentum before collision (the sum of the individual 4-momenta of the two incident protons) and Q_i are the 4-momenta after the collision of the two original protons and the created proton and anti-proton. If we work in the rest frame of one of the two incident protons, so that $P = (m_p c, 0) + (E/c, \mathbf{p})$, we see that

$$P \cdot P = m_p^2 c^2 + m_p^2 c^2 + 2(m_p c)(E/c) = 2m_p^2 c^2 + 2m_p E.$$

We have

$$P \cdot P = (Q_1 + Q_2 + Q_3 + Q_4)(Q_1 + Q_2 + Q_3 + Q_4)$$

and, because of the lemma in the previous section, the magnitude of each of the 16 terms in the expansion of the right hand side of this equation is at least $m_p^2 c^2$. Thus

$$2m_p^2 c^2 + 2m_p E \geq 16 m_p^2 c^2 \quad \text{i.e.} \quad E - m_p c^2 \geq 6 m_p c^2$$

which is the minimum kinetic energy required to form the proton anti-proton pair. This reflects the fact that there must be enough incident energy (at least, in the case considered here of an incident proton colliding with a stationary proton) to account for the recoil kinetic energy of all the particles as well as the rest masses of the created particles. Of course, there is no Newtonian analogue or Newtonian limit for this particular example, because in this example rest mass is not conserved. [42]

(iv) The same as (iii) but using the centre of momentum frame

In the rest frame of one of the incident protons, the total momentum of the two incident protons is $(m_p c, 0) + (E/c, \mathbf{p}) = (m_p c + E/c, \mathbf{p})$ as above.

Suppose now that we transform (without actually doing the transformation) to the centre of momentum frame, in which the total momentum is zero. In this frame, the total momentum of the incident protons is $(E'/c, 0)$ where E' is the (not yet calculated) total relativistic energy. Thus the total relativistic energy of the four protons after the collision, $E_1 + E_2 + E_3 + E_4$ satisfies

$$E' = E_1 + E_2 + E_3 + E_4 \geq 4 m_p c^2 \,.$$

Furthermore, this bound cannot be improved because the inequality becomes an equality if all four protons have zero 3-momentum.

To find out what this inequality says about E, we have to transform back into the original frame. Let us choose axes to that $\mathbf{p}$ is parallel to the x-axis. Then, using our standard Lorentz transformation,

$$\begin{pmatrix} E'/c \\ 0 \end{pmatrix} = \begin{pmatrix} \gamma & \gamma v/c \\ \gamma v/c & \gamma \end{pmatrix} \begin{pmatrix} m_p c + E/c \\ p \end{pmatrix}$$

The second row of this matrix equation tells us the velocity of the transformation and the first row will tell us, using the velocity, we have just found, what E' is in terms of $E + m_p c^2$ and p. But hold on! Do we actually have to do this rather messy tranformation? Could we not just use the scalars?

Indeed we can. The scalar magnitude of the 4-momentum is of course the same in both (all) frames, so

$$E'^2/c^2 = (m_p c + E/c)^2 - p^2$$

and

$$(m_p c + E/c)^2 - p^2 \geq 16 m_p^2 c^2.$$

Now expanding the square on the left hand side, and remembering that $E^2 = p^2 c^2 + m_p^2 c^4$, we find that

$$2m_p E \geq 14 m_p^2 c^2,$$

as before.

But hold on! Surely if we are using 4-vectors, we should be able to avoid using results such as $E^2 = p^2 c^2 + m_p^2 c^4$, because this can be derived using the invariance of $P \cdot P$? And indeed we can. Writing the incoming proton momenta as P_1 and P_2, we have for the total momentum before the collision

$$(P_1 + P_2) \cdot (P_1 + P_2) = P_1 \cdot P_1 + P_2 \cdot P_2 + 2P_1 \cdot P_2 = 2m_p^2 c^2 + 2P_1 \cdot P_2.$$

In the original frame $P_1 = (m_p c, \mathbf{0})$ and $P_2 = (E/c, \mathbf{p})$, so

$$2m_p^2 + 2Em_p = (Q_1 + Q_2 + Q_3 + Q_4)(Q_1 + Q_2 + Q_3 + Q_4) \geq 16 m_p^2 c^2,$$

giving the same result as before.

By working in the centre of momentum, we have established this result rather easily, and we have also proved that no tighter bound is possible.

[42] Note to self: this example could be extended by repeating it in the centre of momentum frame and relating the results by a Lorentz transformation.

(v) Pion decay

We consider the decay of the neutral pion[43] into two photons:

$$\pi^0 \to \gamma + \gamma$$

Suppose the energy of the pion (in some given frame) is E_π and the energy of one of the photons (in the same frame) E_γ. What is the angle, θ, between the path of this photon and the path of the pion?

By conservation of 4-momentum we have, in the obvious notation,

$$P_\pi = P_\gamma + P'_\gamma$$

We know nothing about the 4-momentum P'_γ of the second photon, except that it is a null (light-like) vector and therefore has zero length. Nor are we asked to find out anything about it. We therefore eliminate it by taking lengths. We have

$$P'_\gamma = P_\pi - P_\gamma$$

so

$$\begin{aligned} 0 = \big(P_\pi - P_\gamma\big) \cdot \big(P_\pi - P_\gamma\big) &= P_\pi \cdot P_\pi + P_\gamma \cdot P_\gamma - 2P_\pi \cdot P_\gamma \\ &= m_\pi^2 c^2 + 0 - 2E_\pi E_\gamma / c^2 + 2\mathbf{p}_\pi \cdot \mathbf{p}_\gamma \\ &= m_\pi^2 c^2 - 2E_\pi E_\gamma / c^2 + 2\left(\sqrt{E_\pi^2/c^2 - m_\pi^2 c^2}\right) E_\gamma \cos\theta \end{aligned}$$

(recall that for a photon $E = pc$). In this last equation, everything except θ is known, so θ can be found in terms of the known quantities.

[43]There is a triplet of pions with positive, negative and zero charge ($\pi^\pm$ and π^0). They are mesons, i.e. middle weight particles compared with hadrons (heavy particles such as neutrons and protons) and leptons (light particles such as electrons). The neutral pion has a short life-time: it decays with a half life of 10^{-17} seconds, mostly by the process described in this example.