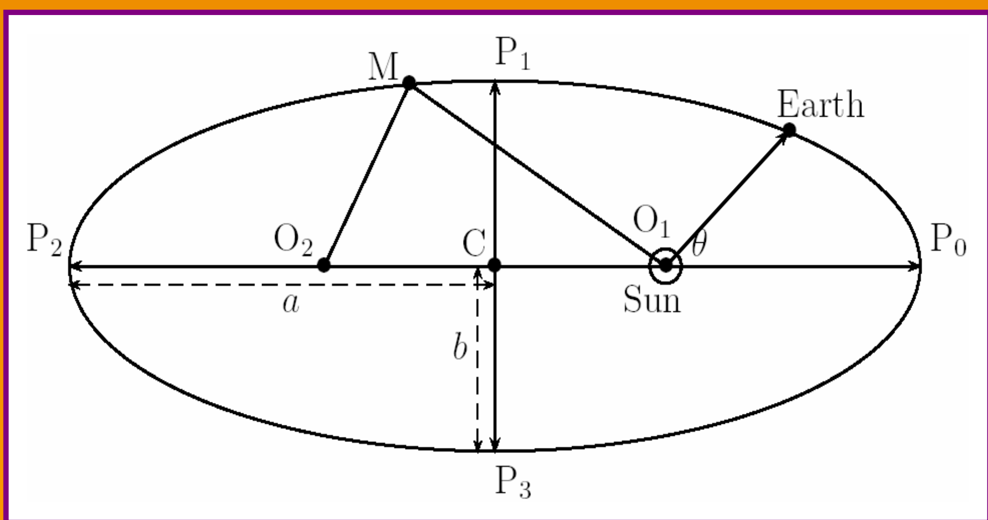
Basic Books in Science

Book 4

Motion and Mass: First Steps into Physics



Roy McWeeny

BASIC BOOKS IN SCIENCE

– a Series of books that start *at the beginning*

Book 4

Mass and motion
– first steps into Physics

Roy McWeeny
Professore Emerito di Chimica Teorica, Università di Pisa, Pisa (Italy)

The Series is maintained, with regular updating and improvement, at
http://www.learndev.org/ScienceWorkBooks.html
and the books may be downloaded entirely free of charge

(Last updated 10 November 2011)

BASIC BOOKS IN SCIENCE

Acknowledgements

In a world increasingly driven by information technology no educational experiment can hope to make a significant impact without effective bridges to the 'user community' – the students and their teachers.

In the case of "Basic Books in Science" (for brevity, "the Series"), these bridges have been provided as a result of the enthusiasm and good will of Dr. David Peat (The Pari Center for New Learning), who first offered to host the Series on his website, and of Dr. Jan Visser (The Learning Development Institute), who set up a parallel channel for further development of the project. The credit for setting up and maintaining the bridgeheads, and for promoting the project in general, must go entirely to them.

Education is a global enterprise with no boundaries and, as such, is sure to meet linguistic difficulties: these will be reduced by providing translations into some of the world's most widely used languages. Dr. Angel S. Sanz (Madrid) is preparing Spanish versions of the books and his initiative is most warmly appreciated.

We appreciate the interest shown by universities in Sub-Saharan Africa (e.g. University of the Western Cape and Kenyatta University), where trainee teachers are making use of the Series; and that shown by the Illinois Mathematics and Science Academy (IMSA) where material from the Series is being used in teaching groups of refugee children from many parts of the world.

All who have contributed to the Series in any way are warmly thanked: they have given freely of their time and energy 'for the love of Science'.

Pisa, 2007 Roy McWeeny (Series Editor)

BASIC BOOKS IN SCIENCE

About this Series

All human progress depends on **education**: to get it we need books and schools. Science Education is of key importance.

Unfortunately, books and schools are not always easy to find. But nowadays all the world's knowledge should be freely available to everyone – through the Internet that connects all the world's computers.

The aim of the Series is to bring basic knowledge in all areas of science within the reach of everyone. Every Book will cover in some depth a clearly defined area, starting from the very beginning and leading up to university level: all will be available on the Internet *at no cost to the reader*. To obtain a copy it should be enough to make a single visit to any library or public office with a personal computer and a telephone line. Each book will serve as one of the 'building blocks' out of which Science is built; and together they will form a 'give-away' science library.

About this book

This book, like the others in the Series, is written in simple English – the language most widely used in science and technology. It builds on the foundations laid in Book 1 (Number and symbols), Book 2 (Space) and Book 3 (Relationships, change – and Mathematical Analysis). Book 4 starts from our first ideas about the world around us: when we push things they usually move and the way they move depends on how 'heavy' or 'massive' they are.

From these simple ideas about mass and motion, and a few experiments that anyone can do, we can lay the foundations of Physics: they are expressed mathematically in the 'laws of motion', which form the starting point for the Physical Sciences. Almost all of Physics and its applications, up to the end of the 19th century, can be understood using only the laws of motion! The rest involves Electricity (to be studied in Book 10) and takes us into Modern Physics and all that has happened during the last 150 years. So we're starting on a very long journey of discovery ....

Looking ahead –

Now you know something about numbers and symbols (Book 1) and about space and geometry (Book 2); and you've learnt (in Book 3) how to use these ideas to study relationships between measurable quantities (how one thing depends on amother). So at last you're ready to start on Physics.

Physics is a big subject and you'll need more than one book; but even with only Book 4 you'll begin to understand a lot about a large part of the world – the *physical* world of the *objects* around us. Again, there are many important 'milestones'....

- Chapter 1 deals with actions like pushing and pulling, which you can *feel* with your body: they are **forces**, which can *act* on an object to make it move, or to change the way it is moving. But every object has a **mass**, which measures how much it *resists* change. By the end of the chapter, you'll know about force, mass, weight and gravity; and the famous laws put forward by Newton. You'll know that forces are *vectors* (which you first met in Book 1) and how they can be *combined.*

- In Chapter 2 you'll think about lifting things: you do **work** and get tired – you lose **energy**. Where has the energy gone? You find two kinds of energy: **potential energy**, which you can *store* in an object; and **kinetic energy**, which is due to its *motion.* The sum of the two is constant: this is the principle of **energy conservation**.

- Chapter 3 extends this principle to the motion of a **particle** (a 'point mass') when it is acted on by a force and moves along and curved path. Energy is still conserved. You learn how to calculate the path; and find that what's good for a small particle seems to be good also for big ones (e.g. the Earth going around the Sun).

- Chapter 4 asks why this can be so and finds a reason: think of a big body as a collection of millions of particles, all **interacting** with each other, and use Newton's laws. One point in the body, the **centre of mass**, moves as if all the mass were concentrated at that point and acted on by a *single force* – the vector sum of all the actual forces applied to the big body. You'll also learn about **momentum** and **collisions**.

- Time to think about **rotational motion**. Chapter 5 shows how to deal with rotation of a many-particle system about its centre of mass. Corresponding to "Force = rate of change of (linear) momentum", there is a new law "**Torque** = rate of change of **angular momentum**". You'll learn how torque and angular momentum are defined; and how the new law applies just as well to both one-particle and many-particle systems. So you can study the Solar System in more detail and calculate the orbits of the planets.

- In Chapter 6 you'll be thinking of a **rigid body**, in which all the particles are joined together so that the distances between them can't change. If the body is moving you are studying **Dynamics**; but if it is in equilibrium (at rest or in *uniform* motion), then you are studying **Statics**. You'll be able to solve many every-day problems and you'll be well prepared for entering the Engineering Sciences.

- Chapter 7 deals with simple machines, illustrating the principles from earlier Chapters, going from levers to water-wheels and clocks.

- The final Chapter 8 carries you to the present day and to the big problems of the future; we all depend on *energy* – for machines and factories, for transporting goods (and people), for digging and building, for almost everything we do. Most of that energy comes from burning fuel (wood, coal, oil, gas, or anything that will burn); but what will happen when we've used it all? We probably need to solve that problem before the end of this century: how can we do it? Do we go back to water-mills and wind-mills, or to the energy we can get from the heat of the sun? In this last chapter you'll find that **mass is a form of energy** and that *in theory* a bottle of seawater, for example, could give enough energy to run a big city for a week! – if only we could get the energy out! This is the promise of **nuclear energy**. Some countries are using it already; but it can be dangerous and it brings new problems. To understand them you'll have to go beyond Book 4. In Book 5 you'll take the first steps into Chemistry, learning something about atoms and molecules and what everything is made of – and inside the atom you'll find the nucleus!

CONTENTS

Notes to the Reader. When Chapters have several Sections they are numbered so that "Section 2.3" will mean "Chapter 2, Section 3". Similarly, "equation (2.3)" will mean "Chapter 2, equation 3".

Important 'key' words are printed in **boldface**, when they first appear. They are collected in the Index at the end of the book, along with page numbers for finding them.

Chapter 1

Mass, force, and weight

1.1 What makes things move?

When you take hold of something (we'll usually call it an 'object' or a 'body') and *pull* it, towards you, or *push* it, away from you, it usually moves – unless it's fixed or too big.

In both cases you 'apply a **force**' to the object. Push and pull are both forces, which can be big or small (depending on how strong you are and on how 'hard' you push or pull). So a force has a *magnitude* (size) and a *direction*: it is a **vector** (see Book 1, Section 3.2) and is often represented by an arrow, pointing one way or another (away from you if you are pushing, towards you if you are pulling); and the length of the arrow is used to show the magnitude of the force it represents (long for a big force, short for a small one).

When you apply a force to a body it will also have a 'point of application', usually *you* – where you take hold of the body – and this can be shown by putting the end of the arrow (not the sharp end!) at that point.

We're now all set to go. Fig.1 represents the force you might apply to a cart full of stones to make it move away from you – in the direction of the arrow. The bold dot marks the point of application.

At first the cart is standing still; it is 'at rest' or 'stationary'. But then, as you keep on *pushing*. it begins to move – very slowly at first, but then faster and faster, until it seems to be going by itself! Even when you stop pushing it keeps on going – until something stops it (perhaps the ground is rough and a wheel gets stuck in a hole).

Another example could be a barge (a flat-bottomed boat for carrying heavy loads), usually *pulled* (or 'towed') by a horse or other strong animal, walking along the 'tow path' at the side of the river or canal (Fig.2). Starting from rest, the barge moves very slowly at first, even when the horse is pulling as hard as it can. But then it goes faster and faster, even when the animal is just walking and seems to be pulling hardly at all – the barge is moving almost by itself.

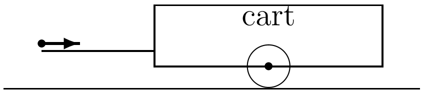


Figure 1

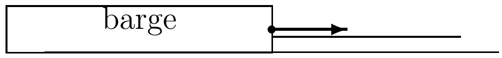


Figure 2

This second example brings in another idea. The ‘pulling’ force in Fig.2 is applied at the dot (•) by means of a rope, connecting the animal to the barge, and the rope is stretched *tight*. We say that there is a **tension** in the rope and that this tension carries or ‘transmits’ the pull from one end of the rope (the animal) to the other end (the barge). The tension is just a special kind of force but it has the important property of being able to carry force from one point to another.

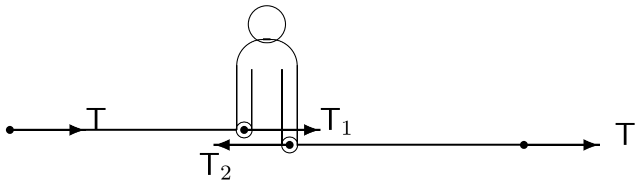


Figure 3

If you cut the rope at any point (P, say) you can keep everything just as it was by holding the two cut ends together (if you’re strong enough!) as in Fig.3. To do that you have to pull the left-hand piece of rope with a force T_1 (*equal* to T, the one at first applied by the horse) and the right-hand piece with a force T_2 just as big as T_1 but pointing in the opposite direction. The forces must be equal, because otherwise you’d be pulled off your feet! The forces you are applying are now $T_1 = T$ (you’re now standing in for the horse!) and $T_2 = -T_1 = -T$, where the minus sign just shows the *direction* of the force (let’s agree that negative means to the left, positive to the right).

Of course the animal is still there, pulling with the same force T on the far end of the right-hand rope, so we show it as the last force vector on the right.

The force applied to the barge (T_1) is called the **action**, while the equal but *opposite* force T_2, is called the **reaction** of the barge against whatever is pulling it.

Since the point P could be *anywhere* in the rope, it is clear that the tension must be the same at all points in the rope and that it can be represented by a pair of arrows in opposite directions $\leftarrow \bullet \rightarrow$. The equal and opposite forces at any point are what keeps the rope tight; and the tension is simply the magnitude of either force.

We can talk about a ‘pushing’ force in a similar way. But you can’t transmit a push with a length of rope or string; it just folds up! You need something *stiff*, like a stick. When you push something with a stick you can describe it using a diagram similar to the one in Fig.3 except that the directions of all the forces are reversed. Just imagine cutting a little piece out of the stick at point P (you can do it in your head – you don’t need a saw!). Then you have the two halves of the stick, separated by the bit you’re thinking of taking away (this takes your place in Fig.3, where you were holding the two pieces of rope together). If you're pushing something on your left (perhaps trying to *stop* the barge that was still moving) then the forces in the stick can be pictured as in Fig.4a (magnified so you can see what's going on): the first bit of stick is pushing the barge with a force $-F$ (i.e. F in the negative direction – to the left) and so feels a reaction $+F$; each piece of stick *pushes* the next one with a force F and is *pushed back* by a force $-F$.

(a) compression (b) tension

Figure 4

Whenever there is a pair of equal but opposite *pushes* at any point in the stick we say it is in **compression**; and it is now the *compression* that transmits the force from one end of the stick to the other. Notice that in Fig.4(a) we've shown the forces acting *on the piece of stick*: they are applied to the 'faces' (i.e. the cut ends) of the piece and point *into* it. To be clear about the difference between compression and tension, look at Fig.4(b) – which shows a little piece of the rope (Fig.3) when it's in tension: the forces acting on this piece of rope are applied at its cut ends and point *out of it*. When you were standing in for that missing piece of rope (Fig.3) these were the forces you could actually feel, as if they were trying to pull you over: they were the *reactions* to the forces you were applying to the barge (on the left) and the horse (on the right).

Tension and compression are two very important ideas, that we'll use a lot.

We now want to bring all these ideas, taken from everyday life, together and to express them in a few simple principles or 'laws' – the laws first proposed by Newton in his 'Principia', a book published in 1686.

1.2 How can we measure a force?

In Physics we need to *measure* all the things we talk about. To measure force we start by looking for a law, in the form of an equation, to express what we've found from everyday experience. Force, let's use the letter F to stand for its magnitude, is what makes a body move; but it does not move *as soon as the force is applied* – you have to *wait* until it starts to move, slowly at first and then faster and faster. So force, applied to a body at rest (not fixed but free to move), will lead to an *increase* in its speed from zero to some value v, where v is the magnitude of the **velocity** (another vector quantity like the force itself). Since v increases as time passes, provided we go on applying the force, we can say v is a 'function of the time t' (see Book 3 Chapter 1) and write $v = v(t)$. The *rate* at which v increases (how much extra speed the body gains in every second) is called the

acceleration of the body, usually denoted by a, and a will also be a function of time: $a = a(t)$. Notice, however, that even when the *magnitude* v of the velocity vector v is not increasing the vector may be changing *direction*, so more correctly acceleration means the rate of *change* of the vector v. (If you swing something round your head, on the end of a string, its speed may be constant but the velocity is continually changing direction. The vector v is constant only if it keeps always the same direction; but you can feel the tension in the string – and that is what pulls it to one side and makes it go in a circle instead.)

Everything we do seems to tell us that "the bigger the force applied to a body, the bigger will be its acceleration": to double the acceleration of the cart, or the barge, we need to double the force applied (e.g. by having two people, instead of one, doing the pushing or the pulling). But the result will also depend on the object being pushed or pulled: a cart full of stones needs a much harder push than one that's empty!

We can now put all we know about moving things into one very simple equation. In words it will say

> The force F needed to make an object move with an acceleration a, is **proportional** to the value of a. The proportionality constant is called the **mass** (m) of the object.

or in symbols

$$F \propto a \quad \text{or} \quad F = ma. \tag{1.1}$$

More fully the quantity m is called the *inertial* mass, being a measure of the 'inertia' or slowness of the object to change its state of motion when acted on by a force.

Equation (1.1) is usually called 'Newton's second law of motion'. It really includes the law he stated first – that any body continues in its state of rest, or uniform motion in a straight line, unless acted on by some 'external agency' (i.e. a force). To see why this is so, we take away the force by putting $F = 0$; from (1.1) this means the acceleration (a) must be zero; and this means the velocity of the body (v) is a *constant* – which includes the specialcase ($v = 0$) when the body is at rest. So a body may be going very fast, even when no force is applied to it! And we've seen examples of this with the cart and the barge – when you stop pushing or pulling, the thing still keeps on going 'by itself', until something stops it. Notice again that 'uniform motion' must be in a straight line, because velocity is really a *vector* and if the motion is in a curve then the direction of the vector v is changing; in other words, the acceleration will *not* be zero! Notice also that v (in special type), is used from now on to indicate the velocity *vector*, which is not the same as its *magnitude* v (shown in ordinary *italic* type).

We've already discovered Newton's third law of motion when we were thinking about tension and compression in the last Section: when two forces act at the same point they must be equal in magnitude but opposite in direction – "action and reaction are equal but opposite". Newton's great idea, however, was much more general than that: he realized that action and reaction together make up the **interaction** between two things. Whenever A acts on B with a force F (the 'action'), B acts on A with a force $-\mathsf{F}$ (the 'reaction'): if A and B interact in any way whatever, then one feels a force F and the other feels a force $-\mathsf{F}$ – and it doesn't matter which one we call the action and which we call the reaction. So Newton's third law of motion can be put as

> "To every action there must be an equal and opposite reaction."

In other words, *there is no such thing as a single force!*

With Newton's three laws we are almost ready to do marvellous things. But not quite – because we still don't know how to measure the force and the mass! So how can we actually use equation (1.1)? To answer this question we need to know something about **gravity**, a word coming from the Latin *gravis*, meaning 'heavy'. (Like many of the great philosophers of the time, Newton wrote in Latin and when he needed a new word that's where he took it from.)

When we let go of an object it falls to the ground, even if we aren't applying any force to it; and it falls with a certain constant acceleration of $a = 0.981\text{m s}^{-2}$, as was noted in Section 1.1 of Book 3. So there must be a force acting on it, even if we aren't even touching it. Where does that mysterious force come from? It is called the 'force due to gravity' and it arises because any two masses *attract each other* even when they aren't connected in any visible way. This doesn't really *explain* what gravity is; and even Einstein's theory of 'general relativity' doesn't tell the whole story. But the general idea is simple: any massive body, like the Earth itself, has a small effect on the space around it; it 'bends' the space very slightly and this bending shows itself as a *field of force*, which is 'felt' by any other mass which enters the field. This has been mentioned already, in the last Chapter of Book 2; but hundreds of years before Einstein, Newton had already proposed, on the basis of observation, what turned out to be the correct (very nearly) *law of universal gravitation* – 'universal' because it seemed possible that it was true for all the stars and planets in the Universe! And the law is again surprisingly simple: in words it says that there is a force of attraction F between any two masses m and M, proportional to the product of the masses and *inversely* proportional to the *square* of the distance, R say, between them. Written as an equation this law becomes

$$F = G\frac{mM}{R^2}, \tag{1.2}$$

where G, the 'constant of gravitation' is a proportionality constant which can only be found by experiment.

The force F in (1.2) is our first example of an interaction between two bodies that doesn't depend on their being in *contact*, or being connected by strings or held apart by sticks. And yet Newton's third law applies: if the force on mass m, produced by mass M, is represented by the vector F pointing from m towards M, then the force on M, produced by m, is represented by the vector $-\mathsf{F}$ pointing in the opposite direction. It may seem strange that an apple falls to the ground (i.e. towards the centre of the Earth) as soon as we let it go, while the Earth doesn't seem to move towards the apple – when both feel the same force of attraction. But that's because the mass of the Earth is many millions of times that of the apple, so according to (1.1) its acceleration towards the apple would be almost zero – even if we consider only the apple and the Earth and not all the other 'small' things (from birds to battleships!) that feel the Earth's gravitational pull and all attract each other in different directions.

We still don't know how to measure force and mass! So where do we go from here?

1.3 Force, mass, and weight

The 'laws' represented in equations (1.1) and (1.2) are not really *laws* at all. They should be called **hypotheses** – proposals, based on everyday experience and guesswork but not proved: they can't be accepted as laws until they've been thoroughly tested and found to be true. But we're doing Physics, and that's the way it goes: we look around us and experiment, we measure things and guess how they may be related, and then we make proposals which can be tried and tested; if they don't work we throw them away and start again; but if they *do* then we accept them and go ahead, taking them as the 'laws of Physics'.

Now that we have a few laws, we can come back to the problem of how to measure forces and masses. Let's use the letter g for the 'acceleration due to gravity' ($g = 0.981$m s^{-2}), the rate at which the speed of a freely falling body increases – nearly 1 metre a second in every second! The most remarkable thing about g is that it *really is a constant*: it's the same for all bodies, pebbles or plastic, people or paper, as long as they're really *free* to fall. You might feel sure that heavier things fall faster, thinking of stones and feathers, but really things falling through the air are not quite free - the air slows them down a bit and this 'bit' is very important for a feather, which just floats slowly to the ground, but not for a heavy stone that just pushes the air out of its way. If we take away the air (doing the experiment in a big glass jar, after sucking out all the air with a pump) we find that everything *does* fall just as fast: g has the same value for all falling bodies. It was the Italian, Galileo Galilei (1564 -1642), who first found this was so and became one of the first people in history to use the 'scientific method' – observing, measuring, proposing a law, and testing it. Before that, people just followed what other people had said. And the famous Greek philosopher Aristotle (384-322 BC) had said that heavier things fell faster, so everyone believed him, for nearly 2000 years, just because he was so famous!

Why is all this so important? If we have two masses (m_1 and m_2) and let them fall, in the gravitational field of the Earth, the second law tells us that the forces acting on the two masses must be $F_1 = m_1 g$ and $F_2 = m_2 g$, respectively. We call the force $F = mg$ associated with any mass m its **weight**. So if $m = 1$ kg is the *standard mass unit*, the 'kilogram', kept in Paris (see Section 1.1 of Book 1), then $w = mg$ will be 1 'kilogram *weight*' or, for short, 1 kg wt. Mass and weight are two different things: weight is the *force* that acts on a mass, because of the gravitational field. And if you take your kilogram to the Moon it will not weigh as much because the Moon's gravitational field is less strong, M in equation (1.2) being much smaller for the Moon than M for the Earth. Perhaps you've seen pictures of astronauts on the Moon jumping to great heights because they don't weigh as much; and this shows that g is much smaller for bodies on the Moon.

All the same, we live on the Earth and the easiest way for us to compare and measure masses is through their weights. This is possible, as we can now see, because the *ratio* of two masses is just the same as the ratio of their weights:

$$\frac{w_2}{w_1} = \frac{m_2 g}{m_1 g} = \frac{m_2}{m_1} \tag{1.3}$$

– the gs cancelling. So the use of a weighing machine in Book 1 as a way of measuring masses (i.e. comparing any mass m with a given unit mass) gives the right answer, wherever you do the experiment – even on the Moon, where g is not the same as it is here – provided you use the weighing machine to *compare* the mass with a standard mass and don't just use the marks on a scale. (Can you say why? - remembering that the 'spring balance' in Book 1 works by stretching a spring.)

Now we have a fairly complete system of units (for measuring masses, lengths, and times, and all quantities depending only on M,L,T) as long as we don't meet electric charges. The units of length and time are the metre (m) and the second (s), respectively. The unit of mass is the kilogram (kg). The unit of force is called the **Newton** (N): it is defined as the force which will give unit mass (1 kg) an acceleration of 1 m s^{-2}. In other words,

$$1 \text{ N} = (1 \text{ kg})\times(1 \text{ m s}^{-2}) = 1 \text{ kg m s}^{-2},$$

which means [force] = MLT^{-2} – "force has the dimensions MLT^{-2}". Since the force acting on a body of mass m is mg, and is the *weight* of the body, we can say

$$1 \text{ kg wt} = 1 \text{ kg}\times(9.81 \text{ m s}^{-2} = 9.81 \text{ kg m s}^{-2} = 9.81 \text{ N}.$$

To convert a force of x kg wt into Newtons we just have to multiply by 9.81, obtaining $9.81x$ N.

1.4 Combining forces

In Section 1 we noted that forces were *vectors* and that when two forces act at a point (e.g. any point in a stretched rope or string) they must be equal and opposite (equal in magnitude but opposite in direction): they are called action and reaction; and the object on which they act (e.g. a little bit of string at point P) doesn't move because the *combined* action of the forces is zero – they are 'in balance' and have a 'resultant' which is the same as no force at all.

Figure 5

Forces are combined, or added, using the law of vector addition: you represent each force vector by an arrow and put the arrows 'head to tail', *without changing their directions* (i.e. by sliding them but *not* turning them). An arrow pointing from the first 'tail' to the last 'head' will then represent the **vector sum** of the forces. We'll find that this is a general rule, however many forces there may be, but for the moment let's talk about just two. Fig.5(a) shows two forces acting on something at point P, while Fig.5(b) shows how they are combined to give a **resultant** which is zero – represented by an arrow with zero length (just a point). When there is no resultant force acting on a body it does not move and we say it is in **equilibrium**.

(You may think all this can't be true: if point P is on the rope that's pulling the barge it will be *moving* so how can it be in equilibrium? But remember that Newton's first law talks about a 'state of rest *or uniform motion in a straight line*' – so there's really no contradiction.)

But what happens if more than two forces act at the same point? They may

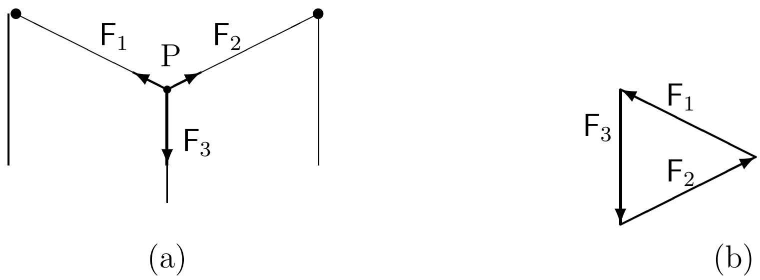


Figure 6

even have different sizes or different directions, so how do we find their resultant?

To have a real-life example, let's suppose the forces F_1, F_2 and F_3 are the *tensions* in three strings tied to a small bead at point P. The tensions are produced by two equal 2 kilogram weights ($w_1 = w_2 = 2$ kg wt), hanging from strings as in Fig.6(a). The strings pass over smooth nails, hammered into a vertical board, so the force $F_1 = w_1$ is transmitted to the bead – the tension being the same at all points on the string – and so is F_2. The third force, $F_3 = w_3$, is the tension in the vertical string that supports the weight w_3. (Notice that the nails are only needed, so as to change the vertical pull of the weights w_1, w_2 into a sideways pull on the bead.)

The directions of the three forces are represented by the by the thick arrows in Fig.6(a); and we suppose the bead has come to rest at point P, which is then its equilibrium position. This means that the resultant force on the bead, the vector sum $\mathsf{F}_1 + \mathsf{F}_2 + \mathsf{F}_3$, should be zero.

To form the vector sum of $\mathsf{F}_1, \mathsf{F}_2, \mathsf{F}_3$ and show that it is zero, we simply put their arrows head to tail (again by sliding them around *without changing their lengths and directions*, which means the vectors will not be changed in any way). The result is shown in Fig.6(b): the vectors now form a *closed triangle* and the fact that it is closed (no distance between the first 'tail' and the last 'head') means the vector sum is zero $\mathsf{F}_1 + \mathsf{F}_2 + \mathsf{F}_3 = \mathsf{0}$. But wait a minute! We didn't say how big the third weight was. If we took $w_3 = 3$ kg wt, for example, the sides would *not* form a closed triangle, the vector sum would not be zero, and the bead at P would not be in equilibrium. In fact, we took $w_3 = 1.789$ kg wt and this is the only value that makes the triangle close exactly.

So what have we done? Since the angles at the corners of a triangle, with sides of known lengths, are easy to calculate by simple geometry (see Book 2), we can actually *calculate* the angles between any three forces that meet at a point and are in equilibrium. Once we know how to work with vectors we don't always have to do experiments – we can do it all in advance, on paper!

1.5 How to work with vectors

To end this Chapter, let's remember the rules for dealing with vectors by introducing their **components**. When we work in three-dimensional space ('3-space', for short), rather than on a flat surface (a 2-space), we sometimes have to do complicated geometry. This can often be made easier by representing vectors in terms of their **components** along three **axes**, the x-axis, the y-axis, and the z-axis, as in Section 5.3 of Book 2. Usually, the axes are chosen **perpendicular** to each other (or **orthogonal**) and are defined by three **unit vectors** $\mathsf{e}_1, \mathsf{e}_2, \mathsf{e}_3$ pointing along the three axes, each with unit length. If you think of the vector v as an arrow, then its components v_1, v_2, v_3 along the three axes are the numbers of steps you have to take along the three directions to express v in the form

$$\mathsf{v} = v_1\mathsf{e}_1 + v_2\mathsf{e}_2 + v_3\mathsf{e}_3, \tag{1.4}$$

where $v_1\mathsf{e}_1$, for example, is the vector v_1 times as long as e_1 and the vectors are added using the usual 'arrow rule'. The order of the terms in the sum doesn't matter; and to add two vectors a, b we simply add corresponding components. So if c is the vector sum of a and b, then

$$\mathsf{c} = \mathsf{a} + \mathsf{b} \quad \leftrightarrow \quad c_1 = a_1 + b_1, \quad c_2 = a_2 + b_2, \quad c_3 = a_3 + b_3 \tag{1.5}$$

- the symbol $\leftrightarrow$ simply meaning that the things it separates are exactly equivalent, the single vector equation is equivalent to three ordinary equations among the numerical components. Sometimes you can go a long way with vector equations (for example, using $\mathsf{F}_1 + \mathsf{F}_2 + \mathsf{F}_3 = \mathsf{0}$ as a condition for three forces at a point to be in equilibrium): but in the end you'll need to get numbers (e.g. the magnitudes of forces and the angles between them) – and then you'll go to the components.

To find the component of a vector along some given axis, all you need do is think of it as an arrow starting at the origin and drop a perpendicular from the tip of the arrow onto the axis: the part of the axis that goes from the origin to the foot of the perpendicular is the **projection** of the arrow on the axis; and its length is the value of the component. In Section 3.2 of Book 2 we noted three important quantities, all relating to the angles in a triangle. If we call the angle θ ('theta'), then the sine, cosine, and tangent of the angle are (see Fig.7(a))

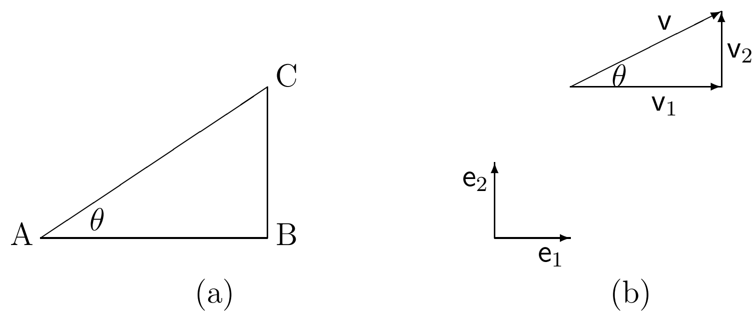


Figure 7

$$\sin\theta = BC/AC, \qquad \cos\theta = AB/AC, \qquad \tan\theta = BC/AB,$$

where AB, BC, AC are the lengths of the three sides. Fig.7(b) shows how the components of the vector v in (1.4) can be expressed in terms of the angles it makes with two unit vectors in the same plane:

$$v_1 = v\cos\theta, \qquad v_2 = v\sin\theta, \tag{1.6}$$

where v is the length of the vector (its magnitude or 'modulus').

Most of the time, we don't need anything else – not even Tables of sines and cosines for all angles – because we know that in the right-angled triangle Fig.7(a) $AC^2 = AB^2 + BC^2$ (Book 2, Chapter1), so given any two sides we can easily get the other side, and then all the ratios.

To see how it all works, let's go back to the example in Fig.6, but making it a bit more difficult: if all three weights are different, the bead is pulled over to one side – so how can we find the new 'equilibrium position'? – and how will this depend on what weights we use? Suppose we choose $w_1 = 2$ kg wt, as in Fig.6, but aren't sure what w_2 and w_3 must be to keep the 'lop-sided' arrangement in Fig.8(a) in equilibrium. How can we decide?

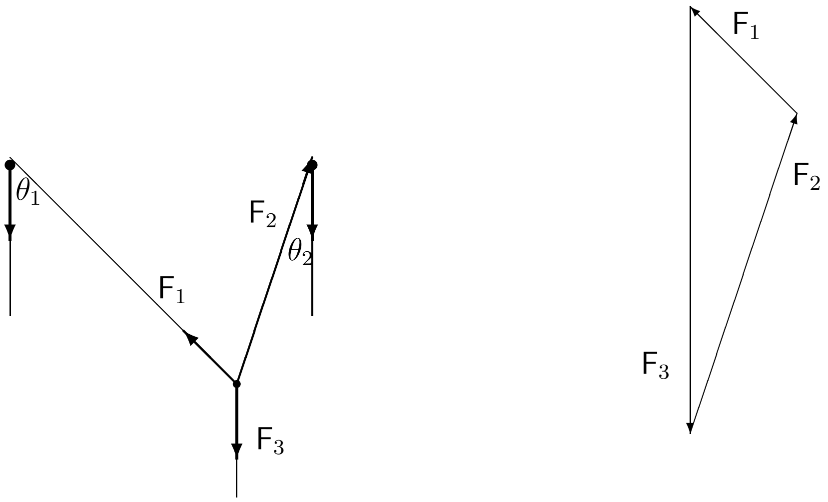
To be in equilibrium, the resultant force acting on the bead at P must be zero – for otherwise it would start moving. So let's resolve the vector sum $\mathsf{F}_1+\mathsf{F}_2+\mathsf{F}_3$ into horizontal and vertical components, as in Fig.7(b), and take each one separately. The force F_3 points vertically downwards and has no horizontal component; but F_1 and F_2 have horizontal components $-F_1\sin\theta_1$ and $F_2\sin\theta_2$, respectively, where F_1, F_2 are the magnitudes of the force vectors and θ_1, θ_2 are the angles shown in Fig.8(a). The horizontal component of the resultant force in the positive (right-hand) direction is thus

$$-F_1\sin\theta_1 + F_2\sin\theta_2 = -w_1\sin\theta_1 + w_2\sin\theta_2$$

– the forces (tensions) in the left-hand and right-hand strings being w_1 and w_2 in units of 1 kg wt. The angles in Fig.7(a) give (think of the triangles with horizontal and vertical sides of length 3,3 for θ_1 and 1,3 for θ_2)

$$\sin\theta_1 = 1/\sqrt{2}, \qquad \sin\theta_2 = 1/\sqrt{10}.$$

With $w_1 = 2$ kg wt we can only prevent the bead moving sideways by choosing w_2 kg wt, so that $2\times(1/\sqrt{2})$ kg wt $= w_2\times(1/\sqrt{10})$. And if you solve this equation (see Book 1) you'll find $w_2 = 4.472$ kg wt.



(a) (b)

Figure 8

For equilibrium we still have to choose w_3 so that the bead will not move up or down; and this means the total force must also have zero component in the *vertical* direction. The positive (upward) component of the forces in the two strings will be

$$F_1 \cos\theta_1 + F_2 \cos\theta_2 = 2 \times (1/\sqrt{2}) \text{ kg wt} + 4.472 \times (3/\sqrt{10}) \text{ kg wt},$$

where we've put in the value of w_2 just found; and this must exactly balance the negative (downward) component due to weight w_3 hanging from the vertical string. If you put in the numbers (do it yourself!) you'll find that w_3 must have the value $w_3 = 5.657$ kg wt.

The 'triangle of forces' is shown in Fig.8(b). Notice how the sides, which represent the force vectors with a scale 1 cm to every kg, are parallel to the strings in Fig.8(a); and that the triangle closes only when the third weight is chosen so that the vector sum of the forces on the bead s exactly zero. Instead of carefully drawing pictures, and sliding the vectors around to form the triangle, we've been able to do all the work using only simple arithmetic. Remember (see Book 2) that the Greeks couldn't work this way because they never quite managed to bring algebra and geometry together.

In the Exercises that follow, you'll find other examples of how to use equilibrium conditions; but they are all solved in the same way – by first of all asking what forces are acting at a point and then resolving them into their components along two perpendicular directions (for forces in two dimensions - a plane); or three directions for forces in three dimensional space.

The science of forces in equilibrium is called **Statics**. When the forces are *not* in equilibrium, and result in movement of the bodies thay act on (usually non-uniform motion), we are dealing with **Dynamics**. Statics and Dynamics together are branches of the science of **Mechanics**. In the next Chapter we begin to think about the way massive objects – from projectiles to planets – move under the influence of forces.

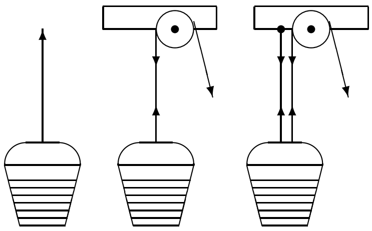
Exercises

1) Express the tensions in the strings (Figures 6 and 8), which keep the bead in equilibrium, in force units (Newtons).

2) How much do you weigh, in Newtons? And what is your mass? How much would you weigh if you were an astronaut, standing on the moon (where the value of g is about 1.70 m s^{-2})?

3) A bucket, hanging from a rope, is used to take water from a well. When empty it weighs 1 kg wt; when full it holds 9 litres of water and every litre has a mass of about 1 kg. What force (in Newtons) is needed to raise the full bucket? (The **litre** is a unit of volume: 1 litre = 10^{-3} m^3.)

4) The bucket (in the last exercise) can be lifted by passing the rope over a wheel, or 'pulley", as in Fig.9(b) below – so you can pull *down*, which is easier, instead of up. Do you have to pull just as hard in (b) as in (a)?



(a) (b) (c)

Figure 9

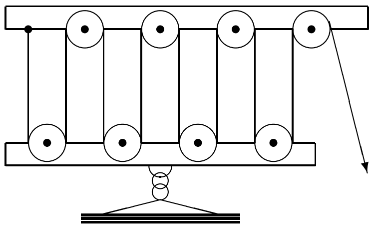


Figure 10

Now pass the rope under the handle of the bucket and fix the loose end to a point (•) on the beam that supports the pulley, as in (c). If you pull down hard, and produce a tension T in the rope, what force will you apply to the bucket? How big must T be, now, to raise the full bucket?

4) Suppose you have to lift a heavy iron bar, weighing 150 kg, which is much too much for any normal person. The last exercise shows you how it can be done, using only a rope and some pulleys. You need eight pulleys, a long rope, and two pieces of wood – and a few 'bits and pieces' for fixing them together as in Fig.10. The top piece of wood just supports four of the pulleys. The other piece of wood carries the other four pulleys; and has a hook and chain under it, for lifting things. Show that, if you can pull with a force of 20 kg wt, then you can lift up to 160 kg! Explain why.

5) A heavy truck is being pulled up a slope, as in Fig.11; its total mass is 1000 kg and it has wheels so it can run freely. The slope is '1 in 10' (1 metre vertically for every 10 metres horizontally). How hard must you pull on the rope (i.e.what tension T must you apply) to keep the truck from running downhill? (Hint: resolve the forces acting into components along and perpendicular to the slope, making things simpler by supposing the forces all act at a point – the middle of the truck. Don't forget that, besides the weight and the tension in the rope, the ground exerts an upward force R (the *reaction* of the ground,

taken perpendicular to the slope) to support part of the weight W. You will also need the sine and cosine of the angle of slope (θ, say): they are $\sin\theta = 0.0995, \cos\theta = 0.9950$. Can you calculate them for yourself?)

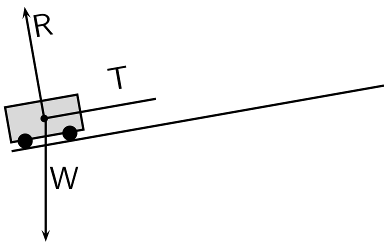


Figure 11

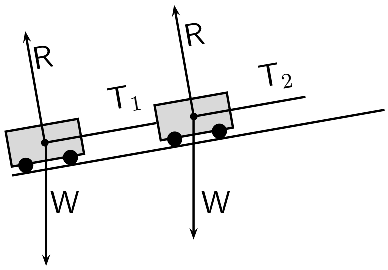


Figure 12

6) Now suppose two trucks are being pulled up the same slope, as in Figure 12. Both have the same weight; but what about the tensions T_1 and T_2 in the two ropes – can you calculate them? (Use the same method as in Exercise 5, writing down the conditions for equilibrium and solving the equations to get the values of T_1 and T_2.)

If the load is too heavy, which rope will break first?

Chapter 2

Work and energy

2.1 What is work?

Now you know something about force, and how it can be used to move things, we can start thinking about some of the other quantities that are important in mechanics – and the first of these is **work**. If you carry something heavy upstairs, or raise a bucket full of water from the well, you are doing *work* – and it makes you feel tired. In both cases you are applying a force (F, say) to an object; and you are moving it (a distance d, say, in the direction of the force). The force you are applying is equal to the weight of the object, mg, but in the opposite direction: taking the positive direction upwards, it is $F = mg$ (m being the mass of the object), while the weight has the same magnitude mg but is downwards.

Let's use W to stand for the work you've done (don't mix it up with *weight*!) and ask how it will depend on F and d. Suppose we double the mass of the object, then we double its weight mg (g being constant), which is also the force $F = mg$ needed to support it. And doubling the weight means doubling the work you have to do to raise the object through a distance d – carrying *two* sacks of flour upstairs, instead of just one, makes you feel twice as tired! In other words, the work done (W) is *proportional* to the force applied (F).

In the same way, doubling the *distance* through which you raise the object (going up *two* floors, instead of one) means doubling the work you have to do: so the work done (W) is also proportional to the distance (d).

To summarize, we suppose that W is proportional to the *product* of F and d: $W \propto F \times d$. And we can choose the unit of work (not yet fixed) so that

$$W = Fd. \qquad (2.1)$$

To be accurate, W is the work done when the force acting, F, "moves its point of application a distance d in the direction of the force". The unit of work is now defined, through (2.1) as the work done when $F = 1$ N and $d = 1$ m: it is 1 Nm, 1 Newton-metre. The physical dimensions of work are thus [W]= $\mathrm{MLT^{-2}} \times \mathrm{L} = \mathrm{ML^2T^{-2}}$. The "Newton-metre" is a big word for a unit; and usually we call it by the shorter name, the "Joule", after

James Joule (1818-1889), one of several people thinking about such things at about the same time.

The formula (2.1) is something we have *guessed*, through thinking about our own experience with lifting weights, but we'll find that it also holds good very generally for small massive objects moving under the influence of all kinds of force. Remember, however, that the W defined in (2.1) is the *physics* definition of work! You might feel you're working hard even just to hold a weight up – without actually lifting it (which makes $d = 0$ and $W = 0$). But in that case work is being done *inside your body*; your muscles are keeping themselves tight, so you can support the weight, and they are using the *chemical* energy that comes from 'burning up' your food. In Physics we're not usually talking about that kind of work, which is very difficult to measure.

Suppose now you've lifted a heavy stone (1 kg wt, say) to a great height (50 metres, say, above the ground). The work you've done is given by (2.1) and is $W = mg \times h = (1\text{kg}) \times (9.81 \text{ m s}^{-2}) \times (50 \text{ m}) = 490.5$ J – which seems quite a lot. But where has all that work gone? The stone doesn't look any different; but you've changed its *position* and it's now in a position to give you all that work back. When it's able to do work we say it has **energy**; and this particular kind of energy, which depends only on the *position* of the stone, is called **potential energy** and is usually denoted by the symbol V

How do we turn that energy back into work? We simply let the stone fall back to the ground: it does work by digging itself a hole in the ground or by breaking anything that tries to stop it! There are many other ways in which a falling weight can turn its potential energy back into useful work: think of a clock, driven by hanging weights – which you wind up at night, giving them enough energy to drive the clock all through the next day. We now want to find out about other forms of energy and how one form can be changed into another, or into useful work.

2.2 Two kinds of energy

Suppose you have a mass m at height h above the ground and you let it fall, from rest (i.e. not moving when you let go). Its potential energy (often we use 'PE', for short) is then $V_0 = mgh$ at the start of its fall. When it's fallen a distance s, however, its PE will be smaller, because its height above the ground will then be $h - s$ of instead of h. So the loss of PE is mgs. Where has it gone? The only thing the stone has got in return is *motion* – it started from rest and now, after time t say, it's going quite fast. We say the lost PE has been changed or 'converted' into **kinetic energy** (the Greek word for motion being *kinos*).

Let's now try to express this kinetic energy (KE for short) in terms of things that have to do with motion. The force acting on the stone, due to gravity, is constant and produces an acceleration g: so in every second its velocity will increase by 9.81 m s^{-1}; and after time t it will be gt. More generally, we can suppose that anything moving with constant acceleration a will have a velocity $v_1 = at_1$ at time t_1 and will have gone a distance s_1; and at some later time t_2 it will have a velocity v_2 and will have gone a distance s_2. So we know the force acting, $F = ma$, and the distance moved in the direction of the force, $s = s_2 - s_1$. All we need now is a formula giving s in terms of t; and this we already know from Book 3, Section 2.1.

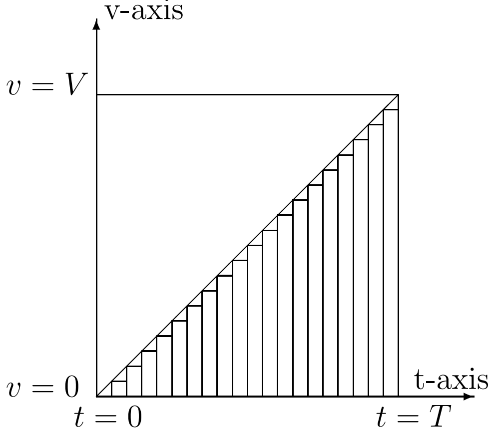


Figure 13

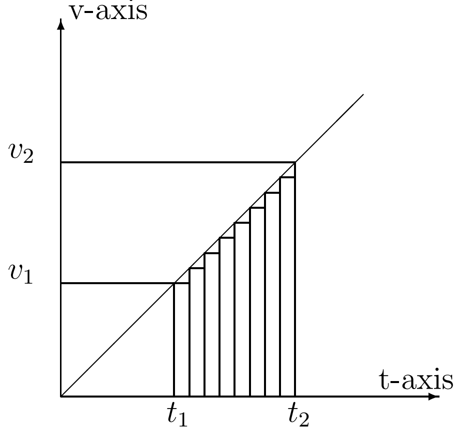


Figure 14

We got the formula by a graphical method: Fig.13 will remind you of it. We used T for the upper limit of time and found that the distance gone as t increases from $t = 0$ to $t = T$ was given by

$$s = \tfrac{1}{2}aT^2. \qquad (2.2)$$

This is represented by the area under the line $v = at$, between the lower and upper boundaries at $t = 0, T$ – which is just half the area of a rectangle with sides T (horizontally) and $V = aT$ (vertically). The result is exact in the limit where the strips, into which the area is divided, become infinitely narrow: it is called the "definite integral of the velocity, with respect to time, between the limits 0 and T".

What we really need is $s_2 - s_1$, the distance moved as t goes from t_1 to t_2. This is represented instead by the shaded area in Fig.14 – which is that of a rectangle (of width $t_2 - t_1$ and height v_1), with a triangle (of the same width, but vertical height $v_2 - v_1$) sitting on top of it. The sum of the two areas thus gives us

$$\begin{aligned} s_2 - s_1 &= (t_2 - t_1)v_1 + \tfrac{1}{2}(t_2 - t_1)(v_2 - v_1) \\ &= (t_2 - t_1)[v_1 + \tfrac{1}{2}(v_2 - v_1)] \\ &= \tfrac{1}{2}(t_2 - t_1)(v_2 + v_1). \qquad (2.3) \end{aligned}$$

Now we can get the work W done by the constant force F during the time $t_2 - t_1$. The force is related to the acceleration by $F = ma$, and since a is the slope of the straight line giving velocity against time we can say

$$F = ma = m \times \frac{v_2 - v_1}{t_2 - t_1}. \qquad (2.4)$$

Thus W = (Force)×(distance) is the product of (2.4) and (2.3):

$$\begin{aligned} W &= m \times \frac{v_2 - v_1}{t_2 - t_1} \times \tfrac{1}{2}(v_2 + v_1)(t_2 - t_1) \\ &= \tfrac{1}{2}m \times (v_2 - v_1)(v_2 + v_1) \\ &= \tfrac{1}{2}mv_2^2 - \tfrac{1}{2}mv_1^2 \end{aligned}$$

– in which the only variables left are mass and velocity! As the time increases from t_1 to t_2, the velocity increases from v_1 to v_2, and the quantity $\frac{1}{2}mv^2$ increases from $\frac{1}{2}mv_1^2$ to $\frac{1}{2}mv_2^2$. This quantity gives us the precise definition of what we have been calling the kinetic energy:

The kinetic energy of a point mass (m), moving with velocity v is

$$\text{Kinetic energy} \quad K = \tfrac{1}{2}mv^2 \tag{2.5}$$

We've now solved the mystery of where the potential energy went! In summary,

> Work done by gravity on a falling point mass
>
> = loss of potential energy ($V_1 - V_2$)
>
> = gain in kinetic energy ($K_2 - K_1$).

So if we use E to stand for the total energy $K + V$, we can say

$$E_1 = K_1 + V_1 = K_2 + V_2 = E_2 \tag{2.6}$$

– the total energy E *does not change* as we go from time t_1, at the beginning of the motion, to t_2, at the end of the motion. We say the total energy is 'conserved' throughout the motion (t_1 and t_2 being arbitrary times at which we make the observations). This result, which we've found only in one special case (for two kinds of energy and for motion under a constant force) is an example of one of the great and universal principles of Physics, that of the **Conservation of Energy**, which we study in more detail in later Sections.

Remember, all this is for a small 'point' mass – like the falling pebble – which we usually call a 'particle'. You can think of a big object as being made up from many small ones: it can then have an extra kinetic energy, coming from its *rotational* motion – but we'll come to that in Chapter 6. Until then we're going to talk only about the motion of single particles; or of things that can be treated *approximately* as just 'big particles' (not asking how? or why? until much later).

Remember also that this is the way science works: you go in small steps, using the simplest 'model' you can imagine, as long as it includes the things (like mass, velocity, force) that seem to be important. Models *in the mind* can easily be thrown away if they don't work! You don't have to make them and then break them up.

2.3 Conservation of energy

A system such as a falling particle, in which the energy is constant – as in (2.6) – is an example of a **conservative system**; no energy is going in or out and the energy it has is *conserved*. With this simple idea we can describe many kinds of particle motion, even without making calculations. All we need do is draw a graph to show how the potential energy of the particle (V) depends on its position (x, say, if it is moving along the x-axis).

Suppose, for example, you throw a stone vertically upwards with a velocity v. Its potential energy, if we use x to mean distance above the ground, will be $V(x) = mgx$. So plotting V against x will give a straight line of slope mg, as in Fig.15 which shows a "potential energy diagram". A horizontal line has been added, at height E in the diagram, to represent the constant total energy. What can this tell us about the motion?

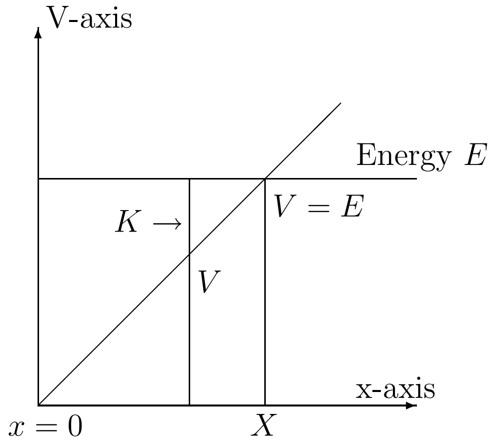


Figure 15

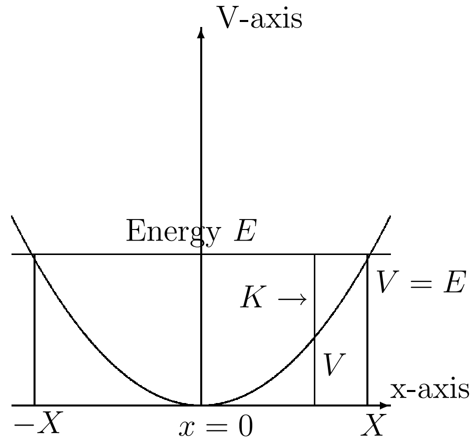


Figure 16

At the start, $E = K_1 + V_1$ is entirely KE, the energy of motion you have given to the stone. $V_1 = V(x_1)$ (the PE) is zero when $x_1 = 0$. At any later time, when the stone is slowing down and has risen to height x_2, we can say $K_2 + V_2 = E$. So the new KE will be $K_2 = E - V(x_2)$ – and this is represented by the distance from the PE curve up to the horizontal line at energy E. Now the important thing to notice is that K must always be *positive*, being proportional to the square of the velocity; when $x = x_2$ is the point where the E-line crosses the PE curve the KE has fallen to zero and the stone stops, for an instant. The value of x_2 is then the maximum height of the stone; it can go no higher for that value of E. After that point, the value of x can only get smaller: going back down the curve you again reach the point $x = x_1 = 0$ and the stone hits the ground again – with all the KE you gave it at the start.

Another example is shown in Fig.16, where the PE curve describes the motion of a pendulum - a small weight on the end of a string, which can swing backwards and forwards. Here $x = 0$ describes the position of the weight when it is hanging vertically, in equilibrium, and non-zero values of x will correspond to displacement of the particle when you push it away from the vertical. When you push it you have to do work and the amount of work done gives you $V = V(x)$, the PE function. As in the first example, the increase in PE as x changes from $x = x_1 = 0$ to $x = x_2$ gives you the energy stored, $V_2 = V(x_2)$; and, if you then let go,

$$K_1 + V(x_1) = K_2 + V(x_2) = E, \tag{2.7}$$

will describe how the balance of energies (between kinetic and potential) can change. If you release the particle from rest at $x = x_1 = X$, the KE will be zero and $E = V(X)$ will fix the constant total energy. At all other points the KE will be $K = E - V(x)$, represented as in the Figure: it can never go negative and so the motion is *bounded* at $x = X$ and $x = -X$, the ends of the swing at which K becomes zero and the motion reverses.

Notice that in both examples we are getting a lot of information about the motion without actually solving (or 'integrating') Newton's equation (1.1) – but that's really because we've done it already in finding the energy conservation equation (2.6)! You'll understand this better in the next Section, where we begin to use what we know about calculus.

First, however, remember that the conservation equation was found only for one particular kind of force, the force due to gravity, which is constant; and that in this case the potential energy is a *function of position*, $V = V(x)$, as indicated in (2.7). Forces of this kind are specially important: they are called **conservative forces** and they can always be derived from a potential function. We look at more general examples in the next Chapter.

2.4 Doing without the pictures – by using calculus

Our starting point for studying the motion of a particle was Newton's second law (1.1): $F = ma$, where F is the force acting on the particle, m is its mass, and a is the rate at which its velocity increases (in the same direction as the force). If you've read some of Book 3, you'll remember that the rate of change of velocity v with respect to time t (which defines the acceleration, a) can be written as

$$a = \frac{dv}{dt}$$

and is the limiting value of a ratio $\delta v/\delta t$ – where δv is a very small change of v, arising in the small time change δt.

Now the mass is just a constant factor, multiplying a, so Newton's law can also be stated as

$$F = m\frac{dv}{dt} = \frac{d(mv)}{dt} = \frac{dp}{dt}, \tag{2.8}$$

where $p = mv$, the mass of the particle times its velocity, used to be called the "quantity of motion" in the particle when it moves with velocity v: nowadays it's usually called the **momentum** of the particle, or more fully the **linear momentum** since it's 'in a line'. The usual symbol for it is p, but others are sometimes used (so watch out!).

Although we've been talking mainly about force, the rate of change of p, the momentum itself is also an important quantity. You'll understand this when we talk about *collisions* in which something massive and moving fast is suddenly stopped: if you're going fast and run into a stone wall it's your momentum that does the damage!

Newton's law in the form (2.8) is a **differential equation**: it determines the momentum p as a function of time (t), provided we know F as a function of time. And we know from

Book 3 Chapter 2 that if we're told the rate of change of something then we can find the 'something' by **integration**. When $\mathrm{d}p/\mathrm{d}t = F(t)$ we say

$$p = \int F(t)\mathrm{d}t, \tag{2.9}$$

where $\int ...\mathrm{d}t$ means "integrate with respect to t" to get p as a function of time. If the force is applied at time t_1 and continues to act until t_2 we can also find the *change* of momentum, $\Delta p = p_2 - p_1$, as the 'definite' integral

$$\Delta p = p_2 - p_1 = \int_{t_1}^{t_2} F(t)\mathrm{d}t, \tag{2.10}$$

between the 'limits' $t = t_1$ (the 'lower' limit) and $t = t_2$ (the 'upper' limit). So (2.9) gives you a function of time, $p = p(t)$, whose derivative is $F(t)$ – for whatever value t may have; while (2.10) gives you a single quantity Δp – the difference of p-values at the end ($p_2 = p(t_2)$) and at the beginning ($p_1 = p(t_2)$) of the time interval.

Where does all this get us if we don't know the force F as a function of time – how can we do the integration? Well, in general, we can't! But so far we were always talking about motion under a *constant* force; and in that very special case we could do the integration, for $F = ma$, a being the constant acceleration. In that case

$$p = \int (ma)\mathrm{d}t = mat = mv,$$

where $v = at$ is the velocity at time t of the particle moving with acceleration a.

Let's do something less easy. At some time, everyone plays a game with a bat (or heavy stick) and ball: you hit the ball with the stick and see how far you can send it. You hit the ball as hard as you can; but the 'hit' lasts only a very short time - a tiny fraction of a second - and the rest of the time the ball is on its journey through space, with only a much smaller force acting on it (gravity, which in the end brings it down to the ground). Imagine what happens at the time of the hit: the stick strikes the ball with a great force, that knocks it out of shape a bit and sends it on its way. As soon as the ball moves it loses contact with the stick and the force drops to zero. Now we ask how the force $F = F(t)$ must look, as a function of time, during that split second when the bat and ball are in contact. Perhaps it will be something like Fig.17:

F will be zero (neglecting the small force due to gravity) *except* between times t_1 and t_2, say, when the bat meets the ball and the ball leaves it. But during the time of contact, perhaps only a thousandth of a second, it will rise very suddenly to a very large value and then drop very suddenly to almost nothing. In other words the 'force curve' will show a very sharp peak; and to make things easy we could think of it as a rectangular 'box' of width Δt and height F_{av} (the 'average' value of the force).

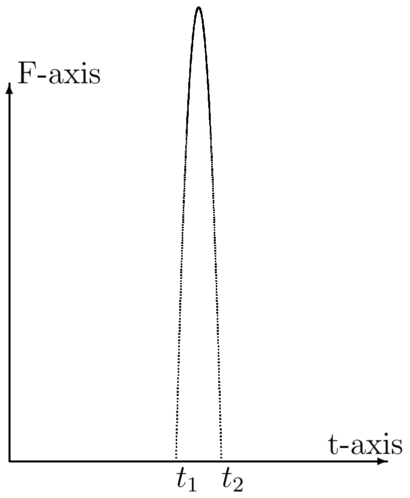


Figure 17

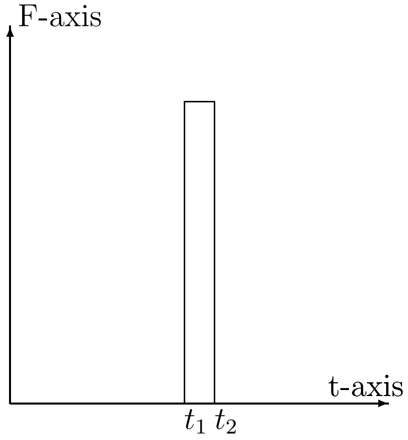


Figure 18

By using this simple 'model' of what's going on, with the *approximate* force curve shown in Fig.18, we can get a good idea of how big the force must be.

Suppose the ball has a mass of 0.2 kg and you give it a velocity of 10 ms^{-1}, starting from rest. Then p_2 in (2.10) will have the value $0.2 \times 10\mathrm{kgms}^{-1}$, which will also be the value of Δp. With the model we're using (Fig.18), this change of momentum is produced in 10^{-3}s. And the definite integral in (2.10) is simply the area under the curve of $F(t)$ between limits t_1 and t_2, which is the area of the 'box' and has the value (height×width) $F_{av} \times 10^{-3}$s. So, according to (2.10),

$$F_{av} \times 10^{-3}\mathrm{s} = 0.2 \times 10\mathrm{kgms}^{-1}$$

and the average force acting on the ball before it leaves the bat will be

$$F_{av} = \left(\frac{0.2 \times 10\mathrm{kgms}^{-1}}{10^{-3}\mathrm{s}}\right) = 0.2 \times 10^4\mathrm{kgms}^{-2} = 2000\mathrm{N}.$$

That's about 200 kg wt! – as if two very heavy men were standing on the ball. And all you did was hit it with a small piece of wood! A force of this kind, which is very large but lasts only a very short time, is called an 'impulsive' force, or just an **impulse**. This kind of force produces a sudden change of *momentum*, which we get by integrating F with respect to the *time*. But in Section 2.2 we found that a force could also produce a change of *kinetic energy*, obtained by integrating F with respect to *distance* over which the force acts (i.e. moves its point of application in the direction of F).

To end this Section let's look at the connection between these two ideas. The change of KE is equal to the work done when the particle is displaced through a distance $\Delta s = s_2 - s_1$ in the direction of the force: it is the definite integral

$$\Delta K = K_2 - K_1 = \int_{s_1}^{s_2} F\mathrm{d}s, \qquad (2.11)$$

where the integrand (the part following the integral sign) represents the work done in the *infinitesimal* displacement ds. But the change of momentum is given in (2.10), written out again here to show how similar the two things look:

$$\Delta p = p_2 - p_1 = \int_{t_1}^{t_2} F\mathrm{d}t.$$

In (2.11), we think of F as a function of distance gone: $F = F(s)$; but the values of any one of the variables s, t, v will determine a particular point on the path, so we can equally well think of F as a function of t, or of v. And in Book 3 we learnt how to 'change the variable', obtaining the rate of change of a function $y = f(x)$, with respect to x, in terms of that for another variable $u = u(x)$: the rule for differentiating was (Chapter 2 of Book 3)

$$\frac{\mathrm{d}y}{\mathrm{d}x} = \frac{\mathrm{d}y}{\mathrm{d}u}\frac{\mathrm{d}u}{\mathrm{d}x}$$

and in Chapter 4 we applied this rule to integration (the inverse process), where it took the form

$$\int f(x)\mathrm{d}x = \int f(v)\frac{\mathrm{d}v}{\mathrm{d}x}\mathrm{d}x = \int f(v)\mathrm{d}v \tag{2.12}$$

– remembering that, in the integral, the name we give the variable doesn't matter.

Now in (2.11), even though we don't know F as a function of s, we can easily introduce the velocity v as a new variable: thus

$$F = ma = m\frac{\mathrm{d}v}{\mathrm{d}t} = m\frac{\mathrm{d}v}{\mathrm{d}s}\frac{\mathrm{d}s}{\mathrm{d}t} = mv\frac{\mathrm{d}v}{\mathrm{d}s}.$$

This means the definite integral in (2.11) can be rewritten as

$$\int_{s_1}^{s_2} F\mathrm{d}s = \int_{s_1}^{s_2} mv\frac{\mathrm{d}v}{\mathrm{d}s}\mathrm{d}s = \int_{v_1}^{v_2} mv\mathrm{d}v = \tfrac{1}{2}m(v_2)^2 - \tfrac{1}{2}m(v_1)^2 = K_2 - K_1.$$

Of course, you'll say, we knew this result already from the graphical method we used in Section 2.2; but the results we got were only for motion under a constant force (giving constant acceleration), as in the case of a freely falling body. But now we know how to handle the *general* case, by using the calculus. If there is a force (e.g. the resistance of the air), trying to slow the particle down, we can still calculate what will happen – provided we know how the force depends on the velocity – even though the energy conservation equation may no longer hold.

2.5 Other forms of energy

So far, we've come across two main kinds of energy: potential energy (PE), which depends on position of a particle in space and not on how fast it is moving; and kinetic energy (KE), which depends only on its velocity. There are other kinds, which may not even involve a particle, which we'll meet in other Chapters. Here we'll introduce only one more kind of PE – the energy of a stretched spring or piece of elastic, which we just call a 'system'. If we change such a system by bending it or stretching it, then we do work on it and the work done is stored as potential energy. When we let go, the system returns to its normal condition and this energy is released. It may turn into KE (the spring may jump into the air) or, if it's a clock spring, it may come out slowly – turning wheels and pointers to show you the time.

Usually, the system is in equilbrium before you do any work on it; and this means that some 'coordinate' (like the length of a spring before you stretch or compress it) has an 'equilibrium value' which can be taken as $x = 0$. And the force you have to apply will be proportional to the amount of the displacement from $x = 0$: we write

$$F(x) = -k|x|, \tag{2.13}$$

where $F(x)$ is the force in the spring when the displacement is x and k is called the "force constant". The modulus $|x|$ (i.e. x without any $\pm$ sign) is used because usually it's only the amount of the displacement that counts – not whether it's in one direction or the other (left or right, up or down). But the $-$ sign before the k means that if x is positive the force $F(x)$ will be *negative*, towards $x = 0$, while if x is negative the force will be in the direction of the *positive* x-axis. In both cases the force is a 'restoring force', trying to bring the system back to its equilibrium condition with $x = 0$. The force law (2.13) is known as "Hooke's law" and the value of k is a property of the system, to be found by experiment.

The energy stored in the spring, for any value of x, can be obtained by integrating the force you have to apply to stretch it. Thus, taking x positive, the force to be applied will be *opposite* to that in the spring and will be $F = +kx$, while the work done in increasing x to $x + \delta x$ will be

$$\delta V = \text{force in positive direction} \times \delta x = kx\delta x.$$

It follows that the potential energy corresponding to a finite displacement x will be the definite integral

$$V(x) = \int_0^x kx\mathrm{d}x = [\tfrac{1}{2}kx^2]_0^x = \tfrac{1}{2}kx^2. \tag{2.14}$$

This function is symmetric about the point $x = 0$, taking the same value when x changes sign, and is in fact the parabola shown in Fig.16 which applies for a swinging pendulum. The same form of PE function holds good for many kinds of energy storage device.

2.6 Rate of working – power

We started to talk about work and energy in Section 2.1 and have come quite a long way, finding important general principles such as the conservation of the total energy $E = K + V$ for any system with only conservative forces. The 'work equation' (2.1) expressed our everyday experiences of carrying sacks of flour upstairs as a simple formula, which led us on to everything that followed. But now we need a new concept. The work you can do is not the only important thing: sometimes *how fast* you can do it is even more important. The *rate* of doing work is called **power**; and the more sacks of flour you can cary upstairs in a given time the more powerful you are! The same is true for machines, of course: you can do more work in the same time if you use a more powerful machine.

All we have to do now is make the definition a bit more precise and decide how to measure power – what will be its units? Suppose that one or more forces act on a system and do an amount of work W (calculated by using (2.1) for each force acting) in a time interval t. Then the ratio W/t will be the average rate of doing work during that interval: it will be called the "average power" consumed by the system and denoted by P. When W is measured in Joules and t in seconds, P will be expressed in units of J s^{-1} (Joules per second). The dimensions of P will be $[P]=\mathrm{ML^2T^{-2}}\times\mathrm{T^{-1}} = \mathrm{ML^2T^{-3}}$ and the unit of power will thus be 1 J s^{-1} or, in terms of the primary units, 1 kg m^2 s^{-3}; this unit is called the Watt (after James Watt, who invented) the steam engine and 1 W = 1 J s^{-1}. The Watt is quite a small unit and the power of small engines in everyday use is very often several thousand Watt, a few kW (kiloWatt).

In science we are usually interested in the *instantaneous* power a machine can give us, not in the average over a long period of time, and this is defined by going to the limit where t becomes infinitesimal:

$$P = \mathrm{d}W/\mathrm{d}t. \tag{2.15}$$

In the Exercises that follow you'll find examples of how all such concepts can be used.

Exercises

1) Look back at Fig.9(b) (end of Section 2.1) which shows a bucket of water being raised from the well. Suppose the bucket, with its water, has a mass of 10 kg and that it has to be raised by 4 m. How much work (W) has to be done? and what force is doing the work?

2) Now look at Fig.9(c), where the bucket seems to be carried by two ropes (even though there's really only one). Why is the tension (T) you have to apply to the rope only half what it was in Exercise 1? Is the work you have to do, to raise the bucket through 4 m, now only half as much as it was? If not – why not?

3) If you let the rope slip when the full bucket is at the top, how much KE will it have when it hits the water? and how fast will it be going?

4) Look at Fig.15 which is an energy diagram for a stone thrown vertically upwards. Suppose the stone has mass 0.1 kg and is thrown up with a speed of 10 m s^{-2}.

How much KE does the stone start with? And how high will it rise before it stops and starts to fall? If you are 1.5 m tall and the stone hits your head on the way down how much KE will it still have?

5) Fig.11 shows a truck being pulled up a slope. The mass of the truck is 1000 kg. Calculate the reaction (R) and the tension (T) in the rope. How much work must you do to pull the truck slowly up to the top, a distance of 10 m? (and why do we say "slowly"?) Where does this work go to?

If the rope snaps, at the top of the slope, what happens to the truck and the forces acting on it? What speed will it have when it reaches the bottom of the slope?

6) In Exercise 4 you calculated the KE (which will also tell you the velocity v of the stone

when it hits your head. What will its momentum be? Now suppose the time of contact is about 0.1 s (before you are knocked out!) and use the same 'model' as in Fig.18 to estimate the average 'contact force' during that short interval. What is its value (i) in Newtons, (ii) in kg wt, and (iii) as a multiple of the weight of the stone when it's not moving.

Chapter 3

Motion of a single particle

A note to the reader

Some parts of this Chapter are difficult (including the first Section); but don't be put off – they are only showing how what we know already, about work and energy and motion, holds good very generally. You'll find that many things start coming together – ideas about space and geometry (from Book 2) and about using the calculus (from Book 3) – and that you can get a good idea of what is happening, even without fully understanding all the details. Later in the Chapter you'll be surprised by how far you can go with nothing more than simple arithmetic.

3.1 What happens if the force on a particle is variable and its path is a curve?

So far we've nearly always been thinking of motion in a straight line, resulting from a constant force. The distance moved (x say) was a function of the time t, $x = x(t)$, and so was the velocity, $v = v(t)$, while the acceleration a was simply a constant. When we turn to motion in "three-dimensional space" ('3-space' for short) things are a bit more difficult because every point in space needs *three* coordinates to describe its position; and every velocity needs three components; and so on. Book 3 has prepared the way for dealing with motion in 3-space – the science of **kinematics** – but now we want to deal with real particles (which have mass, and are acted on by forces) and this takes us into physics.

We have already met scalar quantities (such as distance, speed, kinetic energy, potential energy) which all have magnitudes but do not depend on any particular direction in space; and also vector quantities, which besides having a magnitude are also dependent on a direction. The first vector we meet is the **position vector** of a point in space (we'll nearly always be talking about vectors in 3-space, so the 3 will usually be dropped): it will be the vector that points from the **origin** of a system of coordinates to the point with coordinates x, y, z and can be expressed as (see Fig.19)

$$\mathsf{r} = x\mathsf{e}_1 + y\mathsf{e}_2 + z\mathsf{e}_3 \qquad \text{or} \qquad \mathsf{r} \rightarrow (x, y, z). \tag{3.1}$$

Note that a special type (e.g.r) is used for a vector quantity, as distinct from a scalar.

In the first equation in (3.1), $\mathsf{e}_1, \mathsf{e}_2, \mathsf{e}_3$ are **unit vectors** in the directions of the three coordinate axes (e_1 for the x-axis, e_2 for the y-axis, e_3 for z-axis) and the vector equation $\mathsf{r} = x\mathsf{e}_1 + y\mathsf{e}_2 + z\mathsf{e}_3$ simply means you can get to the point with coordinates x, y, z by taking x unit steps (e_1) in the x-direction, y in the y-direction, and z in the z-direction. Remember (Book 2) it doesn't matter what *order* you take the steps in (see Fig.19) – you get to exactly the same end point, with coordinates x, y, z. The second statement in (3.1) is just another way of saying the same thing: the vector r has associated with it the three numerical components x, y, z.

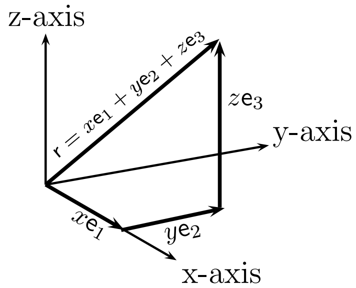


Figure 19

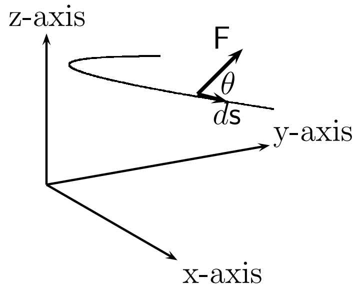


Figure 20

One of the nice things about vector equations is that a sum like $\mathsf{c} = \mathsf{a} + \mathsf{b}$ means the vectors on the two sides of the = sign are equal *component-by-component*. A single vector equation is equivalent to three scalar equations:

$$\mathsf{c} = \mathsf{a} + \mathsf{b} \qquad \text{means} \qquad c_1 = a_1 + b_1, \quad c_2 = a_2 + b_2, \quad c_3 = a_+ b_3. \tag{3.2}$$

Usually, we'll be working in terms of components; but sometimes it helps to use vector language – if you're in trouble go back to Book 2.

As an example of using vectors in dynamics we'll be looking at the motion of a particle (or even a planet, moving round the sun!) when the force acting on it is not constant and its path is not a straight line. How must the things we discovered in Chapter 2 be changed? Does the principle of energy conservation, for example, still hold good when we go from motion along a straight line to motion along some curve in 3-space?

Clearly the work equation (2.1) must be extended from 1 dimension to 3. We'll be needing a general definition of the work done when a particle is moved along some path, through an infinitesimal distance ds, but *not* always in the direction of the force acting. Let's write the corresponding bit of work done as (looking at Fig.20)

$$w = F \cos\theta ds = \mathsf{F} \cdot d\mathsf{s}. \tag{3.3}$$

Here θ is the angle between the force, which is a vector F, and the vector element of path, $d\mathsf{s}$. The work done is thus the magnitude F of the force times the distance moved, $\cos\theta ds$, in the direction of the force; or, equally, the force component $F\cos\theta$ in the direction of the displacement vector $d\mathsf{s}$. The second form in (3.3) shows this quantity as the 'scalar product' of the two vectors F and $d\mathsf{s}$ – which you will remember from Book 2, Section 5.4. (If you don't, just take it as a bit of notation for what we've described in words.) Now if we resolve the vectors F and s into their *components* along x-, y- and z-axes at the point we're thinking of, the element of work done (3.3) becomes simply

$$w = F_x \mathrm{d}x + F_y \mathrm{d}y + F_z \mathrm{d}z, \tag{3.4}$$

where $\mathrm{d}x, \mathrm{d}y, \mathrm{d}z$ are the three components of the path element ds. In other words, each component of the force does its own bit of work and adding them gives you the work done by the whole force – in any kind of displacement!

You'll be wondering why the letter w has been used for the very small element of work done, instead of $\mathrm{d}w$, though
$rmdx, \mathrm{d}y, \mathrm{d}z$ have been used for small *distances*. That's because the distance between two points depends only on where they are (on their *positions*) and not on how you go from one to the other: the small separations are **differentials** as used in Calculus (Book 3). But the work done in going from one point to another is not like that: if you drag a heavy object over a rough surface, going from Point A to point B, you'll soon find that the work you have to do depends on what *path* you follow – the longer the path and the more work you have to do! So it would be wrong to use calculus notation for something that is not a differential. More about this in the last Section of this Chapter.

What force are we talking about in setting up equation (3.4)? In Sections 2.1 and 2.2 we met two kinds of force: one was the weight of a particle and came from the *field* due to gravity (you can't *see* it, but you know it's there because the particle falls; the other was a force you apply to the particle, by lifting it to *feel* the weight. When you just stop it falling the two forces are equal but opposite, the resultant force is zero and the particle is in equilibrium. By moving something slowly (no kinetic energy!) the work you do *on* the particle is stored in the particle as potential energy. But, for a particle moving freely in an orbit (no touching!), the work w is being done by the field and is 'wasted' work in the sense that the particle is *losing* its ability to do any further work (which is its *potential* energy): so $w = -\mathrm{d}V$ and (3.4) can be rewritten as

$$\mathrm{d}V = -(F_x \mathrm{d}x + F_y \mathrm{d}y + F_z \mathrm{d}z). \tag{3.5}$$

(Note that V, defined in Section 2.1 for a very special example, depended only on position – was a function of position – so a small difference could be correctly called dV: now we're thinking of the *general* case and we're going to find the same thing is true.)

If you want to get that PE back then you must take the particle slowly back to where it came from by applying equal but opposite forces at every point on the path: that means changing the signs of F_x, F_y, F_z to get a positive dV – which will then be the *increase* in PE arising from the work *you* have done on the stone.

Now let's get back to the freely moving particle and ask if the total energy (PE plus KE) is conserved during the motion. To do that, we must now look at the kinetic energy K.

The KE is a scalar quantity, $K = \frac{1}{2}mv^2$, where v is the magnitude of the velocity (i.e. the speed), and is easily written in terms of the velocity *components* because $v^2 = {v_x}^2 + {v_y}^2 + {v_z}^2$. Thus

$$K = \tfrac{1}{2}m({v_x}^2 + {v_y}^2 + {v_z}^2). \tag{3.6}$$

To find how K changes with time we can differentiate (Book 3, Section 2.3):

$$\frac{\mathrm{d}K}{\mathrm{d}t} = \tfrac{1}{2}m\left(2v_x\frac{\mathrm{d}v_x}{\mathrm{d}t} + 2v_y\frac{\mathrm{d}v_y}{\mathrm{d}t} + 2v_z\frac{\mathrm{d}v_z}{\mathrm{d}t}\right).$$

But, by the second law, $m(\mathrm{d}v_x/\mathrm{d}t) = ma_x = F_x$, and similarly for the other components. On putting these results into the formula for dK/dt we get

$$\frac{\mathrm{d}K}{\mathrm{d}t} = F_xv_x + F_yv_y + F_zv_z = F_x\frac{\mathrm{d}x}{\mathrm{d}t} + F_y\frac{\mathrm{d}y}{\mathrm{d}t} + F_z\frac{\mathrm{d}z}{\mathrm{d}t} = -\frac{\mathrm{d}V}{\mathrm{d}t}, \tag{3.7}$$

where (3.5) has been used.

This result is the differential form of the energy conservation principle. When a particle moves along any infinitesimal element of path (represented by a displacement vector $d\mathsf{s}$), following Newton's second law, the change in total energy $E = K + V$ is zero:

$$\mathrm{d}E = \mathrm{d}K + \mathrm{d}V = 0. \tag{3.8}$$

A finite change, in which the particle moves from Point 1 on its path to Point 2, will then be a sum of the changes taking place in all the steps $d\mathsf{s}$. And

$$\Delta E = \Delta K + \Delta V = 0, \tag{3.9}$$

where $\Delta K = (\frac{1}{2}mv^2)_2 - (\frac{1}{2}mv^2)_1$ and

$$\Delta V = \int_1^2 -F_x\mathrm{d}x + \int_1^2 -F_y\mathrm{d}y + \int_1^2 -F_z\mathrm{d}z) = V_2 - V_1. \tag{3.10}$$

The integral in (3.10) is called a "path integral", a sum of contributions from all elements $\mathrm{d}\mathsf{s}$ of the path leading from from Point 1 to Point 2. The remarkable thing about this path integral is that it doesn't depend at all on the path itself! It has exactly the same value for any route leading from Point 1 to Point 2. We'll say more about this at the end of the Chapter. But here the important thing is that (3.9) is true for motion of a particle along *any* path, however long and curved it may be, when it moves according to Newton's second law. The principle of energy conservation, which we first met in Section 2.2, evidently applies very generally – as we'll see in the next two Sections.

3.2 Motion of a projectile

Something you throw or shoot into the air is called a "projectile": it could be a small pebble from your catapult, or a bullet from a gun. And it moves, under a constant force (that due to gravity), according to the Newton's second law. The problem is to find its path. This example is different from the one in Section 2.3 – the falling stone – because the motion is now two-dimensional: the projectile may start with a velocity component V_x, in the x-direction (horizontally), and a component V_y in the y-direction (vertically); and the only force acting (leaving out the small resistance of the air) is that due to gravity, which is mg and acts vertically downwards. This is all shown in Fig.21a, where V indicates the velocity vector at the start and • shows the projectile at point P(x, y) at a later time t.

So how does the projectile move?

Let's take $t = 0$ at the start of the motion: then at any later time t the components of position, velocity, and acceleration will all be functions of t; call them (in that order)

$$x(t), \quad y(t), \quad v_x(t), \quad v_y(t), \quad a_x(t), \quad a_y(t). \tag{3.11}$$

At the start of the motion, we can take

$$x(0) = y(0) = 0(\text{the origin}), \quad v_x(0) = V_x, v_y(0) = V_y \quad (\text{given}),$$
$$a_x(0) = 0, a_y(0) = -g \quad (\text{constant acceleration , downward}).$$

Motion with constant acceleration was studied in Chapter 2 of Book 3 and the results were used again here in Section 2.2; to summarize

$$\text{Velocity increase at time } t \text{ is } v = at \quad \text{Distance gone is } s = \tfrac{1}{2}at^2.$$

For the projectile we can use the same results for each of the two components, so we need only change the names of the variables. The equations become

$$v_x(t) = V_x + a_x(t) = V_x, \quad x(t) = V_xt \qquad (\text{x} - \text{component})$$
$$v_y(t) = V_y + a_y(t) = V_y - gt, \quad y(t) = V_yt - \tfrac{1}{2}gt^2 \qquad (\text{y} - \text{component}),$$

where it was remembered that at gives the velocity *increase* as time goes from zero to t and that the starting velocity is, in this example, non-zero – with components V_x, V_y.

We can now plot the path of the projectile: at time t its coordinates will be

$$x = V_xt, \quad y = V_yt - \tfrac{1}{2}gt^2. \tag{3.12}$$

In Fig.21b the whole curve is sketched, up to the point where the projectile hits the ground.

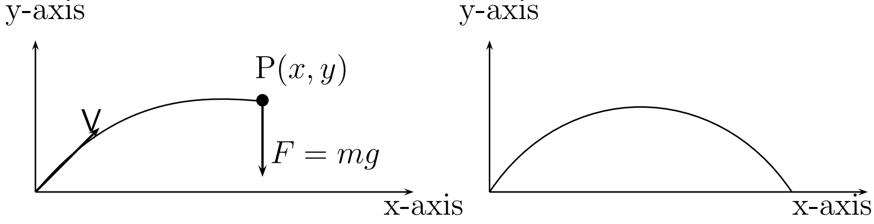


Figure 21a Figure 21b

Usually, when we plot the curve representing a function $y = f(x)$, the value of y (the dependent variable) is given directly in terms of x (the independent variable) by some formula. But here both x and y are expressed in terms of another variable t (the time), which is a *parameter*: together they give a **parametric representation** of the function $y = f(x)$. However, if we want the more usual form, we can easily eliminate the parameter t; because the first equation in (3.12) tells us that, given x, the time must be $t = x/V_x$ – and if we put that value in the second equation we find

$$x = V_x t, \quad y = V_y t - \tfrac{1}{2}gt^2.y = (V_y/V_x)x - \tfrac{1}{2}(g/V_x^2)x^2, \tag{3.13}$$

which is of the second degree in the variable x and describes a **parabola**.

From the equation for the path we can find all we want to know. How far does the projectile go before it hits the ground? Put $y = 0$ in (3.13) and you get

$$x\left(\frac{V_y}{V_x} - \frac{g}{2V_x^2}x\right) = 0$$

One solution is $x = 0$, the starting point, and the other is (get it yourself) $x = 2(V_xV_y)/g$; this is called the **range** – the maximum horizontal distance the projectile can go, for a given initial velocity.

And how high does the projectile go? The maximum value of the function $y = f(x)$ (or at least a value for which the slope of the curve is zero - in this case it will be the top) is reached when the first derivative $(\mathrm{d}y/\mathrm{d}x)$ is zero. So let's put

$$\frac{\mathrm{d}y}{\mathrm{d}x} = \frac{V_y}{V_x} - \left(\frac{g}{V_x^2}\right)x = 0.$$

This tells us that the highest point is reached when $x = (V_xV_y)/g$; and on putting this value in (3.13) (do it!) you'll find the corresponding value of y is $\frac{1}{2}V_y^2/g$.

Before ending this Section, we should note that if we don't want to know the whole path of the projectile – but only to answer questions like "how far?" and "how high?" – it's often quicker to use the energy conservation principle. Thus, when the force acting has only a (vertical) y-component the velocity x-component will not change from its initial value V_x; so its contribution to the KE will always be $\frac{1}{2}mV_x^2$ and the whole KE will be $\frac{1}{2}m(V_x^2 + v_y^2)$. The PE will depend on the height y only and, measured from 'ground' level, will be mgy. Energy conservation then means that the constant total energy (KE + PE) will be

$$E = \tfrac{1}{2}m(V_x^2 + V_y^2) + 0 \quad \text{initially} \quad = \tfrac{1}{2}mV_x^2 + \tfrac{1}{2}mv_y^2 + mgy \quad \text{at any later time}$$

and thus (x-terms cancelling) $\frac{1}{2}mV_y^2 = \frac{1}{2}mv_y^2 + mgy$. To get the maximum height we simply put $v_y = 0$ (upward velocity fallen to zero) and find the corresponding y-value from $\frac{1}{2}mV_y^2 = mgy$ – giving the same result $\frac{1}{2}V_y^2/g$ as before.

3.3 A numerical method

(**A note to the reader**. There's a lot of arithmetic in this Section and the next: you don't have to work through it all – just check a few lines here and there to make sure you understand what's going on.)

In the last Section we found the path of the projectile *analytically* – using the methods of mathematical analysis. It's not always easy, or even possible, to solve problems that way: but if you know the basic equations – in this case Newton's second law – you can always get there by using only simple arithmetic! To show how to do it we'll take the projectile problem again.

The coordinates and velocity components are all functions of time t, so we'll write them as $x(t), y(t), v_x(t), v_y(t)$ and set things going at $t = 0$, with the initial values

$$x(0) = 0, \quad y(0) = 0, \quad v_x(0) = 20, \quad v_y(0) = 20. \tag{3.14}$$

Here we've left out the units, but distances will be in (metres) m, velocities in ms^{-1}, accelerations in ms^{-2}; and we know the units will 'look after themselves' as long as we're careful about physical dimensions (see, for example, Section 2.1).

Let's go step by step from any starting value of t, using what Newton's law tells us and letting $t \to t + \epsilon$, in each step, ϵ being a small time interval (for example 0.1 s). The increases in coordinates x and y are then, respectively, ϵv_x and ϵv_y; and the new coordinates at the end of the step will be

$$x(t + \epsilon) = x(t) + \epsilon v_x, \quad y(t + \epsilon) = y(t) + \epsilon v_y.$$

But what values should we give to the velocity components? – because they will be changing when forces act and produce accelerations. Thus, v_x and v_y will change by ϵa_x and ϵa_y during the step from t to $t + \epsilon$: the value of v_x will be $v_x(t)$ at the beginning of the step and $v_x(t) + \epsilon a_x$ at the end, and similarly for v_y. To allow for the change in velocity components, we'll use the values corresponding to the *mid-point* of the step, with $\frac{1}{2}\epsilon$ in place of the full ϵ; so instead of taking, for example, $x(t+\epsilon) = x(t)+\epsilon v_x(t)$, we take $x(t+\epsilon) = x(t)+\epsilon v_x(t+\frac{1}{2}\epsilon)$. This means the velocity components have to be calculated at times ϵ apart, but *halfway through* successive intervals: to get $v_x(t + \frac{1}{2}\epsilon)$ from $v_x(t - \frac{1}{2}\epsilon)$ we'll simply add $\epsilon\times$acceleration, taking the acceleration at the midpoint which is now $a_x(t)$.

Our working equations for calculating quantities at time $t + \epsilon$ in terms of those at time t will now be – for the coordinates

$$x(t + \epsilon) = x(t) + \epsilon v_x(t), \quad y(t + \epsilon) = y(t) + \epsilon v_y(t) \tag{3.15}$$

– but for the velocities we should use

$$v_x(t + \tfrac{1}{2}\epsilon) = v_x(t - \tfrac{1}{2}\epsilon) + \epsilon a_x(t), \quad v_y(t + \epsilon) = v_y(t - \tfrac{1}{2}\epsilon) + \epsilon a_y(t). \tag{3.16}$$

These last two equations allow us to 'step up' the times by an amount ϵ, going from one interval to the next for as long as we wish. For the first point, $t = 0$ and we don't have values of $v_x(-\frac{1}{2}\epsilon)$ – as there's no interval before the first – but we can safely use $v_x(0 + \frac{1}{2}\epsilon) = v_x(0) + \frac{1}{2}\epsilon a_x(0)$ (velocity = time$\times$ acceleration) to get a reasonable start. And after that we can simply go step by step, using (3.15) and (3.16).

You'll see how it works out when we start the calculation. To do this we make a Table to hold the working equations:

t	$x(t+\epsilon) = x(t) + \epsilon v_x(t+\frac{1}{2}\epsilon)$ $v_x(t+\frac{1}{2}\epsilon) = v_x(t-\frac{1}{2}\epsilon) + \epsilon a_x(t)$	$y(t+\epsilon) = y(t) + \epsilon v_y(t+\frac{1}{2}\epsilon)$ $v_y(t+\frac{1}{2}\epsilon) = v_y(t-\frac{1}{2}\epsilon) + \epsilon a_y(t)$

and then one to hold them when we've put in the numerical values we know:

t	$x \rightarrow x + 0.1 \times v_x$ $v_x \rightarrow v_x + 0.1 \times (0) = v_x$	$y \rightarrow y + 0.1 \times v_y$ $v_y \rightarrow v_y - 0.1 \times (-10) = v_y - 1$

Here $\rightarrow$ is used to mean "replace by"; and we've chosen a 'step length' $\epsilon = 0.1$. The (constant) acceleration due to gravity is $a_y \approx -10$ with horizontal component $a_x = 0$. Note that the line which holds the velocity components gives the new values (on the left of the $\rightarrow$) at time $t+\epsilon$ – in terms of values two lines earlier, at time $t-\epsilon$. To remind us of this, the lines are labelled by the t-values used in the calculation.

Now we'll make a similar table to hold the numbers we calculate, using the rules above and the special starting values for the entries at $t = 0$. First, however, we note that some 'true' values of the (x, y) coordinates, calculated from the formulas in (3.12) at times $t = 0.4,\ 0.8,\ 1.2,\ 1.6,\ 2.0$ are (respectively)

$$(8.0, 7.2) \quad (16.0, 12.85) \quad (24.0, 16.80) \quad (32.0, 19.20) \quad (40.0, 20.00)$$

The first few (double)-lines in our Table of approximate values come out as:

$t = 0.0$ (0.05)	$x = 0.0$ $v_x = 20.0 + 0.05 \times 0.0 = 20.0$	$y = 0.0$ $v_y = 20.0 + 0.05 \times (-10) = 19.5$
$t = 0.1$ (0.15)	$x \rightarrow 0.0 + 0.1(20.0) = 2.0$ $v_x \rightarrow 20.0 + 0.1(0.0) = 20.0$	$y \rightarrow 0.0 + 0.1(19.5) = 1.95$ $v_y \rightarrow 19.5 + 0.1(-10) = 18.5$
$t = 0.2$ (0.025	$x \rightarrow 2.0 + 0.1(20.0) = 4.0$ $v_x \rightarrow 20.0 + 0.1(0.0) = 20.0$	$y \rightarrow 1.95 + 0.1(18.5) = 3.80$ $v_y \rightarrow 18.5 + 0.1(-10) = 17.5$
$t = 0.3$ (0.35)	$x \rightarrow 4.0 + 0.1(20.0) = 6.0$ $v_x \rightarrow 20.0$	$y \rightarrow 3.80 + 0.1(17.5) = 5.55$ $v_y \rightarrow 17.5 + 0.1(-10) = 16.5$
$t = 0.4$ (0.45)	$x \rightarrow 6.0 + 0.1(20.0) = 8.0$ $v_x \rightarrow 20.0$	$y \rightarrow 5.55 + 0.1(16.5) = 7.20$ $v_y \rightarrow 16.5 + 0.1(-10) = 15.5$
$t = 0.5$ (0.55)	$x \rightarrow 8.0 + 0.1(20.0) = 10.0$ $v_x \rightarrow 20.0$	$y \rightarrow 7.20 + 0.1(15.5) = 8.75$ $v_y \rightarrow 15.5 + 0.1(-10) = 14.5$
$t = 0.6$ (0.65)	$x \rightarrow 10.0 + 0.1(20.0) = 12.0$ $v_x \rightarrow 20.0$	$y \rightarrow 8.75 + 0.1(15.0) = 10.25$ $v_y \rightarrow 14.5 + 0.1(-10) = 13.5$
$t = 0.7$ (0.75)	$x \rightarrow 12.0 + 0.1(20.0) = 14.0$ $v_x \rightarrow 20.0$	$y \rightarrow 10.25 + 0.1(13.5) = 11.60$ $v_y \rightarrow 13.5 + 0.1(-10) = 12.5$

If you continue (try it!), you'll get the nice smooth curve shown in Fig.22.

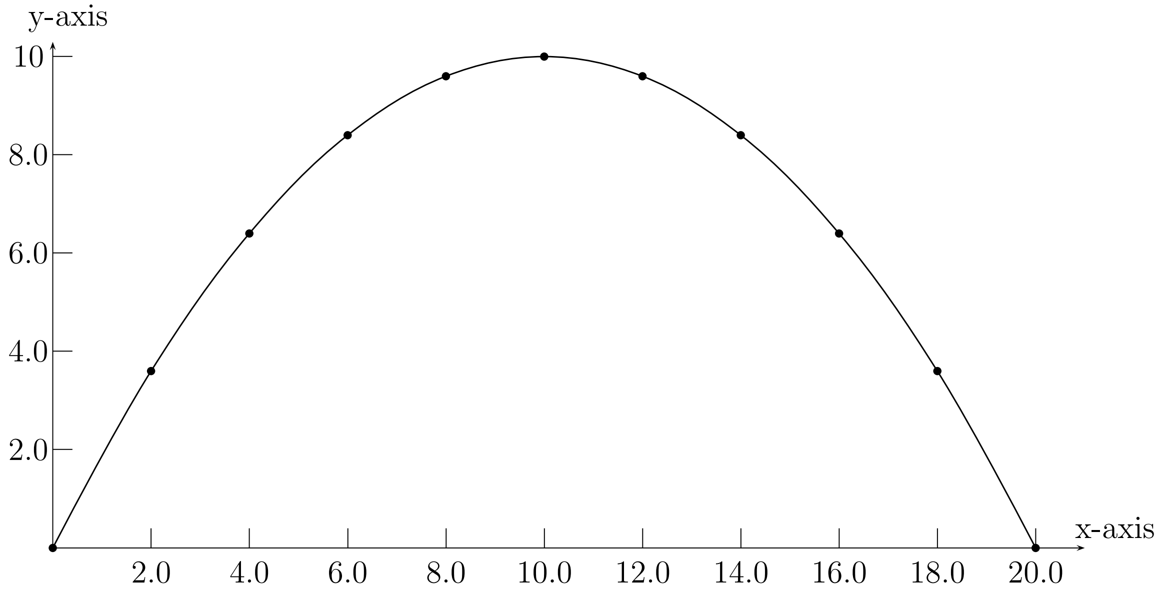


Figure 22

That was easy, and the results agree perfectly (can you say why?) with the exact results in Section 3.2, but what if we want to talk about a planet moving round the sun? In the next Section we'll find it's just as easy.

3.4 Motion of the Earth around the Sun

Suppose we have a single 'particle' (anything from a small pebble to the Moon, or the Earth!) moving along some path – like the one shown in Fig.20 – under the action of some force F. All we need to know, to find the path, is how the force depends on position of the particle; along with its position and velocity components, $x(0), y(0), z(0)$ and $v_x(0), v_y(0), v_z(0)$, at any time $t = t_0$ – which we call the 'starting time' and usually put equal to zero, $t_0 = 0$. It all worked out nicely in the last Section, where we used Newton's law (acceleration = force/mass) for each of the two components, to estimate how the velocity and position changed as the time increased by a small amount $t \rightarrow t + \epsilon$. But now we're going to do something a bit more exciting: will the same equations and methods work just as well when the 'particle' is the Earth – the whole of our world – on its journey round the Sun? If they do, and allow us to *calculate* that the journey will take about 365 days, then we can feel pretty sure that the law of gravity really is a *universal* law, applying throughout the Universe!

In Section 1.2 the force of attraction between two point masses, m and M, was given in equation (1.2) as $F = GmM/r^2$, where r was the distance between them and G was the 'gravitational constant'. But now we're working in three dimensions, using vectors and components, this must be written in a different way. Suppose the big mass M (the Sun) is used as the origin of coordinates: then the position vector of m (the Earth) will be

$$\mathsf{r} = x\mathsf{e}_1 + y\mathsf{e}_2 + z\mathsf{e}_3,$$

as in Fig.19. The force F is along the line of r, but is directed the opposite way – towards the origin – and must therefore be a multiple of $-\mathsf{r}$. Since r (the length of r) is simply $|\mathsf{r}|$, we can now write the force *vector* as

$$\mathsf{F} = \frac{GmM}{r^2}\left(\frac{-\mathsf{r}}{r}\right) = -\frac{GmM}{r^3}\mathsf{r}, \tag{3.17}$$

where the factor $-\mathsf{r}/r$, in the middle, is the *unit vector* pointing from the Earth to the Sun. When we express the final vector r in the form $\mathsf{r} = x\mathsf{e}_1 + y\mathsf{e}_2 + z\mathsf{e}_3$, the result is

$$\mathsf{F} = F_x\mathsf{e}_1 + F_y\mathsf{e}_2 + F_z\mathsf{e}_3,$$

where the components are

$$F_x = -\frac{GmM}{r^3}x, \quad F_y = -\frac{GmM}{r^3}y, \quad F_z = -\frac{GmM}{r^3}z. \tag{3.18}$$

We're now ready to write down the equations of motion for the Earth as it moves around the Sun, just as we did for the projectile in the last Section. This path will be the **orbit** of the Earth; and it will lie in a single plane – for if you take this as the xy-plane then the force F will never have a z-component to pull it out of the plane. So let's suppose the orbit is in the xy-plane with the Sun at the origin and the Earth at the point (x, y). The form of the orbit is shown in Fig.23a, which also shows the force vector F – directed always towards the Sun. We'll start the calculation at time $t = 0$, when the Earth is at the point (•) labelled 'Start', with coordinates $x(0), y(0)$.

The equation (mass)×(acceleration)=(force acting) then becomes, for the two components,

$$a_x = \frac{\mathrm{d}v_x}{\mathrm{d}t} = -GM\frac{x}{r^3}, \qquad a_y = \frac{\mathrm{d}v_y}{\mathrm{d}t} = -GM\frac{y}{r^3}, \tag{3.19}$$

where the mass m has been cancelled from each equation and we remember that, in terms of the coordinates x, y, the distance from the Sun is $r = \sqrt{x^2 + y^2}$.

The calculation will follow closely the one we made for the projectile, the main difference being that the acceleration is *not constant*, its components both being non-zero functions of position (x, y). Instead of (3.14), however, we take a starting point ($t = 0$) on the orbit, with

$$x(0) = R_0, \quad y(0) = 0, \quad v_x(0) = 0, \quad v_y(0) = V_0, \tag{3.20}$$

where R_0 is the initial distance from the Sun and V_0 is the intial velocity, in a direction perpendicular to the position vector.

Another difference, however, is that it's no longer sensible to work in units of kilogram, metres and seconds when we're talking about bodies with masses of many millions of kilograms, moving at thousands of metres every second. Wouldn't it be easier to use, for example, days or months? We know how to change from one set of units to another, provided we know the physical dimensions of the quantities we're talking about (see Section 1.3). Velocity, for example, has dimensions of distance ÷ time; so we write $[v] = LT^{-1}$ and if we multiply the unit of length by a factor k then we must *divide* the measure of any length by k – and similarly for the time factor.

Suppose we choose an 'astronomical' unit of length as $L_0 = 1.5 \times 10^{11}$ m, which the astronomers tell us is the average distance of the Earth, in its orbit, from the Sun; and the Month as the unit of time – 1 Month $\approx$ 30 days = $30 \times 24 \times 3600$ s = 2.592×10^6 s. The observed value of the velocity of the Earth in its orbit is also well known: it is about 30,000 m s^{-1} and we'll take this as the starting value of V_0.

To express V_0 in our new units we simply multiply the value in ms^{-1} by two factors: $(1.5 \times 10^{11})^{-1}$ for the length; and $((2.592 \times 10^6)^{-1})^{-1}$ (i.e. 2.592×10^6) for the time. The result is

$$V_0 = (3 \times 10^4)\left(\frac{2.592 \times 10^6}{1.5 \times 10^{11}}\right) L_0 \text{ Month}^{-1} \approx 51 \times 10^{-2}\, L_0 \text{ Month}^{-1}.$$

To summarize, a reasonable value of the start velocity seems to be about 0.51 distance units per month.

The only other quantity to express in our new units is GM in (3.9): this has dimensions (check it for yourself, using the data for G in Section 1.3) $[GM] = \mathrm{L}^3\ \mathrm{T}^{-2}$. The value of GM looks enormous in standard units (the Sun's mass alone is about 1.99×10^{30} kg!), while G has the value – measured in experiments here on the Earth – 6.67×10^{-11} m^3 s^{-2} kg. Expressed in the new units (check it yourself!) you should find

$$GM = 0.264\ L_0^3 \text{ Month}^{-2}.$$

And now, at last, we can start the calculation! – and we can just use the numbers, from now on, knowing that the units are sure to come out right in the end.

Our working equations for calculating quantities at time $t + \epsilon$ in terms of those at time t will now be just like those in (3.15) and (3.16): for the coordinates,

$$x(t+\epsilon) = x(t) + \epsilon v_x(t), \quad y(t+\epsilon) = y(t) + \epsilon v_y(t) \tag{3.21}$$

– but for the velocities we should use

$$v_x(t + \tfrac{1}{2}\epsilon) = v_x(t - \tfrac{1}{2}\epsilon) + \epsilon a_x(t), \quad v_y(t+\epsilon) = v_y(t - \tfrac{1}{2}\epsilon) + \epsilon a_y(t). \tag{3.22}$$

The only difference between these equations and those for the projectile is that the acceleration components a_x, a_y are given by (3.19): they must be calculated in every step instead of being constants (0 and -g). These last two equations allow us to 'step up' the times by an amount ϵ, going from one interval to the next for as long as we wish. Again, for the first point, $t = 0$ and we don't have values of $v_x(-\frac{1}{2}\epsilon)$ – as there's no interval before the first – but we can safely use $v_x(0 + \frac{1}{2}\epsilon) = v_x(0) + \frac{1}{2}\epsilon a_x(0)$ (velocity = time$\times$ acceleration) to get a reasonable start. And we can do exactly the same for the y-component. After that we simply go step by step, using (3.21) and (3.22).

Let's take $\epsilon = 0.2$, which is 6 days ($\frac{1}{5}$ Month in our working units) and make the first few time steps, starting from the values in (3.20) with $R_0 = 0.5$ and $V_0 = 0.51$. We then get the Table on the next page.

t		
$t = 0$	$x = 1.0$	$y = 0.0$
	$v_x = -0.0264$	$v_y = 0.5100$
$t = \epsilon$	$x = 0.9947$	$y = 0.1020$
	$r = \sqrt{.9947^2 + .1020^2} = 0.9999$	$r^3 = 0.9997, \quad 1/r^3 = 1.0003$
	$a_x = -0.264(0.9947)(1.0003) = -0.2627$	$a_y = -0.264(0.1020(1.0003) = -0.0269$
	$v_x \rightarrow -0.0264 - 0.2627(0.2) = -0.0789$	$v_y \rightarrow 0.5100 - 0.0269(0.2) = 0.5046$
$t = 2\epsilon$	$x = 0.9789$	$y = 0.2029$
	$r = \sqrt{.9789^2 + .2029^2} = 0.9994$	$r^3 = 0.9988, \quad 1/r^3 = 1.0012$
	$a_x = -0.264(0.9789)(1.0012) = -0.2587$	$a_y = -0.264(0.2029)(1.0012) = -0.0536$
	$v_x \rightarrow -0.0789 - 0.2587(0.2) = -0.1306$	$v_y \rightarrow 0.5046 - 0.0536(0.2) = 0.4939$
$t = 3\epsilon$	$x = 0.9528$	$y = 0.3017$

Notice that, at any given time (e.g. $t = 2\epsilon = 0.4$), the two lines after the calculation of the coordinates (x, y) are used in getting the corresponding components of the acceleration (a_x, a_y) – which are then used to get the average velocity components (next line) for calculating the distances gone in the next time step (up to $t = 3\epsilon = 0.6$).

If you keep going for 60 time steps you'll reach the point marked 'Day 360'. The (x, y)-values obtained in this way are plotted in Fig.23a. Each time step represents 6 days and the orbit has closed almost perfectly in 360 days – that's not a bad approximation to 1 year when you remember that we're doing a rough calculation, using only simple arithmetic (even though there's quite a lot of it!).

It may seem unbelievable that, starting from measurements made by the astronomers and a value of G obtained by measuring the force of attraction beteen two lead balls in the laboratory, we can *calculate* the time it will take the Earth to go round the sun – 150 million kilometres away! So you should check the calculation carefully.

To help you on your way, the last few steps, leading from $t = 58\epsilon$ up to $t = 60\epsilon$ (360 days), are given below:

t		
58ϵ	$x = 0.9784$ $a_x = -0.2585$	$y = -0.2059$ $a_y = -0.0544$
59ϵ	$x = 0.9944$ $a_x = -0.2626$	$y = -0.1050$ $a_y = 0.0277$
60ϵ	$x = 1.000$	$y = -0.0030$

Only the acceleration components are given, for time $t = 58\epsilon$, so you'll need the velocities v_x and v_y for $t = 57\epsilon$: these are $v_x = 0.1321$ and $v_y = 0.4935$. Now you should be able to fill in the missing values, just as in going from $t = 2\epsilon$ to $t = 3\epsilon$.

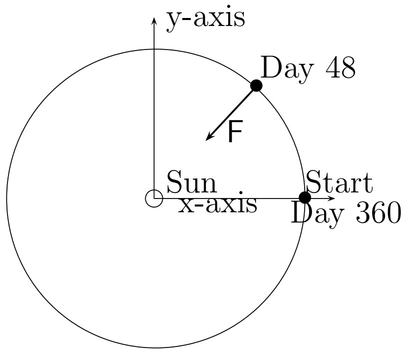


Figure 23a

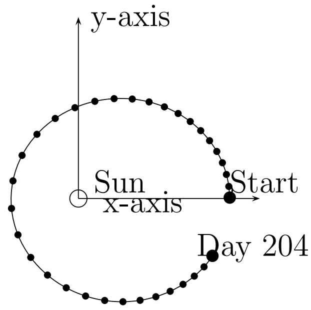


Figure 23b

You should note that the orbit is not exactly a circle, but rather an ellipse (with a 'long diameter' and a 'short diameter'); but the diameters differ only by less than one part in fifty. However, if you use different starting conditions at $t = 0$, you can get very different results: Fig.23b, for example, shows the effect of changing the starting velocity of the Earth in its orbit from 0.51 units to 0.40. In that case it would be drawn into a more 'lopsided' orbit, getting much closer to the Sun for much of its path; and the length of the year would be very different. Other planets, like Mars and Venus, have orbits of this kind: but more about such things in other Books of the Series.

A very important final conclusion is that once you know the equations of motion, and the values of the coordinates and velocity components at any given time t_0, then the way the system behaves *at all future times* is completely determined: we say that the equations are **deterministic** – nothing is left to chance! This is a property of many of the key equations of Physics.

3.5 More about potential energy

In Section 4.1 we found that the idea of 'conservation of energy' applied even for a particle moving along a curved path (not only for the up-down motion studied in Section 2.1), provided Newton's second law was satisfied. In all cases the gain in kinetic energy, as the particle went from Point 1 to Point 2, was exactly the same as the loss of potential energy – defined as the work done by the forces acting. The differential form of this result, $\mathrm{d}K + \mathrm{d}V = 0$, where $\mathrm{d}K$ and $\mathrm{d}V$ are infinitesimal changes in K and V then led us to the result $K + V =$ constant at all points on the path. (If you're not sure about using *differentials* look back at Section 2.3 in Book 3.)

This result means that the forces acting on the particle must have a special property: they are said to be **conservative forces** and when this in this case it is possible to define a potential energy function $V(x, y, z)$ – a function whose value depends *only on position* and whose differential $\mathrm{d}V$ appears in (3.5). And for forces of this kind it is possible to get the force components at any point in space from the single function $V(x, y, z)$.

To see how this can be done, think of the differentials $\mathrm{d}V, \mathrm{d}x, \mathrm{d}y, \mathrm{d}z$ simply as very small related changes, when you pass from point (x, y, z) to an infinitely close point $(x+\mathrm{d}x, y+\mathrm{d}y, z+\mathrm{d}z)$. We can then define a derivative of V with respect to x as the limit of the ratio $\mathrm{d}V/\mathrm{d}x$ when only x is changed: it is called a *partial* derivative and is written $\partial V/\partial x$, with a 'curly' d. Thus

$$\frac{\partial V}{\partial x} = \lim_{\mathrm{d}x \to 0} \frac{\mathrm{d}V}{\mathrm{d}x}, \quad (y, z \text{ constant}). \tag{3.23}$$

You will have met partial derivatives already in Book 3: there's nothing very special about them except that in getting them you change only one variable at a time, treating the others as if they were constants. When you have a function of three variables, like the PE with $V = V(x, y, z)$, you have three partial derivatives at every point in space – the one given in (3.23) and two more, with the 'special' direction being the y-axis or the z-axis.

The force components in (3.5) can now be defined as partial derivatives of the potential energy function:

$$F_x = -\frac{\partial V}{\partial x}, \quad F_y = -\frac{\partial V}{\partial y}, \quad F_z = -\frac{\partial V}{\partial z}, \tag{3.24}$$

where it is understood that the variables not shown, in each derivative, are kept constant.

Now, after all that work (brain work!), we can give a general definition of the potential energy of a particle. We start from any point O (calling it the 'origin' or the 'zero of potential energy') and carry the particle from O to any other point P. The PE given to the particle is then the work done in moving it from O to P:

$$V_P - V_O = W = \int_O^P (F_x \mathrm{d}x + F_y \mathrm{d}y + F_z \mathrm{d}z).$$

If we label the points as '0' and '1', this can be written

$$V_1 - V_0 = W(0 \to 1) = \int_0^1 (F_x \mathrm{d}x + F_y \mathrm{d}y + F_z \mathrm{d}z), \tag{3.25}$$

so $V_1 = V_0 + W(0 \to 1)$, the PE at Point 1, will always contain a constant V_0, which can have any value whatever (an '*arbitrary*' value depending on where we choose to start from. You might ask what use is a definition like that – if you can never say what value the PE really has at any point in space! But the fact is that the only things we need are *differences* between the values of V at any two different points; and

$$V_2 - V_1 = (V_0 + W_{0\to 2}) - (V_0 + W_{0\to 1}) = W_{0\to 2} - W_{0\to 1},$$

where the arbitrary constant V_0 has disappeared.

Finally, suppose we carry the particle from Point 0 to Point 1 and then back again, from 1 to 0, along the same path – as indicated in Fig.24a. The whole change in PE will be zero (we're back at the starting point, as if we'd never set off), but it is the sum of two parts:

$$W(0 \to 1) = \int_0^1 (F_x \mathrm{d}x + F_y \mathrm{d}y + F_z \mathrm{d}z) \quad \text{(outward journey)}$$

and

$$W(1 \to 0) = \int_1^0 (F_x \mathrm{d}x + F_y \mathrm{d}y + F_z \mathrm{d}z) \quad \text{(outward journey)}.$$

Since the sum must be zero, the second path integral must be the negative of the first: in words, *changing the direction of the path changes the sign of the work done.*

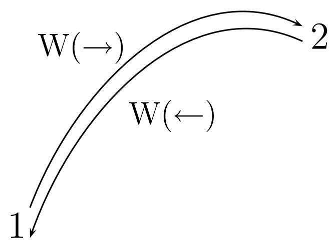


Figure 24a

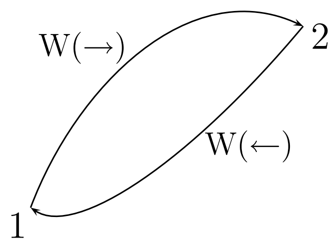


Figure 24b

Now it doesn't matter what names we give the two points: if we call them 1 and 2 we can say

$$\int_2^1 (F_x \mathrm{d}x + F_y \mathrm{d}y + F_z \mathrm{d}z) = -\int_1^2 (F_x \mathrm{d}x + F_y \mathrm{d}y + F_z \mathrm{d}z). \tag{3.26}$$

This is also a known result from calculus (Book 3): interchanging the upper and lower limits in a definite integral reverses the sign. But suppose now we make the return journey by a *different* route (as in Fig.24b). The work done, being independent of the path from Point 2 to Point 1, will still be the negative of the work done in the outward journey: but it now follows that the work done by the applied force in going round any closed path or 'circuit' is *zero*. Mathematicians often use a special symbol for this kind of path integral, writing it as

$$\oint (F_x \mathrm{d}x + F_y \mathrm{d}y + F_z \mathrm{d}z) = 0. \tag{3.27}$$

Forces that can be derived from a potential energy function, as in (3.24) are said to be "conservative". Now we have another definition: conservative forces are those for which the path integral of the work they do, taken round any closed circuit is zero.

As you must know, not all forces are conservative. If you slide an object over a rough surface it doesn't go easily, even if the surface is horizontal and the motion is not opposed by gravity: the motion is resisted by **friction** and the force arising from friction opposes any force you might apply. 'Push' or 'pull', the frictional force is always in the opposite direction; so however you go round a closed circuit, the work integral (3.27) must be non-zero – work is always done and you never get it back. Another example is the frictional force arising when a fast-moving object pushes its way through the air; the frictional force is always in a direction opposite to the direction of motion. There are other examples of non-conservative forces; but we won't meet them for a long time (Book 12). Until then, we'll usually be assuming that friction can be neglected, at least in a first approximation – which can later be improved by adding terms that will allow for it.

Exercises – in preparation.

Chapter 4

From one particle to many – the next big step

4.1 Many-particle systems

Suppose we have a collection of *many* particles, instead of just one. How will they move when forces act on them? This is an important question because we nearly always want to know about *big* systems, like the trucks in Fig.20 or the whole Earth, moving around the Sun (Fig.23); and even if they are small compared with the whole Universe we can hardly call them "particles". Yet we've treated them just as if they were single mass points, each body being at some *point* in space (with a position vector r) and having a certain *mass* (m). It seems like a miracle that everything came out right – that the Earth went round the Sun in about 360 days and so on – that Newton's second law worked so well. Now we want to know *why*.

So instead of asking how $m\mathsf{r}$ changes when a mass point is acted on by forces, let's ask the same question about a collection of mass points with mass m_1 at point r_1, m_2 at point r_2, and so on. The total mass of the whole collection, which we're going to think of as a single body, will then be

$$M = m_1 + m_2 + m_3 + \ ... = \Sigma_i m_i \tag{4.1}$$

where we use the usual shorthand notation Σ_i to mean the *sum* of all similar terms, with the index i taking values $1, 2, 3,$ ... for however many particles we have. And just as M will take the place of a single mass m, we'll *define* a quantity

$$M\mathsf{R} = m_1\mathsf{r}_1 + m_2\mathsf{r}_2 + m_3\mathsf{r}_3 + \ ... = \Sigma_i m_i \mathsf{r}_i \tag{4.2}$$

to take the place of $m\mathsf{r}$. We'll ask how this quantity changes when forces act on the system.

Suppose a force f_1 is applied to the mass m_1, a force f_2 to m_2, etc., these forces being 'external' to the system (e.g. forces due to gravity, or pushes and pulls applied 'by hand'). All this is shown in Fig.26(a), for a set of particles in a plane, but everything we do will apply just as well in three dimensions. Each particle will move, according to Newton's second law, m_1 starting with an acceleration a_1 such that $m_1\mathsf{a}_1 = \mathsf{f}_1$, and so on. In calculus notation this means

$$m_i\mathsf{a}_i = m_i\frac{\mathrm{d}^2\mathsf{r}_i}{\mathrm{d}t^2} = \mathsf{f}_i \quad \text{(for all } i\text{)}.$$

The quantity R, defined in (4.2) is the position vector of the **centroid**, or **centre of mass** of the system. It is

$$\mathsf{R} = \frac{m_1\mathsf{r}_1 + m_2\mathsf{r}_2 + m_3\mathsf{r}_3 + \ ...}{m_1 + m_2 + m_3 + \ ...} = \frac{\Sigma_i m_i \mathsf{r}_i}{\Sigma_i m_i}. \tag{4.3}$$

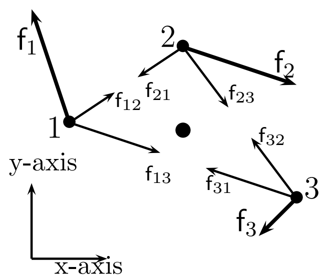


Figure 26a

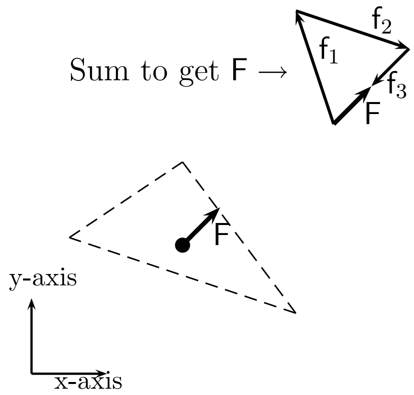


Figure 26b

As usual, if we want to use coordinates instead of vectors, we remember that a single vector equation corresponds to three equations for the separate x-,y- and z-components. If $\mathsf{r}_1 = x_1\mathsf{e}_1 + y_1\mathsf{e}_2 + z_1\mathsf{e}_3$ and $\mathsf{R} = X\mathsf{e}_1 + Y\mathsf{e}_2 + Z\mathsf{e}_3$, then the coordinates of the centroid will be

$$X = \frac{m_1x_1 + m_2x_2 + m_3x_3 + \ ...}{m_1 + m_2 + m_3 + \ ...} = \frac{\Sigma_i m_i x_i}{\Sigma_i m_i}, \tag{4.4}$$

with similar equations for Y and Z.

Now for the miracle! If you differentiate (4.3) twice, with respect to time, and remember that all the masses are simply numerical constants, you find that

$$\frac{\mathrm{d}^2\mathsf{R}}{\mathrm{d}t^2} = \frac{m_1}{M}\frac{\mathrm{d}^2\mathsf{r}_1}{\mathrm{d}t^2} + \frac{m_2}{M}\frac{\mathrm{d}^2\mathsf{r}_2}{\mathrm{d}t^2} + \ ... = \frac{1}{M}\mathsf{f}_1 + \frac{1}{M}\mathsf{f}_2 + \ ... \tag{4.5}$$

When you write F for the vector sum of all the forces acting on the particles of the whole system, and A for the sum of their separate acceleration vectors i.e.

$$\mathsf{F} = \mathsf{f}_1 + \mathsf{f}_2 + \ ... \ , \qquad \mathsf{A} = \mathsf{a}_1 + \mathsf{a}_2 + \ ... \ , \tag{4.6}$$

what do you get from (4.5)? It becomes simply, multiplying both sides of the equation by M,

$$\mathsf{F} = M\mathsf{A} = M\frac{\mathrm{d}^2\mathsf{R}}{\mathrm{d}t^2} \tag{4.7}$$

– force = mass × acceleration. But now the 'force' is something you *calculate*, as the vector sum of the forces acting on all the separate particles, and so is the acceleration – which refers to a point in space (the 'centre of mass') and not to the motion of any *real* particle. That is the miracle: Newton's second law tells us how the whole system would move if we could put all its mass at a point that we have *invented* and called the 'centroid' or 'mass centre'. But – you will say – it can't really be so easy. We've been talking about *independent* particles, each one of them feeling only its own 'external' force, like gravity or a push applied from outside.The particles inside any real object must also feel some kind of *internal* forces, which hold them all together, and we don't know anything about them. Or do we? It's here that Newton's *third* law comes to the rescue: for every action there is an exactly equal but opposite reaction. So if we put all those forces into the vector sum in (4.6), of all the forces acting on all the particles, *they must all cancel in pairs*! We don't have to worry about them. Fig.25a shows three particles acted on by three external forces $\mathsf{f}_1, \mathsf{f}_2, \mathsf{f}_3$, along with three *pairs* of internal forces – the force f_{12} which pulls Particle 1 towards Particle 2, the equal and opposite reaction f_{21} which pulls Particle 2 towards Particle 1, and so on. To get the sum of all the force vectors you have to put all the arrows head-to-tail, by shifting them but without changing their directions, and the resultant sum is then represented by the arrow that points from the first tail to the last head. The order in which you take the arrows doesn't matter (vector addition is 'commutative', as you will remember from Book 2) so you can follow each action with its reaction – to get a zero vector, which does not change the sum. In the end, only the external forces (shown as the thicker arrows, $\mathsf{f}_1, \mathsf{f}_2, \mathsf{f}_3$, in Fig.25a) contribute to the vector sum. They are equivalent to the single force $\mathsf{F} = \mathsf{f}_1 + \mathsf{f}_2 + \mathsf{f}_3$ shown in Fig.25b – and this is the force which, if applied to a mass $M = m_1 + m_2 + m_3$ sitting at the centroid $\bullet$, will tell us how the whole 3-particle system will move from one point in space to another, according to equation (4.7).

That is the second part of the miracle: a collection of particles, acted on by external forces, moves through space as if its particles were all squeezed together into a single point mass at the centroid – *even when the particles interact*. The interactions may be due to their gravitational attraction or to sticks or strings that fasten them together: it doesn't matter. That's why, in the last Section, we were able to treat our whole world – with all its mountains and seas, forests and cities (and you and me!) – like a single enormously heavy pebble travelling around the sun!

There's one thing we do need to worry about, however. We've been thinking about bodies moving from one place to another, all their parts moving in the same direction: this is called **translational motion**. But there can still be other kinds of motion: even if the vector sum of all the forces acting is zero and the centroid of the system is not moving, the forces may tend to *turn* the system around the centroid, producing **rotational motion**. In a later Section we'll find how to deal with rotations; but until then we'll think only of translational motion.

4.2 Conservation of linear momentum

In earlier Sections we found a principle of *energy* conservation, the total energy E (KE + PE) being conserved in *time*: in other words, E after an interval of time $(\Delta t) = E$ before (i.e. at $\Delta t = 0$, provided the forces acting were of a certain kind. Another important principle refers to **collisions** in which two or more particles may be involved: it states that the total momentum of the particles after a collision is the same as that before, the total momentum being

$$M\mathsf{V} = m_1\mathsf{v}_1 + m_2\mathsf{v}_2 + m_3\mathsf{v}_3 + \ ... = \Sigma_k m_k \mathsf{v}_k \qquad (4.8)$$

as follows from (4.2) on differentiating with respect to time and putting $d\mathsf{R}/\mathrm{d}t = \mathsf{V}$ and $\mathrm{d}\mathsf{r}_i/\mathrm{d}t = \mathsf{v}_i$.

When there are no external forces acting on the system of particles, $\mathsf{F} = \mathsf{0}$, and (4.7) tells us that $\mathsf{A} = \mathsf{0}$ and hence V in the last equation must be a constant vector. In this case the vector sum of the particle momenta in (4.8) must have exactly the same value before and after the collision:

$$m_1\mathsf{v}_{1i} + m_2\mathsf{v}_{2i} + m_3\mathsf{v}_{3i} + \ ... = m_1\mathsf{v}_{1f} + m_2\mathsf{v}_{2f} + m_3\mathsf{v}_{3f} + \ ... \ , \qquad (4.9)$$

where the labels 'i' and 'f' mean 'initial' and 'final' values (before and after).

The only collision we've studied so far was that between a bat and a ball in Section 2.4. There, the important thing was that the force involved was an **impulse**, creating almost at once a sudden change of momentum, and that other forces were so small they could be neglected. But this is generally true in collisions; the forces acting do so only at the moment of contact; they produce changes of *momentum*; and the conservation of momentum, expressed in (4.9), is the key equation to use.

Let's go back to a similar example, where a ball of mass m is struck by a hammer of mass M (as in the game of 'croquet', pictured in Fig.27(a).

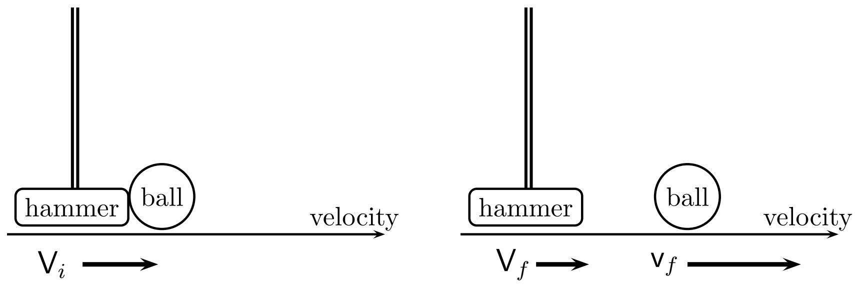


Figure 27a Figure 27b

In Section 2.4 the mass of the ball was taken as $m = 0.2$ kg and the blow was enough to give it a velocity of 10 ms^{-1}. The hammer was not considered; but we'll suppose it has a mass $M = 1$ kg and is travelling at 10 ms^{-1} when it hits the ball; and we'll take the left-right direction as positive for all velocities. We then have (still using subscripts i, f for 'initial' and 'final') :

Initially, total momentum $= MV_i + mv_i = (1\text{kg}) \times (10\ \text{ms}^{-1}) + 0 = 10\ \text{kg ms}^{-1}$

Finally, total momentum $= MV_f + mv_f = MV_f + (0.2\text{kg}) \times (10\ \text{ms}^{-1})$

On equating the two values of the total momentum we get $MV_f = 10\text{kg ms}^{-1} - 2\text{kg ms}^{-1} = 8\text{kg ms}^{-1}$, so, since $M = 1$ kg, the hammer velocity is reduced from 10 to 8 m s^{-1} (shown by the shorter arrow).

What about the kinetic energy, before and after the collision?

Initially, total KE $=\frac{1}{2}Mv_i^2 = \frac{1}{2}100\ \text{kg m}^2\text{s}^{-2} = 50\ \text{kg m}^2\text{s}^{-2}$

Finally, total KE $= \frac{1}{2}MV_f^2 + \frac{1}{2}mv_f^2 = \frac{1}{2}64\ \text{kg m}^2\text{s}^{-2} + \frac{1}{2}(0.2\text{kg}) \times (100\ \text{m}^2\text{s}^{-2})$.

So before the collision the KE is 50 kg m^{2}s^{-2}, or 50 Joules (with the named units first used in Section 2.1); but after collision the KE is reduced to 42 J. Where has the lost KE gone? Well, the forces acting in a collision don't *have to be* conservative: there is no potential energy function and no principle of conservation of the total energy. A collision usually makes a loud noise and it generates heat (the hammer and the ball can both get quite warm); and both are forms of energy – even if 8 Joules is hardly enough to heat a spoonful of water.

4.3 Elastic and inelastic collisions

Are there any kinds of collision in which no kinetic energy is lost? – at least in good approximation. An example will remind you that there are. You must at some time have bounced a rubber ball on a stone pavement. When you drop it, its downwards velocity increases (PE turning into KE) and when it hits the pavement it's going quite fast; then it bounces back, coming almost up to your hand; then down again and so on. If it came all the way back, the upward velocity after the bounce (collision) would be the same as the downward velocity when it hit the pavement. And you could say the collision was 'perfectly elastic' – there would be no loss of KE. Nothing is quite perfect, or the ball would go on bouncing forever! But the example gives us the idea: we *define* an **elastic collision** as one in which there is no loss of kinetic energy. And for such collisions we can use the principle of energy conservation in addition to that of momentum conservation.

On the other hand, if you try to bounce a lump of wet clay it just doesn't play! it simply says "shlop" and sticks to the surface. And if two lumps of wet clay collide they just become one; and you have an example of a perfectly **inelastic collision**.

To see how important the *kind* of collision can be, we can go back to the croquet hammer and ball. But let's not suppose the ball goes away with a velocity of 10 ms^{-1} (the velocity of the hammer when it struck the ball) – which was only a guess anyway. Conservation of momentum then requires that $mv_i + MV_i = mv_f + MV_f$ or

$$m(v_i - v_f) = M(V_f - V_i), \qquad (a)$$

in which both final values are now unknown. To get them we need a second equation (to solve for two unknowns we need two equations); so we try assuming the collision is perfectly elastic, which means the total kinetic energy will also be conserved. This gives us a second condition: $\frac{1}{2}m_1v_{1i}^2 + \frac{1}{2}m_2v_{2i}^2 = \frac{1}{2}m_1v_{1f}^2 + \frac{1}{2}m_2v_{2f}^2$. And this means

$$m(v_i^2 - v_f^2) = M(V_f^2 - V_i^2),$$

which can also be rearranged to give

$$m(v_i - v_f)(v_i + v_f) = M(V_f - V_i)(V_f + V_i). \qquad (b)$$

The initial velocities are given, $v_i = 0, V_i = 10$ ms^{-1}, and the two equations, (a) and (b), are now enough to give us both of the final velocities. Divide each side of equation (b) by the corresponding side of (a) (the two sides being in each case equal!) and you get

$$v_i + v_f = V_f + v_i \quad \text{or} \quad v_f - V_f = V_i - v_i. \qquad (c)$$

But (a) and (c) together make a pair of **simultaneous linear equations**: both must be satisfied at the same time and they are linear in the two unknowns, which we can call $x = v_f$ and $y = V_f$ – so as to see them more clearly. Thus, (c) in the first form can be written

$$y = v_i + x - V_i,$$

while (a) becomes

$$m(v_i - x) = M(y - V_i).$$

We can get rid of y in this last equation by substituting the value $y = v_i + x - V_i$ from the one before it. And then we only have to get out the x by untangling the messy thing that's left. That's a bit of simple algebra (see Book 1, Chapter 3) so you can do it yourself: you should get

$$x\ (= v_f) = \left(\frac{m - M}{m + M}\right) v_i + \left(\frac{2M}{m + M}\right) V_i. \qquad (4.10)$$

If you do it the other way round, substituting for x instead of y, you will find the solution for y:

$$y\ (= V_f) = \left(\frac{2m}{m + M}\right) v_i + \left(\frac{2M}{M - m}\right) V_i. \qquad (4.11)$$

On putting in the numerical values we now find $v_f = (50/3)$ ms^{-1} and $V_f = (20/3)$ ms^{-1}: so if the collision is perfectly elastic the ball takes *more* than the initial velocity of the hammer (indicated by the long arrow in Fig.27(b)). And we can be sure that the total kinetic energy will still be just what it was before the hit (because we made it so! – by supposing the collision to be elastic). You can check the numbers: you should find (250/9) J for the ball and (200/9) J for the hammer, giving altogether the 50 J before the ball was hit.

And what if the collision is perfectly *inelastic*? – if the hammer strikes a ball of wet clay. The two things, hammer and ball, then stick together and become one. To see what difference it makes, suppose the masses are the same as in the last example and that the hammer has the same initial velocity. In this case we have:

Initially, total momentum $= mv_i + MV_i = (1\text{ kg}) \times (10\text{ ms}^{-1})$

Finally, total momentum = $(m+M)V_f = (1.2\text{kg}) \times V_f$,

where V_f is the final velocity of hammer plus clay, moving as one, and is the only unknown. There must be no change of total momentum, so

$$V_f = \frac{(1\text{ kg}) \times (10\text{ ms}^{-1})}{1.2\text{ kg}} = (25/3)\text{ms}^{-1}$$

As for the kinetic energy, it started with the value 50 J but is now $\frac{1}{2}(m+M)V_f^2 = \frac{1}{2}$ 1.2 kg $\times$ $(25/3)^2$ $(\text{ms}^{-1})^2$ = (125/3) kg m^2s^{-2} = 41.667 J – so more than 8 J of the initial KE is lost, without any useful result (the ball is still sticking to the hammer).

For the present, that's all you need to know about the conservation of momentum; but, remember, we're talking about *linear* momentum and motion in a straight line. Sometimes we'll need to talk about the momentum of, say, a wheel, spinning around an axis. That will be *angular* momentum and we'll begin to think about it in the next Chapter.

Exercises (in preparation)

Chapter 5

Rotational motion

5.1 Torque

Suppose we have a sytem of particles moving through space with constant velocity V (which may also be zero) and want to know what goes on *inside* the system. The vector V refers to the centre of mass, which moves according to (4.8) when external forces are applied to the system. (Note that capital letters, like $M, \mathsf{V}, \mathsf{F}$, will now be used as in Section 4.1 for quantities that refer to the whole system – not to a single particle.) When the vector sum of these forces is zero

$$\frac{\mathrm{d}}{\mathrm{d}t}(MV) = m_1\frac{\mathrm{d}^2\mathsf{r}_1}{\mathrm{d}t^2} + m_2\frac{\mathrm{d}^2\mathsf{r}_2}{\mathrm{d}t^2} + \ ... \ = 0 \tag{5.1}$$

but this does not mean that $m_1(\mathrm{d}^2\mathsf{r}_1/\mathrm{d}t^2) = 0$ etc. for each separate term in (5.1). It only means that one point R with coordinates X, Y, Z (given in (4.4) etc.) will move with constant (or zero) velocity. We can take it as a new origin and call it O. What else can happen? The system can *turn* around O, which from now on we'll think of as a fixed point.

Let us take two axes, e_1 and e_2 in the plane of r_1, so that (see Fig.28a)

$$\mathsf{r} = x\mathsf{e}_1 + y\mathsf{e}_2, \tag{5.2}$$

The vector r is the position vector of the point with coordinates (x, y) *relative to the centre of mass*, which we'll often call the 'CM'. And we'll suppose one of the particles, of mass m, is at point (x, y). When particle labels are needed they are sometimes put in the upper position, so they don't get mixed up with indices for different vectors or components; but for the moment let's just leave them out.

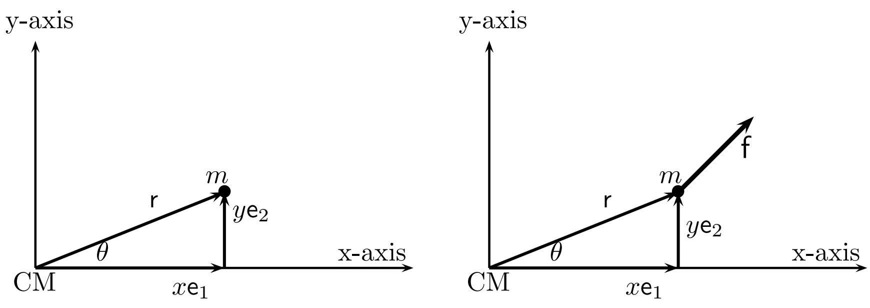


Figure 28a Figure 28b

A third axis, along the unit vector e_3, can be chosen using the 'corkscrew rule' (Book 2, Section 5.4): e_3 shows the direction in which a corkscrew would move in a turn that sends e_1 towards e_2.

When $m(\mathrm{d}^2\mathsf{r}/\mathrm{d}t^2) \neq 0$ it measures the force f acting on the mass m (Fig.28b), which is trying to move it so that its position vector r will turn around the axis e_3. You know quite a lot about rotations from Chapter 4 of Book 2. A **rotation** around e_3 turns a unit vector e_1 so that it will point along r in Fig.28b and is measured by the rotation angle θ, counted positive when e_1 turns towards e_2.

A turning force is called a **torque**. How do we measure a torque? Suppose you have to loosen a nut on a bolt that sticks upwards out of an iron plate – in the plane of the paper in Fig.29a.

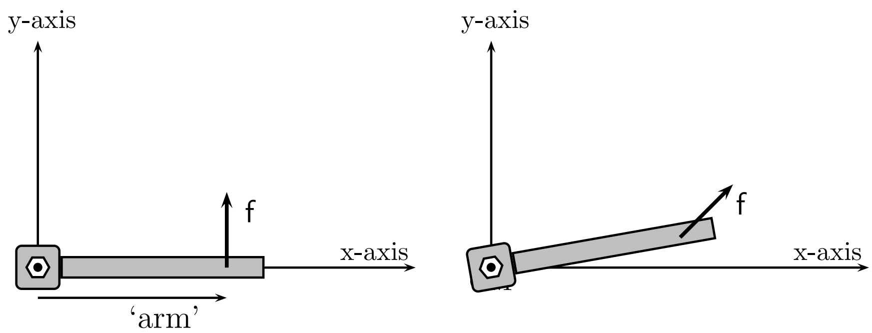


Figure 29a Figure 29b

You can use a 'key' or 'wrench', which fits over the nut and has a long handle, to which you can apply a force – as shown in the Figure. The key lies along the x-axis, while the nut you're trying to loosen is at the origin; and the force is in the direction of the y-axis.

The longer the 'arm' of the key, the greater the torque, and the easier it is to turn the nut; the arm 'magnifies' the turning effect of the force. When the force is perpendicular to the arm it has the greatest effect. So let's try a definition: the torque of a force around an axis is measured by the product

Torque = (force applied) × (perpendicular distance of its line of action from the axis)

How does this translate into symbols? If we use a for the length of the arm and f for the magnitude of the force, then the torque of f about the z-axis e_3 will be the product af.

We can't call it t, because t always stands for the *time* – and we find it everywhere. So let's use the corresponding Greek letter, τ ('tau'), and then we won't get mixed up. Now the arm a, in Fig.29a, is the x-coordinate of the point (x, y) at which the force is applied; f is the y-component (the only one) of the vector f; and τ is going to be a component of the torque *around the z-axis*. So let's add the labels and write

$$\tau_z = xf_y. \tag{5.3}$$

This component of the torque is also called the **moment** of the force about the z-axis. But to get a more general definition we have to look at the case shown in Fig.29b, where the key does not lie along one of the axes and the applied force is not perpendicular to it. Of course you can use the same definition in words, but then you have to work out the perpendicular distance from the origin to the line of action of the force. There's a simpler way, which is quite general and looks much nicer.

You know from Section 4.1 that any force can be expressed as the vector sum of two other forces acting at the same point: so f in Fig.29b is exactly equivalent to

$$\mathsf{f} = f_x\mathsf{e}_1 + f_y\mathsf{e}_2 = \mathsf{f}_x + \mathsf{f}_y, \tag{5.4}$$

where $\mathsf{f}_x, \mathsf{f}_y$ are the *vectors*, parallel to the two axes, with magnitudes f_x, f_y. The torque applied by the f_y, which is perpendicular to the x-axis, is xf_y – exactly as in (5.4) – and is in the positive (anticlockwise) sense. But the torque due to f_x, perpendicular to the y-axis, has an arm of length y and acts in the *negative* (clockwise) sense – giving a turning force $-yf_x$. The two forces together give the torque about the z-axis:

$$xf_y - yf_x = \tau_z, \tag{5.5}$$

which works for any directions of the key, and the force acting at point (x, y), in the xy-plane. Remember the order of the x,y,z and you can't go wrong: x (first term) turning towards y (second term) gives the z-component τ_z.

The beauty of this result is that it holds even in *three dimensions*! This must be so, because if you rotate the whole system about the origin – so that the axes x,y,z turn into new axes pointing along the y,z,x directions – everything will look just the same from inside the system. (This is what we called an "invariance principle" in Book 2: there's nothing special about different directions in space, so rotating everything will make no difference to our equations.) In this way you will find, instead of (5.5), three equations, which can be collected into

$$xf_y - yf_x = \tau_z, \qquad yf_z - zf_y = \tau_x, \qquad zf_x - xf_z = \tau_y. \tag{5.6}$$

It's enough to remember xyz → yzx → zxy, in 'rotating' the labels x,y,z. But remember also that the *order* matters; if you swap only x and y, for example, that doesn't correspond to a pure rotation of the axes in space, but rather to a reflection in which only two axes are interchanged. If you know about *vector products*, from Book 2 Sections 5.4 and 6.3, you may have noticed that (5.6) says that τ_x, τ_y, τ_z are the three components of a **vector product**:

$$\tau = \mathsf{r} \times \mathsf{f}. \tag{5.7}$$

This is not a 'true' vector. In addition to having an axis in space, it has a sense of rotation around the axis, like a screw: it is called a **pseudo-vector**, but here we use it only as a convenient notation for the the three equations (5.6).

5.2 Angular momentum and torque

Just as the torque is expressed as the *moment* of the force vector f around an axis through the origin, the moment of any other vector can be defined in a similar way. If a particle of mass m, at point (x, y), moves with velocity v and has linear momentum $\mathsf{p} = m\mathsf{v}$, then its **moment of momentum**, or **angular momentum** is defined as (again using the Greek letter λ instead of 'el' – which sometimes gets mixed up with '1')

$$\lambda = \mathsf{r} \times \mathsf{p}. \qquad (5.8)$$

In terms of components, this means

$$xp_y - yp_x = \lambda_z, \qquad yp_z - zp_y = \lambda_x, \qquad zp_x - xp_z = \lambda_y. \qquad (5.9)$$

We know that force produces linear momentum: can it be that torque produces angular momentum?

To answer this question we write Newton's second law in the component form

$$f_x = \frac{\mathrm{d}p_x}{\mathrm{d}t}, \qquad f_y = \frac{\mathrm{d}p_y}{\mathrm{d}t}, \qquad f_z = \frac{\mathrm{d}p_z}{\mathrm{d}t}, \qquad (5.10)$$

and ask if there is a parallel relation between components of the torque (τ_x, τ_y, τ_z) and the rates of change of the three angular momentum components $(\mathrm{d}\lambda_x/\mathrm{d}t,\ \mathrm{d}\lambda_y/\mathrm{d}t,\ \mathrm{d}\lambda_z/\mathrm{d}t.)$ In fact, we'd like to know if $\tau_x = \mathrm{d}\lambda_x/\mathrm{d}t$ and similarly for the other two components.

At first sight this doesn't look very promising, because the components of λ in (5.9) contain products of both coordinates and momentum components – and all of them depend on time. Differentiating might just give us a mess! But let's try it, differentiating λ_x with repect to the time t:

$$\frac{\mathrm{d}\lambda_x}{\mathrm{d}t} = \frac{\mathrm{d}}{\mathrm{d}t}(yp_z - zp_y) = y\frac{\mathrm{d}p_z}{\mathrm{d}t} + p_z\frac{\mathrm{d}y}{\mathrm{d}t} - z\frac{\mathrm{d}p_y}{\mathrm{d}t} + p_y\frac{\mathrm{d}z}{\mathrm{d}t}$$

But now remember that $p_z = mv_z = m(\mathrm{d}z/\mathrm{d}t)$ and $p_y = mv_y = m(\mathrm{d}y/\mathrm{d}t)$ and put these values in the line above. You'll get, re-arranging the terms,

$$\frac{\mathrm{d}\lambda_x}{\mathrm{d}t} = y\frac{\mathrm{d}p_z}{\mathrm{d}t} - z\frac{\mathrm{d}p_y}{\mathrm{d}t} - p_y\frac{\mathrm{d}z}{\mathrm{d}t} + p_z\frac{\mathrm{d}y}{\mathrm{d}t}.$$

The last two terms on the right are $-mv_yv_z$ and $+mv_zv_y$, respectively, and therefore cancel; while the first two terms together give $yf_z - zf_y = \tau_x$ – the middle equation in (5.6). So we've done it: the result we hoped to find, and two others like it, are

$$\tau_x = \frac{\mathrm{d}\lambda_x}{\mathrm{d}t}, \qquad \tau_y = \frac{\mathrm{d}\lambda_y}{\mathrm{d}t}, \qquad \tau_z = \frac{\mathrm{d}\lambda_z}{\mathrm{d}t}. \qquad (5.11)$$

These results are very similar to Newton's second law in the form (5.10): it's enough to change a *force* component, such as f_x, into a *torque* component (τ_x); and a *linear* momentum component, such as p_x, into an *angular* momentum component (λ_x) – and you get (5.11).

In Section 4.1, we extended Newton's second law to a whole system of particles, however many, and found the same law applied to a single *imaginary* particle of mass $M = m_1 + m_2 + \ ...$, located at a single *imaginary* point with position vector R, defined in (4.2). In fact, $\mathsf{F} = \mathrm{d}\mathsf{P}/\mathrm{d}t$ where P is the *total* linear momentum.

In these last two Sections, however, we've been thinking about rotational motion, in which a force is applied to one particle, at point (x, y), as it turns around an axis through the origin of coordinates. Instead of Newton's law for *translational* motion, we've now obtained equations (5.11) which describe *rotational*, or angular, motion: but they still apply only to a single point mass, such as a particle moving in an orbit. What we need now is a corresponding law for whole system of particles, possibly moving around is centre of mass – which may be at rest at point R, or may be travelling through space according to equation (4.7).

To get the more general equations all we have to do is add up all the equations for the single particles; and in doing this we remember that only the *external* torques need be included – because every action/reaction pair will consist of equal and opposite forces with the same line of action, giving the same equal and opposite moments about the origin and hence zero contribution to the total torque. Thus, the equation $(\mathrm{d}\lambda/\mathrm{d}t) = \tau$ becomes, on summing,

$$\mathrm{d}\mathsf{L}/\mathrm{d}t = \mathsf{T}, \tag{5.12}$$

where the capital letters stand for the sums over all particles (i) of single-particle contributions: $\mathsf{L} = \sum_i \lambda(i)$ and $\mathsf{T} = \sum_i \tau(i)$. The particle label (i) is shown in parentheses so that it doesn't get mixed up with the labels (x, y, z) for coordinate axes. When we remember that the angular momentum and torque are each 3-component quantities, and that each vector equation corresponds to three equations for the components, this all begins to look a bit messy. But equation (5.12) only says that

$$\frac{\mathrm{d}L_x}{\mathrm{d}t} = T_x, \qquad \frac{\mathrm{d}L_y}{\mathrm{d}t} = T_y, \qquad \frac{\mathrm{d}L_z}{\mathrm{d}t} = T_z, \tag{5.13}$$

where

$$T_x = \sum \tau_x, \qquad T_y = \sum \tau_y, \qquad T_z = \sum \tau_z, \tag{5.14}$$

are total torque components (leaving out the particle label i in the summations) and

$$L_x = \sum \lambda_x, \qquad L_y = \sum \lambda_y, \qquad L_z = \sum \lambda_z, \tag{5.15}$$

are total angular momentum components.

We know that if the vector sum (F) of all the external forces applied to a system of particles is zero the CM of the system will either remain at rest or will travel through space with constant velocity; but we have now found that in either case the system may still *turn* around the CM, provided the applied forces have a non-zero *torque*. The next great principle we need applies to this *rotational* motion: it says simply that if the total torque – or turning force – is zero then the system will either have no angular momentum or will go on rotating with *constant angular momentum*. In that case,

$$L_x = L_y = L_z = \text{ constant.} \tag{5.16}$$

In other words, there is a principle of **conservation of angular momentum**, which corresponds to that of *linear* momentum for a system on which no external *force* acts.

In the next Section we begin to see how important this principle can be.

5.3 Another look at the solar system

In Section 3.4, we were able to find how the Earth moves around the Sun – using nothing but simple arithmetic to get an approximate solution of the 'equation of motion', which followed from Newton's second law. But the results may have seemed a bit strange: for centuries people had believed that the path of the Sun, its orbit, was a circle; but our results gave an orbit which was *not quite* circular. And astronomers knew long ago, from their observations, that some of the other planets moved in orbits which were far from circular. Why this difference?

According to the principle of energy conservation, the sum of $\frac{1}{2}mv^2$ (the kinetic energy) and V (the potential energy) should give the constant total energy E. When the force of attraction towards the Sun is given by (1.2), as $F = -GmM/r^2$ (putting in the minus sign to show the force is in the negative direction along the vector r), the PE must be such that $F = -(\mathrm{d}V/\mathrm{d}r)$ – the rate of decrease of V with r. Thus V must satisfy the differential equation

$$\frac{\mathrm{d}V}{\mathrm{d}r} = \frac{GmM}{r^2}.$$

The solution is easy to see, because (Book 3, Chapter 3) the function $y = x^n$ has the derivative $(\mathrm{d}y/\mathrm{d}x) = nx^{n-1}$ and thus, for $y = r^{-1}$, it follows that $\mathrm{d}y/\mathrm{d}r = -r^{-2}$. This does not necessarily mean that $V = -GmM \times r^{-1}$ (which gives the right derivative), because you can add any *constant* c and still get the right function F when you form $(\mathrm{d}V/\mathrm{d}r)$. What it does mean is that

$$V = -\frac{GmM}{r} + c$$

is a general solution, whatever the value of c. To choose c we must agree on a 'zero of potential energy' – for what value of r shall we take $V = 0$? The usual convention is to count V as zero when the two masses, m and M, are an *infinite* distance apart: this gives at once $0 = 0 + c$, and so c must also be zero. The PE of a planet of mass m at a distance r from the Sun is therefore chosen as

$$V = -\frac{GmM}{r}. \tag{5.17}$$

The energy conservation equation now requires $\frac{1}{2}mv^2 - (GmM/r) =$ constant; but if the orbit is not exactly circular, so the distance of the Earth from the Sun is not fixed, the separate terms (KE and PE) cannot both be constant. When the Earth goes closer to the Sun (smaller r) its PE must become more negative and its KE, and velocity, must increase; and when it goes further away its PE will become less negative and it will travel more slowly.

To understand what's happening, we need to use the other great principle: the conservation of angular momentum, which applies to any orbital motion in the presence of a **central field** i.e. a force directed to one fixed point (in this case the Sun, taken as origin). The central force F has zero torque about the origin (see Fig.30, which lies on its line of action. (Note that F, shown as the bold arrow in the Figure, lies on top of the position vector r but has the opposite direction – towards the Sun.)

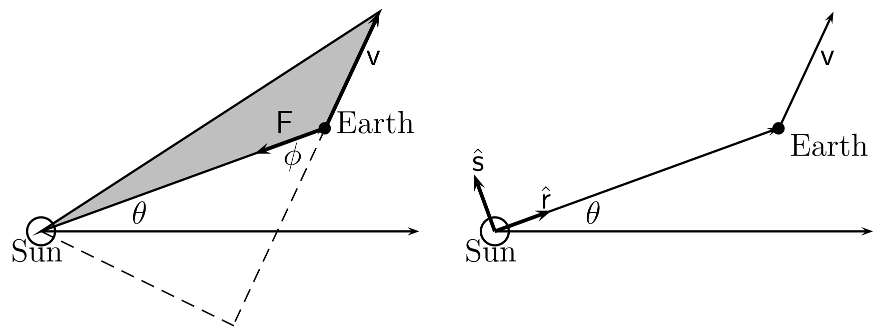


Figure 30 Figure 31

The Earth's angular momentum around the Sun (i.e. its *moment* of momentum) must therefore remain constant. Here we're thinking of the Earth as a single particle, but equation (4.23) applies for any number of particles – so we can keep the same notation even for a *one*-particle system, using the symbols L and T instead of λ and τ. The conservation principle for L may then be written, with the vector product notation used in (5.8),

$$\mathsf{L} = \mathsf{r} \times m\mathsf{v} = \text{constant} = mh\hat{\mathsf{n}} \tag{5.18}$$

– a vector of constant magnitude in a direction normal to the plane of the orbit, written as a *unit* vector $\hat{\mathsf{n}}$ multiplied by a constant h. In fact, h is the product of the magnitudes of the vectors r and v multiplied by the sine of the angle between them: $h = rv \sin\phi$, as shown in Fig.30

The constant h has a simple pictorial meaning. The base of the shaded triangle in the Figure has a length v, which is the distance moved by the Earth in unit time; and the height of the vertex (at the Sun) is the length of the perpendicular, $p = r\sin\phi$. The area of this triangle is thus Area=$\frac{1}{2}$ base $\times$ height (as we know from simple geometry, Book 2, Chapter 3). So $h = vp = rv\sin\phi$ is just twice the area of the shaded triangle – twice the area 'swept out' by the radius vector r, in unit time, as the Earth makes its journey round the Sun.

We say that h is twice the **areal velocity**; and what we have discovered is that the areal velocity is a **constant of the motion** for the Earth and all the other planets as they move around the Sun.

This important result was first stated four hundred years ago by Kepler, on the basis of astronomical observations, and is usually called "Kepler's second law". His "first law", published at the same time, stated that the orbit of any planet was an **ellipse**, not a circle, and his "third law" concerned the **period** of the planet – the time it takes to go round the Sun. Of course, Kepler didn't have Newton's laws to guide him, so his discoveries were purely *experimental*. Much later, in fact, Newton used Kepler's observations to show that the force of attraction between two bodies must be given by an equation of the form (1.2), an inverse square law. The interplay of experiment and theory is what leads to continuous progress in Physics and makes it so exciting; you never know what's coming when you turn the next corner!

5.4 Kepler's laws

From the two principles we have – conservation of energy and conservation of angular momentum – we can now get all the rest! First we note that the velocity v, being a vector, can be written as the sum of two perpendicular components. There will be a *radial* component, along the direction of the unit vector $\hat{\mathsf{r}}$ (shown in Fig.31), and a *transverse* component, in the transverse direction $\hat{\mathsf{s}}$. As r sweeps out the shaded area in Fig.30, both $\hat{\mathsf{r}}$ and $\hat{\mathsf{s}}$ will change. In a small increase of the angle θ, call it $\mathrm{d}\theta$, the tip of the (unit-length) arrow representing $\hat{\mathsf{r}}$ will move through $\mathrm{d}\theta$ *in the transverse direction*. And if this change takes place in a small time interval $\mathrm{d}t$ the *rate* of change of $\hat{\mathsf{r}}$, as it rotates, will thus be $(\mathrm{d}\theta/\mathrm{d}t)\hat{\mathsf{s}}$. The rate of increase of the angle θ, with time, is the modulus of the **angular velocity** and the corresponding increase in the unit vector $\hat{\mathsf{r}}$ is written $\mathrm{d}\hat{\mathsf{r}}/\mathrm{d}t = (\mathrm{d}\theta/\mathrm{d}t)\hat{\mathsf{s}}$. If you next think about the way $\hat{\mathsf{s}}$ changes (look again at Fig.31), you'll see the length of the unit vector changes by the same amount – but in the negative direction of r! These two important results together can now be written, denoting the angular velocity by $\mathrm{d}\theta/\mathrm{d}t = \omega$,

$$\frac{\mathrm{d}\hat{\mathsf{r}}}{\mathrm{d}t} = \omega\hat{\mathsf{s}}, \qquad \frac{\mathrm{d}\hat{\mathsf{s}}}{\mathrm{d}t} = -\omega\hat{\mathsf{r}}. \tag{5.19}$$

Notice that the two rates of change are written with the notation of the usual **differential calculus**; but each is the derivative of a *vector*, with repect to time, and although the time is a scalar quantity (measured by a single number t) the vectors $\hat{\mathsf{r}}$ and $\hat{\mathsf{s}}$ are not. The vectors change with time and are said to be "functions of a scalar parameter" t. But the definition of the time derivative of a vector is parallel to that of any scalar quantity $y = f(t)$: just as $(\mathrm{d}y/\mathrm{d}t)$ is the limiting value of the ratio $\delta y/\delta t$, when the changes are taken smaller and smaller, so is $\mathrm{d}\mathsf{v}/\mathrm{d}t$ the limiting value of the change $\delta\mathsf{v}$ divided by the number δt. Notice also that the magnitude of the angular momentum vector, $|\mathsf{L}| = h$ in (5.18) is simply a multiple of the angular velocity ω:

$$\mathsf{L} = \mathsf{r} \times m\mathsf{v} = h\hat{\mathsf{n}}, \qquad |\mathsf{L}| = h = mrv = mr(r\omega) = mr^2\omega. \tag{5.20}$$

We're now ready to find the two missing laws.

Kepler's first law

The *first* law states that the orbits of all the planets in the solar system are ellipses. An ellipse is shown in Fig.32: it is the figure you get if you knock two pegs into the ground and walk around them with a 'marker', tied to a long loop of string passing over the pegs and kept tightly stretched, as in the Figure. If O_1 and O_2 are the positions of the pegs, and M is the position of the marker, then the loop of string (constant in length) makes the triangle O_1MO_2. The points O_1 and O_2 are called the **foci** of the ellipse; and Kepler noted that, for all the planets, the Sun was always to be found at one of the foci. If the two foci come together, to make a single focus, the orbit becomes a circle.

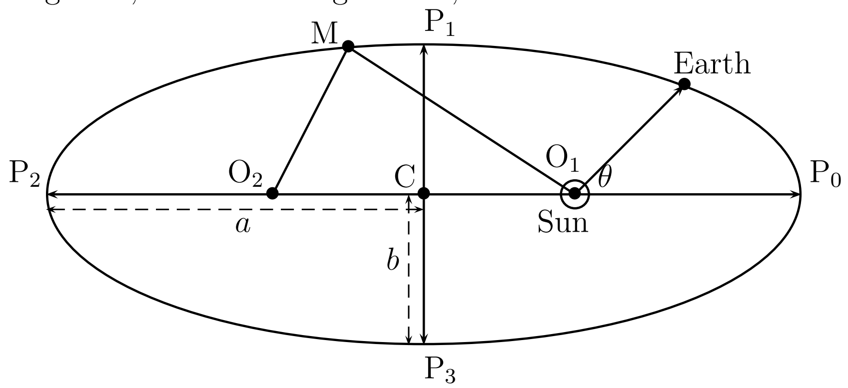


Figure 32

(**Note**. The proof that follows is quite difficult! Don't worry about the details – you can come back to them when you're ready – but look at the equation (5.25), which will tell you how to *calculate* the ellipse.)

To show that the orbit is, in general, an ellipse we start from the fact that the angular momentum (5.20) is a constant of the motion, while the Earth still moves according to Newton's second law $m\mathsf{a} = \mathsf{F}$. The acceleration is the rate of change of the velocity vector with time; and such a rate of change is often indicated just by putting a dot over the symbol for the vector. With this shorthand, $\mathsf{a} = \dot{\mathsf{v}}$ and similarly $\dot{\mathsf{r}}$ will mean $\mathrm{d}\mathsf{r}/\mathrm{d}t$. In the same way, two dots will mean differentiate twice: so $\mathsf{a} = \dot{\mathsf{v}} = \mathrm{d}\dot{\mathsf{r}} = \mathrm{d}^2\mathsf{r}/\mathrm{d}t^2$.

From (5.19) the velocity vector can be written in terms of its radial and transverse components, in the directions of unit vectors $\hat{\mathsf{r}}$ and $\hat{\mathsf{s}}$. Thus

$$\mathsf{v} = \frac{d}{\mathrm{d}t}(r\hat{\mathsf{r}}) = \frac{\mathrm{d}r}{\mathrm{d}t}\hat{\mathsf{r}} + r\frac{\mathrm{d}\hat{\mathsf{r}}}{\mathrm{d}t} = \dot{r}\hat{\mathsf{r}} + r\dot{\theta}\hat{\mathsf{s}}. \tag{5.21}$$

$$\mathsf{a} = \dot{\mathsf{v}} = \left(\frac{\mathrm{d}^2 r}{\mathrm{d}t^2}\hat{\mathsf{r}}\right) + \left(\frac{\mathrm{d}r}{\mathrm{d}t}\omega\hat{\mathsf{s}}\right) + \left(\frac{\mathrm{d}(r\omega)}{\mathrm{d}t}\right)\hat{\mathsf{s}} + (r\omega)(-\omega\hat{\mathsf{r}}). \tag{5.22}$$

At the same time, by (1.2), the force vector is $\mathsf{F} = -(GmM/r^2)\hat{\mathsf{r}}$.

Newton's second law (mass × acceleration = force) then equates two vectors, $m\mathsf{a}$ and F, both lying in the plane of the orbit – whose normal is the constant vector $\mathsf{L} = h\hat{\mathsf{n}}$, according to (5.20). Now let's take the *vector product* of both sides of the equation with $\mathsf{L} = h\hat{\mathsf{n}}$. Why? Because we know the motion must be in the plane and the vectors must therefore have no normal components: taking the vector product of the equation with $\hat{\mathsf{n}}$

will simply 'kill' any components *normal* to the plane because the vector product of a vector with itself is zero! The *in-plane* components, which we want, will be obtained by solving the equation that remains; and this becomes $m\mathsf{a} \times h\hat{\mathsf{n}} = -(GmM/r^2)\hat{\mathsf{r}} \times h\hat{\mathsf{n}}$. But

$$m\mathsf{a} \times h\hat{\mathsf{n}} = m\frac{d}{\mathrm{d}t}(\mathsf{v} \times h\hat{\mathsf{n}})$$

while, using $\mathsf{L}(= \mathsf{r} \times \mathsf{v})$ in place of $h\hat{\mathsf{n}}$,

$$-(GmM/r^2)\hat{\mathsf{r}} \times h\hat{\mathsf{n}} = -(GmM/r^2)\hat{\mathsf{r}} \times (r^2\omega)\hat{\mathsf{n}} = -(GmM)\omega\hat{\mathsf{s}},$$

since the three unit vectors $\hat{\mathsf{r}}$, $\hat{\mathsf{s}}$ and $\hat{\mathsf{n}}$ form a right-handed basis with $\hat{\mathsf{r}} \times \hat{\mathsf{s}} = \hat{\mathsf{n}}, \hat{\mathsf{s}} \times \hat{\mathsf{n}} = \hat{\mathsf{r}}, \hat{\mathsf{n}} \times \hat{\mathsf{r}} = \hat{\mathsf{s}}$. (Look back at Book 2, Section 5.4, if you're not sure about vector products.)

On taking away a common factor m, and remembering that $\mathrm{d}\hat{\mathsf{r}}/\mathrm{d}t = \omega\hat{\mathsf{s}}$, the equation of motion can be written

$$\frac{\mathrm{d}}{\mathrm{d}t}(\mathsf{v} \times h\hat{\mathsf{n}}) = GM\frac{\mathrm{d}\hat{\mathsf{r}}}{\mathrm{d}t} = \frac{\mathrm{d}}{\mathrm{d}t}(GM\hat{\mathsf{r}}).$$

But if the derivatives of two vector quantities are equal the quantities themselves can differ only by a *constant vector*, call it $e\hat{\mathsf{a}}$ – a numerical multiple of the unit vector $\hat{\mathsf{a}}$, which fixes a direction in space. We can therefore write

$$(\mathsf{v} \times h\hat{\mathsf{n}})/(GM) = \hat{\mathsf{r}} + e\hat{\mathsf{a}}. \tag{5.23}$$

Now let θ be the angle between the position vector r and the constant vector a; and take the scalar product of the last equation with r. The result will be, putting the right-hand side first,

$$r + er\cos\theta = \mathsf{r} \cdot (\mathsf{v} \times h\dot{\mathsf{n}})/(GM). \tag{5.24}$$

The final step depends on the result (Book 2, Section 6.4) that in a 'triple-product', like that on the right, the order of the 'dot' and the 'cross' can be interchanged; so the last equation can be re-written in the standard form

$$r(1 + e\cos\theta) = (\mathsf{r} \times \mathsf{v}) \cdot h \times \mathsf{n})/(GM) = h^2/GM. \tag{5.25}$$

There are two numerical parameters in this equation: e is called the **eccentricity** and determines whether the ellipse (Fig.32) is long and thin, or shorter and 'fatter'; the other, h^2/GM, gives half the 'length' of the ellipse, the value of a in the Figure. Suppose we are told the values of these parameters. Then any pair of values of the *variables* r and θ that satisfy equation (5.25) will fix a point on the ellipse (see Exercise xxx). By starting with $\theta = 0$ and increasing its value in steps of, say, 45 degrees, calculating corresponding values of r from (5.25), you'll find a series of points $\mathrm{P}(r,\theta)$ (like P_0, P_1, P_2, P_3 in Fig.32) which fall on the ellipse.

We can find the value of a (which is called the "length of the **semi-major axis**) from the values of r at points P_0 and P_2: they are easily seen to be $r_0 = l/(1+e)$ and $r_2 = l/(1-e)$, from which it follows that $a = l/(1-e^2)$. (Check it yourself!) With a bit more geometry (see Exercise xxx) you can find the length of the other axis: b in Fig.32 is the **semi-minor**

axis. To summarize, the length and width of the ellipse are determined as $2a$ and $2b$, with

$$a = \frac{l}{1-e^2}, \qquad b = \frac{l}{\sqrt{1-e^2}}. \tag{5.26}$$

Knowing all this about the ellipse, we can come back to the last of Kepler's famous laws.

Kepler's third law

The third law answers the question: How long does it take a planet to complete its journey round the Sun? This time (T) is called the **period**: the period of the Earth is about 365 days, while that of the moon as it goes around the Earth is about 28 days. The orbits of the planets are all ellipses, even though their masses (m) may be very different, because equation (5.25) does not contain m: they differ only in having different values of the parameters $l = h/GM$ and e – and these are fixed once we know one point P on the orbit, along with the corresponding velocity vector. We take the parameter values as 'given' because they must have been determined millions of years ago, when the solar system was being formed, and they change very very slowly as time passes. What Kepler wanted to do was to find a rule relating T to the form of the orbit. And, as a result of careful measurements, he found that, for all the known planets, the *square of the period is proportional to the cube of the long axis of the orbit*; in other words $T^2 \propto a^3$.

Now that we know the forms of the orbits, we should be able to *prove* that Kepler's third law will correctly describe how the length of the year, for any of the planets, varies with the half-length (a) of the orbit.

First let's remember that, from (5.18) and the paragraph that follows it, that h is twice the areal velocity of the planet (the shaded area in Fig.30, which is swept out in unit time by the vector r): so if we know the area of the whole ellipse we can simply divide it by $\frac{1}{2}h$ and that will give us the number of time units taken to sweep over the whole area.

You may think the area of an ellipse will be hard to find: but it's not. You known from Book 2 that the area of a *circle* is πr^2; and an ellipse is only a 'squashed' circle. Think of a circle of radius a and imagine it cut into thin horizontal strips of width d, so that $n \times d = a$. To squash the circle you simply squash every strip, so that the width is reduced to d', without changing the number of strips or their lengths. When $nd' = b$ you'll have an ellipse of half-length a and half-width b, as in Fig.32. And because every strip has been reduced in area (i.e. width×length) by a factor b/a, the same factor will give the change in the whole area: the area πa^2 for the circle will become $\pi a^2 \times (b/a)$ for the ellipse. Thus,

$$\text{area of an ellipse} = \pi ab \quad (a = \text{halflength},\ b = \text{halfwidth}). \tag{5.27}$$

On dividing the whole area of the orbit by the amount swept over in unit time ($\frac{1}{2}h$) we get the number of time units needed to complete the whole orbit:

$$T = \frac{\pi ab}{\frac{1}{2}h} = \frac{2\pi ab}{\sqrt{GMl}},$$

where we have used the definition of the parameter l, namely $l = h^2/(GM)$, just after equation (5.25).

That's all right: but we can't get l just by looking at the sky! On the other hand we *can* observe the length and width of an orbit; and from (5.26) we can get l in terms of a and b. Thus

$$\frac{b^2}{a} = \left(\frac{l^2}{1-e^2}\right)\left(\frac{1-e^2}{l}\right) = l. \tag{5.28}$$

And now, by substituting this value in the equation for the period T, we find

$$T = \frac{2\pi ab}{\sqrt{GM}}\frac{\sqrt{a}}{b} = \left(\frac{2\pi}{GM}\right)a^{3/2}. \tag{5.29}$$

This is Kepler's third law. In words, it states that the square of the period of a planet is proportional to the cube of the half-length of its orbit. But now this result has been *proved*, from Newton's laws, we have obtained the actual value of the proportionality constant: it is $2\pi/(GM)$ and depends only on the mass of the Sun, which is the same for all planets in the solar system, and the gravitational constant G, which we can find for ourselves by measuring the force of attraction between two heavy bodies in the laboratory – here on Earth! Another thing – our Sun has been in the sky for a long time (about 5000 million years is the estimated age of the solar system) and the Sun's mass M is slowly changing, because it burns up fuel in producing sunshine. As M gets smaller, the orbit's half-length a gets bigger and so does the period T: the planet spends longer far away from the Sun and takes longer to go round it. And it's the same for all the planets: it may be another 5000 million years before the Sun uses up all its fuel and the solar system dies – but, when it does, we'll all go together!

Exercises (in preparation)

Chapter 6

Dynamics and statics of rigid bodies

6.1 What is a rigid body?

We've spent a long time in space – thinking about the planets moving around the Sun, as if each one was a single particle, moving independently of the others. Now we'll come back to Earth and to systems of *many* particles, not moving freely but all joined together, somehow, to make a *rigid structure*. A simple example was studied in Section 4.1, where Fig.26a showed three massive particles joined by light sticks (so light that we could forget they had any weight): the sticks are needed only to keep the distances between all pairs of particles *fixed*. The fact that the distances between all the mass points stay fixed, even when they may be moving, is what makes the structure *rigid*. The structure shown in Fig.26a lies in a plane – it is flat – but all our arguments about internal and external forces, actions and reactions, and so on are unchanged if there are *many* particles and they are not all in one plane. All we must do is use three coordinates for every mass point and three components for every vector (force, velocity, etc.). So what we did in Section 4.1 was very general.

This model of a rigid structure can be extended to rigid *bodies* in which there may be millions of particles – all so close together that no space (almost) is left between them. If you cut out, from a flat sheet of metal, a shape like that in Fig.26a, you will get a plane triangular **lamina**, or plate: think of this as made up from an enormous number of tiny bits of metal, all joined together so that the distance between any two bits doesn't change as the lamina moves. You then have a **rigid body** in the form of a lamina. And we don't have to ask about *how* the bits are joined together because (as we saw) Newton's third law tells us that actions and reactions always come in equal pairs – and that's enough! If we number all the particles and suppose there is a bit of stuff with mass m_1 at the point with coordinates (x_1, y_1), one with mass m_2 at (x_2, y_2), and so on, then we can define things like the coordinates of the centre of mass just as we did in Section 4.1. The total mass of the body (M, say) will be given by (4.1) and the position vector (R) of the centre of mass will follow from (4.3). Let's repeat the equations:

$$M = m_1 + m_2 + m_3 + ... \tag{6.1}$$

is the mass of the whole body, while the centre of mass, or centroid, is at the point

$$\mathsf{R} = \frac{m_1\mathsf{r}_1 + m_2\mathsf{r}_2 + m_3\mathsf{r}_3 + \ ...}{m_1 + m_2 + m_3 + \ ...} = \frac{\Sigma_i m_i \mathsf{r}_i}{\Sigma_i m_i}. \tag{6.2}$$

The *coordinates* (X, Y, Z) of the centroid (components of R) are given by

$$X = \frac{m_1x_1 + m_2x_2 + m_3x_3 + \ ...}{m_1 + m_2 + m_3 + \ ...} = \frac{\Sigma_i m_i x_i}{\Sigma_i m_i}, \tag{6.3}$$

with similar equations for Y and Z.

In a *real* 'rigid body', made out of some continuous material like metal or hard plastic, there will be too many particles to count – an infinite number. But we can imagine the body cut into small pieces, each one being given a number, and go ahead in the same way using this 'model' of the continuous body. If you've studied the Calculus (in Book 3) you'll be able to guess how the equations need to be changed. Suppose, for example, you have a long iron bar and want to find its CM. Measure distances along the x-axis, so that the ends of the bar are at $x = 0$ and $x = L$; and suppose every unit of length has a mass m_d, which is called the **mass-density** – just the mass per unit length. Then a piece of the bar between points x and $x + \mathrm{d}x$, with length $\mathrm{d}x$, will have a mass $m_d(x)\mathrm{d}x$, where in general m_d may depend on position and so is written as a function of x. The total mass of the bar will be the sum of the masses of all the bits, in the limit where the bits get smaller and smaller but you have infinitely many of them. This is something you've met in calculus; and the limit of the sum in an equation like $M = \sum_i m_i$ is written

$$M = \int_0^L m_d(x)\mathrm{d}x,$$

which is a **definite integral** taken between $x = 0$ and $x = L$, the two ends of the bar. In the Exercises at the end of the Chapter, we'll see how such integrals can be evaluated, but for the moment we'll just suppose it can be done.

6.2 Rigid bodies in motion. Dynamics

To get some 'feeling' for what happens when a rigid object moves through space, we can start with a very simple system, thinking first of just two masses (m_1, m_2) at the two ends of a light stick (Fig.33a). We'll call it a "stick-object". In this case, measuring distances (x) from the position of mass m_1, so that $x_1 = 0, x_2 = l$ (the length of the stick), the centroid will have x-coordinate

$$x_c = (m_1 \times 0 + m_2 \times l)/(m_1 + m_2) = \frac{m_2 l}{m_1 + m_2}$$

– where we use x_c instead of X, because this is the distance from m_1, a point fixed in the body, not one of the coordinates (X, Y, Z) of the centroid as it moves through space. The centroid is indicated in Fig.33a by the 'bullet' •, for the case where the first mass is twice as heavy as the second: $m_1/m_2 = 2$, which gives $x_c = l/3$ – one third of the way along the stick.

If you now throw the stick-object into the air it will move as if all its mass is concentrated at the centroid, whose coordinates X, Y, Z will change with time. But as the centroid moves the stick will also *rotate*. As soon as you let go of the stick it will move under the influence of gravity; and, as we know form Section 3.2, the centroid (moving like a single particle) will follow a parabola. Fig.33b shows where the centroid has got to after some time t; and also shows how the stick might have rotated during that time.

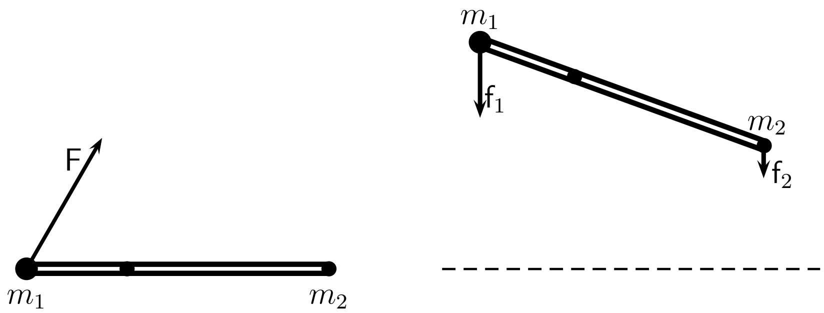


Figure 33a Figure 33b

What can we say about the rotational motion? The details will depend on *how* the object is thrown, on the force we apply before letting go. And, to be simple, we'll suppose the force (F in Fig.33a), which is an *impulsive* force, is applied to m_1 in the vertical plane – so that the rotating stick always stays in the vertical plane and we need only think about x- and y-components of the forces acting. The stick, after it leaves your hand, will go on rotating about a horizontal axis through the CM; and the only forces acting on it will be f_1, f_2, as shown in Fig.33b. How will they change its rotational motion?

In Section 4.4 we discovered a very general principle: that angular momentum is conserved, according to (5.12), when no torque is acting on a system; so we need to know how much torque (if any) is produced by the forces $\mathsf{f}_1, \mathsf{f}_2$. Remember we're thinking of the torque (the moment of the forces) *around the centre of mass* and that each moment can be written nicely as a *vector product* (5.7). Thus

$$\mathsf{T} = \tau_1 + \tau_2 = \mathsf{r}_1 \times \mathsf{f}_1 + \mathsf{r}_2 \times \mathsf{f}_2, \tag{6.4}$$

where all distances are measured from the CM; so r_i is the position vector of mass m_i *relative to the mass centre*. If we use this reference frame to calculate the position of the centroid, using (4.2), it must of course come out as $\mathsf{R} = \mathsf{0}$ because we're there already! In other words,

$$(\sum_i m_i)\mathsf{R} = \sum_i m_i \mathsf{r}_i = \mathsf{0}.$$

Now look at the total torque in (6.4), putting $\mathsf{f}_i = m_i g \hat{\mathsf{f}}$, where $\hat{\mathsf{f}}$ is a unit vector pointing vertically downwards: it can be written

$$\mathsf{T} = \sum_i \tau_i = \sum_i \mathsf{r}_i \times (m_i g)\hat{\mathsf{f}} = \left(\sum_i m_i \mathsf{r}_i\right) \times g\mathsf{f} = \mathsf{0}, \tag{6.5}$$

– since we have just seen that $\sum_i m_i\mathsf{r}_i = \mathsf{0}$, and if a vector is zero then so is its product with any other vector.

We've shown that the torque around the centroid of any object, due to gravity, is zero – that there is no resultant turning force that would produce a rotation. This is one of the most important properties of the centre of mass and is the reason why the CM or centroid used to be called the "centre of gravity". We know from Chapter 4, Exercise (xxx) that the CM of a uniform bar is at its midpoint, so if we hang it from that point there will be no turning force to make it tip over to one side or the other. Similarly If you support the stick-object in Fig.33a at one point only, directly below the CM (•), it will stay horizontal as long as $\mathsf{F} = \mathsf{0}$ even though one of the weights at its two ends is twice as big as the other; and we say it is "in balance". There's more about the principle of the balance in the next Chapter, where we talk about making weighing machines.

But now we're talking about the *motion* of the stick-object when it is thrown in the air. When it is in free flight (Fig.33b) there is no applied torque and according to (5.12) the angular momentum L must stay constant with the value you gave it before letting go. So the stick-object will go on rotating, around a horizontal axis through the CM, as it makes its journey through space. That's all supposing you threw the stick 'straight', so the motion started off in the vertical plane, with the axis horizontal: otherwise the stick would 'wobble', with the rotation axis continually changing direction, and that gives you a much more difficult problem. So for the rest of this Section we'll think only about rotation of a body around an axis in a fixed direction.

Motion around a fixed axis is very important in many kinds of machinery and it's fairly easy to deal with. But it still seems that angular motion is very different from motion through space. We know that Newton's second law applies directly to a rigid body, the total applied force F giving a rate of change of linear momentum P according to (4.7), namely $\mathsf{F} = (\mathrm{d}\mathsf{P}/\mathrm{d}t) = M(\mathrm{d}\mathsf{V}/\mathrm{d}t)$ – just as if the total mass was all at the CM and moved with linear velocity V. But for the rotational motion we have instead $\mathsf{T} = (\mathrm{d}\mathsf{L}/\mathrm{d}t)$ and there seems to be no similar last step, relating L to the *angular* velocity ω. Wouldn't it be nice if we could find something like (5.12) in the form

$$\mathsf{T} = \frac{\mathrm{d}\mathsf{L}}{\mathrm{d}t} = I\frac{\mathrm{d}\omega}{\mathrm{d}t},$$

where ω is the angular velocity produced by the torque L and I is a new proportionality constant? To show that this is possible we only need to express L in terms of the angular momentum vectors for all the separate particles that make up the body.

Remember that L is a vector (of a special kind) and is simply a sum of one-particle terms λ_i, given in (5.8) with components in (5.9). We want to find L in the easiest possible way, so let's take the axis of rotation as the z-axis and evaluate λ_z, given in (5.9), for the particle with mass m_i. But haven't we done this already? In talking about the Earth (with mass m) going around the Sun we wrote the angular momentum as

$$\mathsf{r} \times \mathsf{p} = \mathsf{r} \times (m\mathsf{v}) = mrv\sin\phi$$

where r was the distance from the axis (through the Sun and normal to the plane of the orbit) to the point with coordinates x, y; and ϕ was the angle between the vectors r and

v. It's just the same here: the z-component of λ will be $\lambda_z = mrv$ because this is a *rigid* body, with $r = \sqrt{x^2 + y^2}$ (fixed) and v perpendicular to the position vector r. We can also introduce the angular velocity, as we did in (5.20), because $v = r\omega$ and therefore

$$\lambda_z = mrv = mr(r\omega) = m(x^2 + y^2)\omega. \tag{6.6}$$

To get the total angular momentum we simply take one such term for every particle (of mass m_i at point (x_i, y_i) – z_i not appearing) and add them all together. The total angular momentum around the axis of rotation is then

$$L_z = \sum \lambda_z = \sum_i m_i(x_i^2 + y_i^2)\omega = I_z\omega, \tag{6.7}$$

where I_z is property of the rotating body, evaluated for a given axis (in this case called the z-axis) from the formula

$$I_z = \sum_i m_i(x_i^2 + y_i^2) = \sum_i m_i r_i^2, \tag{6.8}$$

where r_i^2 is simply the square of the distance of the mass m_i from the given rotation axis.

Note that the moment of inertia of an object is not determined once and for all time – like the mass – just by weighing it: it depends on the shape of the object, the masses of all its parts and where they are placed, and on what axis you choose. If you have an egg-shaped object, for example, there will be different moments of inertia for spinning it around its long axis and a short, transverse, axis. For a three-dimensional object there are three of them, all calculated in the same way, called **principal moments of inertia**; and often the principal *axes* are 'axes of symmetry' (see Chapter 6 of Book 1) around which you can rotate the object without making any change in the way it looks (as in the case of the egg – where any transverse axis gives the same value of the moment, so there are only two *different* principal values). All this will be clear when you try to calculate a moment of inertia, as in some of the Exercises at the end of the Chapter. But let's start things off with an example –

An Example A bicycle wheel spinning around its axis

Put the wheel in a horizontal plane, with its axis vertical. From above it will look like Fig.34a, below; or, if you add masses (•) on the rim, like 34b.

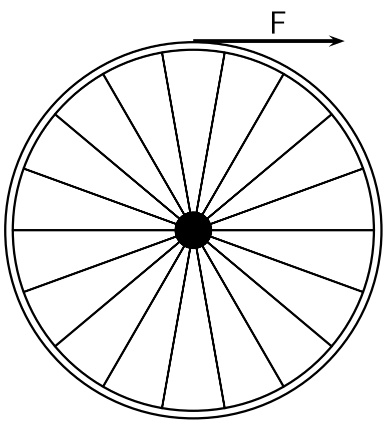


Figure 34a

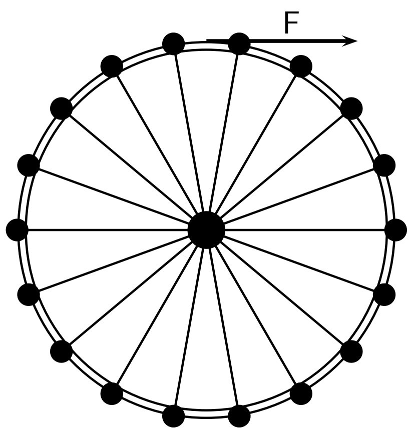


Figure 34b

The masses on the rim won't apply any torque to the wheel, because each mass will feel only a downward force mg, parallel to the axis, with zero torque around the axis). You can spin the wheel, in either case, by applying a horizontal force F to the rim. The magnitude of the applied torque is then $T = F \times R$, where R is the radius of the wheel.

To say what will happen to the wheel you need to know its moment of inertia, I. That's easy: you just use (6.8). If any bit of the rim has mass ΔM it will contribute $R^2 \times \Delta M$ to the moment of inertia about the axis; and since all the bits are the same distance from the axis the total moment of inertia for the whole wheel will be $I_0 = R^2 M$ (forgetting about the wire spokes, which are very light). If you apply the force F for a short time the wheel will start spinning about its axis.

But after adding the masses, as in Fig.34b, the wheel will not start spinning so easily – it will have much more *inertia*. If each lump of stuff has a mass m, and there are 18 of them, the loaded wheel will have a moment of inertia $I = I_0 + 18R^2 m$. You'll now have to apply a much larger torque, and perhaps for a longer time, to get the wheel moving (i.e. to increase its angular momentum) – as follows from equation (5.12). And once you've got the wheel moving it will be much harder to slow it down – as you'll discover it you put a stick between the spokes!

One last thing about rotational motion of a body around a fixed axis: since every bit of mass is moving, because it has an *angular* velocity, as well as the velocity due to its *translation* through space, it will have a kinetic energy. Just the kinetic energy of translation is $\frac{1}{2}MV^2$, we might expect something similar for the rotational kinetic energy, but with an *angular* velocity ω in place of the velocity V of the CM and a *moment of inertia* I in place of the total mass M. And that's exactly what we find.

To get the kinetic energy of rotation we can use the same method as in getting the angular momentum (moment of momentum) around the axis (the 'z-axis'). Every element of mass m_i is moving with a linear velocity $v_i = r_i\omega$, where ω has the same value for all points in the body, and therefore has a kinetic energy

$$\tfrac{1}{2}m_i v_i^2 = \tfrac{1}{2}m_i r_i^2 \omega^2.$$

If we add all contributions, remembering that ω is the same for all of them, the result will be

$$T_{rot} = \tfrac{1}{2}I_z\omega^2, \qquad (6.9)$$

where I_z is the moment of inertia as defined in (6.8).

Even when the axis of rotation is not fixed, but is free to turn and twist in any direction about one fixed point, a similar calculation can be made (though it is much harder). And even when the body is completely free and has no point fixed in space the total kinetic energy can be written as

$$T = T_{trans} + T_{rot}, \qquad (6.10)$$

where T_{trans} is the KE of the total mass M, as if it were concentrated at the CM and moving with velocity V, while T_{rot} is the extra KE due to rotational motion of all elements *relative* to the CM, as if it were fixed. So the 'separation' of the motion of a rigid body into translational and rotational parts is very general indeed – even though it's all in our minds, to help us to think and calculate!

6.3 Rigid bodies at rest. Statics

In Chapter 1, when we first started to talk about forces acting on a particle, we were mainly interested in *equilibrium* – where the forces were 'in balance' and didn't produce any motion. And we noted that Statics and Dynamics were the two main branches of the Science of Mechanics. Since then, we've nearly always been studying bodies *in motion* (i.e. Dynamics). Whatever happened to Statics, which is what most books do first? We did it that way because Statics is only a 'special case' of Dynamics, in which the bodies move with zero velocity! So once you've done Dynamics you can go straight to Statics without having to learn anything new. You only need ask for the conditions under which all velocities are zero - and stay zero.

These conditions follow from the two equations (4.7) and (5.12) which state, respectively, that (i) force produces linear momentum (and hence translational motion), and (ii) torque produces angular momentum (and hence rotational motion). The conditions for no motion at all – for **equilibium** – can therefore be stated as

$$\mathsf{F} = \sum \mathsf{f} = 0 \tag{6.11}$$

– the vector sum of all the forces acting on the body must vanish – and

$$\mathsf{T} = \sum \mathsf{r} \times \mathsf{f} = 0 \tag{6.12}$$

– the vector sum of all the torques acting on the body must also vanish. Here the terms have not been labelled, but it is understood that, for example, $\sum \mathsf{r} \times \mathsf{f}$ means $\mathsf{r}_1 \times \mathsf{f}_1 + \mathsf{r}_2 \times \mathsf{f}_2 + ...$, where force f_i is applied at the point with position vector r_i and the sum runs over all particles ($i = 1, 2, 3, ...$) in the body. If these two conditions are satisfied at any given time, then the forces will produce no changes and the conditions will be satisfied permanently, the body will stay in equiibrium.

So far, we have supposed that the torque refers to any axis *through the centre of mass* O. This was important in Dynamics; but in Statics there is no need to use the CM. If we take a new origin O′, with position vector R relative to the CM, the position vector of any point P will become $\mathsf{r}' = \mathsf{r} - \mathsf{R}$, instead of r. (If you're not sure about this make a diagram with dots at points O, O′, P and draw the arrows R (from O to O′), r (from O to P), and r' (from O′ to P): you'll see that r' is the vector sum of r and $-\mathsf{R}$). The torque about an axis through O′ and perpendicular to the plane of the vectors r, R, will then be

$$\mathsf{T}' = \sum \mathsf{r}' \times \mathsf{f} = \sum (\mathsf{r} - \mathsf{R}) \times \mathsf{f} = \sum \mathsf{r} \times \mathsf{f} - \mathsf{R} \times (\sum \mathsf{f}) = \mathsf{T},$$

as follows from (6.12). In statics, you can take moments about just any old point and the equilibrium condition $\mathsf{T}' = \mathsf{T} = 0$ is always the same!

Since the total force F and the total torque T are both 3-component quantities, and will only vanish if all their components are separately zero, the vector equations (6.11) and (6.12) are equivalent to two sets of ordinary (scalar) equations:

$$F_x = 0, \quad F_y = 0, \quad F_z = 0 \tag{6.13}$$

and, for the torque components,

$$T_x = yF_z - zF_y = 0, \quad T_y = zF_x - xF_z = 0, \quad T_z = xF_y - yF_x = 0, \qquad (6.14)$$

where, as usual, we use Cartesian coordinates where the x-, y-, and z-axes are all perpendicular to each other. Remember that the order of the labels (x, y, z) in each of the torque equations is *cyclic* – in the first equation, the x-component (on the left) depends on y- and z-components (on the right), while in the second you replace xyz by yzx and similarly in the third you again move the first letter to the end (yzx $\rightarrow$ zxy).

In the Examples, we'll see how these six scalar equations can be solved to find the kinds of equilibrium that can result.

Example 1 - a loaded bench

The Figure below represents a wooden plank, of weight w, supported at two points, P_1 and P_2.

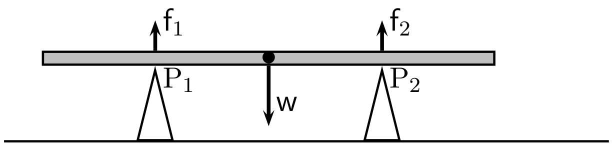


Figure 35

We can also add a number of loads (e.g. people sitting), of weights $w_1, w_2, \dots$, say, at distances $x_1, x_2, \dots$ to right or left of the midpoint $\bullet$. When the system is in equilibrium, what will be the values (in kg wt) of f_1, f_2, the upward forces exerted by the two supports?

As we know already, the force on the plank due to gravity can be represented by a vector of length w pointing vertically down from its midpoint.

Two other forces, with their points of application, are also shown in the Figure. Notice that any two forces with the same line of action are exactly equivalent: the point of application doesn't matter – you can slide the force along the line without changing its effect (its moment around any point stays the same). This is sometimes called the "principle of transmissibility of force". So the up arrows f_1 and f_2 are drawn as if the forces are applied at the top surface of the bench, while w starts from the underneath surface (not the CM, which is inside): this just makes the drawing clearer,

Let's now write down the two conditions, (6.12) for zero total force and (6.13) for zero torque. The first means (taking x-axis left-right along the bench, y-axis vertically upwards, and z-axis pointing towards you out of the paper)

$$f_1 + f_2 - w = 0. \qquad \text{(A)}$$

There is only one equation, for the force components in the y-direction. The x- and z-components are zero.

The second condition (6.13) also gives only one equation (can you say why?), where the support-points, P_1 and P_2, are at $x = -X$, relative to the mid-point as origin. This is (noting that w has zero moment about the origin)

$$-Xf_1 + Xf_2 = 0, \quad \text{(B)}$$

which means $-f_1 + f_2 = 0$; and if we solve these two *simultaneous equations* (see Book 1) by adding them together, it follows that $2f_2 - w = 0$. So $f_2 = \frac{1}{2}w$. The second unknown follows on putting this result back in the first equation: $f_1 + \frac{1}{2}w - w = 0$ and therefore $f_1 = \frac{1}{2}w = f_2$. As we'd expect, each support carries just half the weight of the bench.

Remember that the zero-torque condition applies for *any* choice of axes, when we calculate the moments. So we can check our results by taking moments around, say, an axis through point P_1. Again, the forces give only z-components of torque, but now instead of equation (B) we find

$$2Xf_2 - Xw = 0,$$

since f_1 has zero moment about the new axis, while w has a negative (clockwise) moment. Thus, $f_2 = \frac{1}{2}w$ and you get the same result as before.

It's more interesting to ask what happens if a heavy person, of weight W, sits on the bench, with his CM at a distance x from the mid-point. In this case the conditions (A) and (B) are changed: they become

$$f_1 + f_2 - w - W = 0, \qquad \text{(A)}$$

for zero vertical force, and

$$-Xf_1 + Xf_2 + xW = 0, \quad \text{(B)}$$

for zero torque about the z-axis through the origin. Suppose now the person weighs four times as much as the bench, so $W = 4w$, and sits down half way between the left-hand support (P_1) and the middle. In that case, $x = \frac{1}{2}X$ and the new equations are

$$f_1 + f_2 = 5w, \qquad \text{(A)} \quad \text{and} \quad -Xf_1 + Xf_2 = -2Xw. \quad \text{(B)}$$

Again, the two equations can be solved to give the values of the two unknowns, f_1, f_2. Cancelling a common factor X from (B) and adding the result to (A), gives $2f_2 = 3w$ so $f_2 = (3/2)w$; and if we put this result back in (A) we get $f_1 = 5w - (3/2)w = (7/2)w$.

That's the complete solution: the right-hand support carries three times the weight it carried before the person sat down, while the left-hand support carries *seven* times as much. Things like this are going to be very important if you ever have to build a bridge, with heavy trucks running over it. You'll want to know how strong the supports must be and how many of them will be needed.

Example 2 - a lifting device

The Figure below shows a device for moving heavy loads.

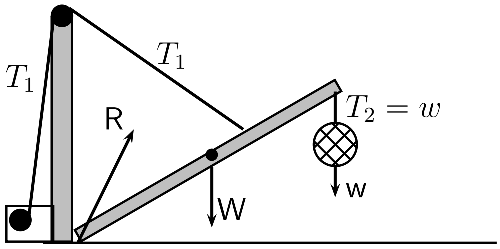


Figure 36

The load hangs from a beam (shown shaded grey), carried by a strong wire cable. At the start, the load will lie on the ground; but the beam can be pulled into a more vertical position by a second wire cable, which passes over a smooth bar and can be wound round a drum (shown as a circle on the left). What you'll want to know are the *tensions* T_1 and T_2 in the two cables; and also the *reaction* R of the ground – which acts against the *pressure* applied to it by the beam. So there are three unknowns, one of which is a vector (and we don't even know which way it points).

In the Figure, the load has been lifted clear of the ground and the beam is in equilibrium when all the forces acting on it satisfy the conditions (6.12) and (6.13). First we have to say exactly what they are. As you'll remember from Chapter 1, the tension in a string or cable is the same at all points: there are two forces, equal but acting in opposite directions. We'll usually just show the magnitude of the tension, without putting in all the arrows. T_1 is produced by the winding machine, T_2 by the weight it has to support.

As usual, we'll resolve all forces into horizontal and vertical components, so that T_1 acting on the beam will have components T_h (pointing to the left) and T_v (pointing vertically upwards). Similarly R will have components R_h (pointing to the right) and R_v (pointing directly upwards). The other forces are vertical – the weight W of the beam and the tension T_1, which has the same magnitude as the weight w it supports.

For equilibrium, the total horizontal force on the beam must be zero and so must the total vertical force. So we can write the conditions on the magnitudes of the forces as

$$\text{(a)}: \quad R_h - T_h = 0, \quad \text{(b)}: \quad R_v + T_v - W - w = 0.$$

Now take moments about a horizontal axis through the top end of the beam (which, we'll suppose, has length L and makes an angle θ with the ground). Counting anticlockwise torques as positive, this gives a third condition:

$$\text{(c)}: \quad -R_v(L\cos\theta) + R_h(L\sin\theta) + W(\tfrac{1}{2}L\cos\theta) - T_v(D\cos\phi)a - T_h(D\sin\phi) = 0,$$

where ϕ is the angle that the 'lifting' cable makes with the horizontal.

Now let's put in some numerical values, taking

$$L = 4 \text{ m}, w = 20 \text{ kg}, W = 40 \text{ kg}, \theta = 30\,\text{deg}, D = L/4 = 1 \text{ m}$$

– D being the distance from the top end of the beam to the point where the 'lifting' cable is attached. (To make the arithmetic easier, we'll choose the height of the wall so that

$\phi = \theta$, but if you have a pocket calculator you can use other values.) The three conditions then become

$$(a): \quad R_h - T_h = 0, \quad (b): \quad R_v + T_v = 60 \text{ kg wt},$$

and, for zero torque around a horizontal axis (pointing towards you out of the plane of the Figure),

$$\begin{aligned}(c): \quad & -R_v(4\cos 30\,\text{deg}) + R_h(4\sin 30\,\text{deg}) + W(2\cos 30\,\text{deg}) \\ & = T_v(3/2)\cos 30\,\text{deg} \;+ T_h(3/2)\sin 30\,\text{deg}\,.\end{aligned}$$

There are then four things we don't know, two components of the reaction R and two components of the tension T_1, which is a vector – even though we've only used its magnitude $T_1 = |\mathsf{T}_1|$. And there's a golden rule that to find n 'unknowns' you must have n independent conditions. We have only three – so something is missing. We must find another equation. What can it be?

You'll remember, from Chapter 1, that a string or cable can't apply a 'push'. It can only feel a tension – and this force can only be *along the string.* But here we have resolved the force (the vector T_1) into two components, one (horizontal) of magnitude $T_h = T\cos\theta$, the other (vertical) of magnitude $T_v = T\sin\theta$; and we were hoping to solve our equations as if the two components were independent. In fact, they are not: the ratio T_v/T_h is fixed by the direction of the string – so here is our missing equation. It is

$$(d): \quad \frac{T_v}{T_h} = \frac{T\sin\theta}{T\cos\theta} = \tan\theta = \tan 30\,\text{deg}\,.$$

And now we can solve the four equations (a),(b),(c), and (d).

(Remember that the sine, cosine and tangent of 30 deg can be obtained from an equilateral triangle, all angles being 60deg. Take each side of length 2 units and drop a perpendicular from one corner to the opposite side: each half of the triangle then has one angle of 90deg and sides of lengths 1,2 and $\sqrt{3}$ units. If you draw it, you'll see that $\sin 30\,\text{deg} = \frac{1}{2}$, $\cos 30\,\text{deg} = \frac{1}{2}\sqrt{3}$, and $\tan 60\,\text{deg} = 1/\sqrt{3}$.)

On putting in the numerical values and dividing all terms by $4\cos 30\,\text{deg}$ equation (c) becomes

$$-R_v + R_h(1/\sqrt{3}) - T_v(1/4) - T_h(1/4\sqrt{3}) = -20 \text{ kg wt}.$$

But from (a) $R_h = T_h$ and from (d) $T_v = T_h/\sqrt{3}$; so on putting these values in the last equation above we get (check it!)

$$-R_v + (1/2\sqrt{3})T_h = -20 \text{ kg wt}.$$

We're nearly there! Equation (b) told us that $R_v + T_v = 60$kg wt, which is the same as $R_v + (\sqrt{3})T_h = 60$kg wt. If you add this to the last equation above, you get rid of one unknown, R_v, which cancels out. So you are left with

$$(3/2\sqrt{3})T_h = 40\text{kg wt}, \quad \text{or} \quad T_h = (80/\sqrt{3})\text{kg wt}.$$

That's the first result. The next follows at once because (a) told us that $R_h = T_h$. And we know from (d) that $T_v = T_h/\sqrt{3}$. So $T_v = (80/3)$kg wt. Finally, putting the values of R_v and T_v into (b), we get $R_v = (60 - 80/3)$ kg wt $= (100/3)$ kg wt. And we're done!

Notice that the units (the metre and the kilogram weight) have been kept throughout, but if your equations are right you can safely leave out the units – they will 'look after themselves'. We kept them in just to make sure that everything was OK: the numbers that come out at the end are

$$R_h = 46.189, \quad R_v = 33.333, \quad T_h = 46.189, \quad T_v = 26.667$$

and as they all refer to forces they are correctly measured in 'kg wt'. The magnitude of the tension in the cable is also important and comes out as $T = \sqrt{T_h^2 + T_v^2} = 80(2/3) = 53.333$ – again in 'kg wt'.

Example 3 - equilibrium with friction

Sometimes we've talked about strings passing over *smooth* pegs and things sliding down *smooth* surfaces, as if there was nothing to stop the motion or to slow it down. But we know that real life is not like that: something may be hard to move because it is resting on a *rough* surface and, however hard you push, it never gets started. Or if it is already moving it doesn't go on forever – eventually it stops. But Newton's first law told us that an object could only change its "state of uniform motion in a straight line" if some *force* was acting on it – to make it move faster or to slow it down. Even when something is falling 'freely' through the air – perhaps a man falling from an aircraft, before he opens his parachute – the constant force due to gravity doesn't produce an acceleration that goes on forever: there is a 'terminal velocity' and when the speed stops increasing the total force on the body must be zero. In other words, the force applied by gravity, or by pushing or pulling, must be opposed by some kind of *resistance*; and when the two forces are equal and opposite the state of motion will stop changing. In Dynamics, we've usually left this resistance out of our calculations, saying it was so small we could forget about it and that our equations would be a 'good approximation'. But in Statics, where the forces acting are 'in balance' and result in *equilibrium*, we can't neglect anything, however small. The resistance to motion offered by a rough surface, or by the air being pushed out of the way by a falling body, is called **friction**. It opposes any kind of motion and it's not easy to make theories about it because it depends on very small details of the 'interface' between things in contact. But it's so important that life would be very different without it. You wouldn't even be able to walk without it! Try walking on a very smooth slippery surface: if you step forward with one foot, the other one goes backwards and your centre of mass stays where it was – without friction you have nothing to push against. And you wouldn't be able to write, because without friction the pen would slip through your fingers!

So how do we deal with friction in our theories? We have to fall back on experiment, which can give us 'empirical' laws, that can then be expressed mathematically. For hundreds of years the laws of friction have been known; and they're very simple to write down and apply. A body like, say, a brick, resting on a horizontal surface (Fig.37) will feel only the

the downward force (its weight W) due to gravity, and an upward reaction ($\mathsf{N} = -\mathsf{W}$) exerted by the surface. Suppose you now apply a horizontal push F, as shown in the Figure.

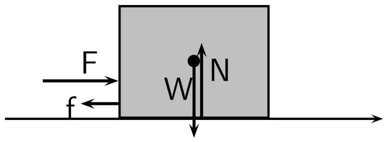


Figure 37

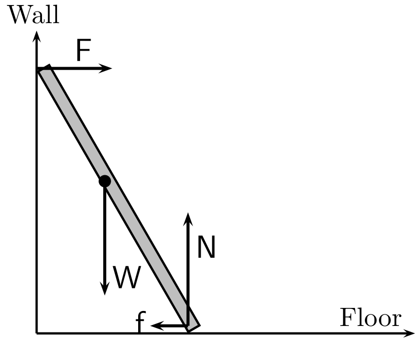


Figure 38

What you'll find is this: At first nothing happens; but then, when F reaches a certain value F_0, the equilibrium is broken and the brick begins to slide. At that point, F is exactly opposed by a **frictional force** f, which arises from the contact between the two surfaces – the underneath of the brick and the surface that supports it. So we can say $f \leq F_0$, where the equality applies just when the brick begins to move. Moreover, F_0 depends only on the nature of the two surfaces and N, the modulus of the normal force N that presses them together: if you double N then you double F_0 – the maximum frictional force you can get is *proportional* to N. Putting all this together, the basic law of friction can be written

$$f \leq \mu_s N, \tag{6.15}$$

where the proportionality constant μ_s is called the "coefficient of (static) friction". Once the block starts moving, the frictional force usually becomes a bit smaller, but the same relationship holds except that μ_s is replaced by μ_d – the "coefficient of (dynamical) friction". Note that both coefficients, relating one force to another, are numbers – without physical dimensions – and that they relate only the *magnitudes* of the forces. For any given surfaces they can be found only by experiment. Also N is always *normal* to the contact surface, while f is perpendicular to N and opposes the force F producing the motion.

Equation (6.15) is not an exact law; but it usually holds in good approximation and is easy to apply. Let's have a go.

Figure 38 shows a ladder propped against a vertical wall (the y-axis): without friction between the foot of the ladder and the horizontal floor (the x-axis), there could be no equilibrium (can you say why?). The ladder would just slide down before you could even start to climb it. Suppose the ladder has length L, with its CM at the midpoint, and that its foot is a the point $(X, 0)$ while its top is at $(0, Y)$. The labelled arrows indicate the forces acting:

W = weight of the ladder,

N = normal reaction from the floor,

f = frictional force ($f \leq N$),

F = normal reaction from smooth wall.

Resolving all forces into their x- and y-components, equilibrium requires that

$$\text{(a)} \quad F - f = 0, \quad \text{(b)} \quad N - W = 0,$$

for no motion in the x- and y-directions.

Now take a horizontal axis through the foot of the ladder: the condition for zero torque around the axis is,

$$\text{(c)} \quad -F \times Y + W \times (\tfrac{1}{2}X) = 0.$$

The first question to ask is: Will the ladder stay up, or will it slip? And of course we'll need to know how long and heavy it is; and also what angle it makes with the wall and what is the value of the coefficient of friction. To make the arithmetic easy let's take the length as $L = (13/2)$ m and put the foot of the ladder (5/2) m away from the wall. Suppose also that $W = 80$ kg wt and that the coefficient $\mu_s = 0.4$.

There are now three conditions (a,b,c) and the things we don't know are f, N, F – so we have enough equations to find them. It's easy: from (b) we have $N = W = 80$ kg wt; and from (c) $F \times 6 = W \times (5/4)$, which gives $F = 50/3$ kg wt. Finally, from (a), $f = F = (50/3)$ kg wt.

Now the greatest possible value of f is $0.4 \times N = 0.4 \times 80 = 32$ kg wt. So we've answered the question: we only need a frictional force of (50/3) =16.667 kg wt to keep the ladder in equilibrium, so all is well – f can go up to 32 kg wt before the ladder will slip!

The next question to ask is: How far can I climb up the ladder before it slips? Try to answer this for yourself. Put in your own weight, w, and suppose you go a horizontal distance x from the foot of the ladder. Then ask how equations (a),(b) and (c) must be changed and finally solve them.

Exercises (in preparation)

Chapter 7

Some simple machines

7.1 Levers

A **machine** or **device** is some kind of tool that will help you to do a particular job, like lifting a heavy weight or digging a hole in hard ground. One of the simplest machines you can imagine is probably a **lever**, which can take many forms depending on what you want to do; and the simplest lever is just a strong iron bar – strong so it won't bend when you use it. Two kinds of lever are shown in Fig.39, each being used for a similar purpose – to topple a heavy block of wood.

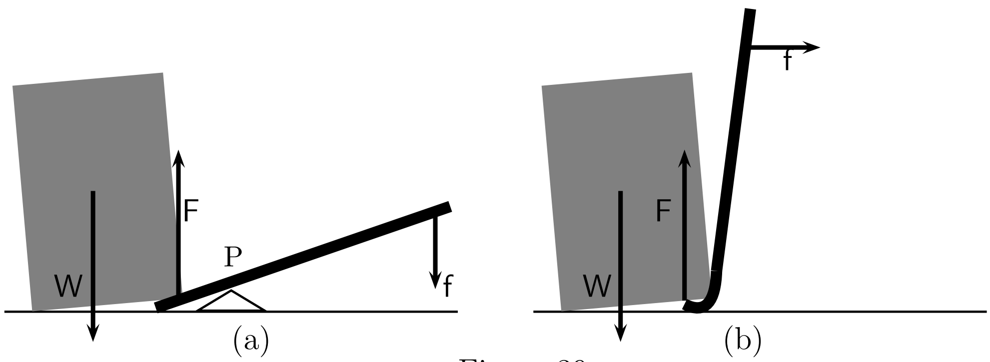


Figure 39

In Fig.39a a long bar is supported (at some point P) by a **pivot**, strong enough to carry a heavy weight without sinking into the ground and 'sharp' enough to allow the bar to turn easily around the particular point P. Suppose you want to push over the big block of wood, on the left in Fig.39a, and it's much too heavy to move by hand. You can do it by putting one end of the lever under the block (you may have to dig a small hole if there isn't enough space to get the bar in) and then pressing down with all your weight on the other end. How does it work?

Suppose the horizontal distance from the pivot P to the end of the bar, under the block, is d_1 and that from P to the point where you apply the force f is d_2. Then the force F applied to the block when it just starts to move will be equal but opposite to its reaction $-\mathsf{F}$ on the bar. When the bar is 'in balance' we can take moments about P and say that the anticlockwise moment of $-\mathsf{F}$ plus the clockwise moment of our applied force f must be zero; and that means that $-Fd_1 + fd_2 = 0$.The ratio of the magnitudes of the two forces is thus $R = F/f = d_2/d_1$: the force you can apply to the block is R times as big as the force you have to apply with your muscles. This ratio is called the **mechanical advantage** of the device – with good reason, because if the bar is 2 m long and you put the pivot 10 cm from the edge of the block you can apply an upward force to the block of twenty times your own weight!

Figure 39b shows another kind of lever, in which the bar is *bent* at one end: that's the end you use, by putting it under the edge of the block and this time *pulling* the free end of the bar (which is nearly vertical) towards you. Sometimes it's easier to pull than to push; and also there's no need to supply a separate pivot – the bent end of the bar acts as its own pivot, provided you put a bit of iron plate under it so that it doesn't sink into the ground. In both cases, whether you push or pull, the mechanical advantage is $R = d_2/d_1$; but notice that in Fig.39b the distance d_1, from the lifting end to the pivot (the point of contact between bar and ground - or plate) can be very small, making the ratio R correspondingly larger.

There are many other kinds of lever, but the idea is always the same: There is a Load, a Pivot, and an Applied Force; and if the distance from Load to Pivot is d_1, while that from Applied Force to Pivot is d_2, then your mechanical advantage is the ratio d_2/d_1 – the force you apply is 'magnified' by this factor.

7.2 Weighing machines

At the beginning of Book 1, the idea of weighing things was introduced. The thing to be weighed was put in a 'pan', which moved a pointer over a scale (marked in kilogramme weight units) to show how much the object weighed. The weight is a *force* and two forces are equal if they move the pointer to the same point on the scale. The weighing machine must be **calibrated** by putting, in turn, 1,2,3,... standard units (of 1 kg) in the pan and marking the 'pointer readings' on the scale in the same way; and if an object put in the pan moves the pointer to half way between the points marked '2 kg' and '3 kg' then we say it weighs 2.5 kg.

The basic operation in weighing is that of *comparing* two weights. The easiest way of doing that is to make a simple **balance**: all you need is a wooden board and something to act as a pivot as in Fig.40. Near each end of the board you draw a line to show where the weights must be placed. Before putting anything on the board you must make sure it stays 'in balance' (i.e. in *equilibrium*) when you rest it on the pivot, placed halfway between the two lines (if it doesn't, you can add a lump of clay on one side or the other until it does). When it's right, as near as you can make it, go ahead –

To use the balance you need a standard set of weights: perhaps a small plastic bag of sand would weigh about 100 gm, so you'll need ten of them to make 1 kg. Make sure they are equal by putting one on each side (on top of the line) and checking that they stay in balance: then, if the distance from the pivot is d, one weight will have an anticlockwise moment $d \times w_1$ and the other a clockwise moment $-d \times w_2$; so the total torque will be zero only when $d \times (w_1 - w_2) = 0$ and the units are equal, $w_1 = w_2$. Now you can weigh anything up to 1 kg by putting it on one side and seeing how many 100 gm units you must put on the other side to get equilibrium. When the weight W is balanced by five units $W = 500gm$. If 5 units are not enough, but six are too much, you'll need a set of sub-units (each of, say, 20 gm) and you can go ahead in exactly the same way; if you have to add two sub-units to get things to balance then the unknown W will be 540 gm or 0.54 kg.

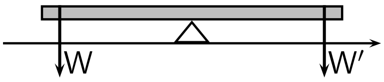


Figure 40

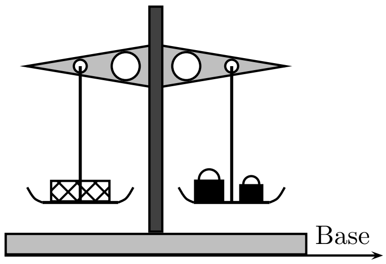


Figure 41

Of course such a simple 'machine' is not going to give accurate results (can you give some reasons?) but it can easily be improved. A more accurate type of balance is indicated in Fig.41. It is usually made of metal and contains more than one pivot. The central column supports the horizontal arm on a 'knife-edge'; and two equal 'pans' hang from the two ends of the arm, each being supported on its own knife-edge. There's usually a pointer, with a scale behind it, to show when things are exactly in balance – but *not* for showing the weight. The standard weights to be used are now usually very accurately made pieces of metal, coming in units of 100 gm, 50 gm, 20 gm, 10 gm etc. down to 1 gm, 0.5 gm (5 milligrammes), and even smaller sub-units, depending on what the balance is being used for. This kind of balance was used for many years in weighing chemicals, in shops and laboratories, but nowadays you nearly always find electronic devices which automatically show the weight in figures.

Another kind of weighing machine is still widely used in markets, for weighing heavy things like sacks of vegetables – or even people. In its simplest form it is made from a long iron bar, hanging from a strong beam (as shown in Fig.42) on a hook which acts as the pivot. Not far from the pivot hangs a scale pan on which you put the thing you want to weigh. The standard weights are usually heavy metal slabs going from, say, 5 kg down to 1 kg that you can put on a metal plate hanging from the bar – on the other side of the pivot – as in the Figure . So if you put two of the big weights and three of the smallest on the plate you'll have a 13 kg weight.

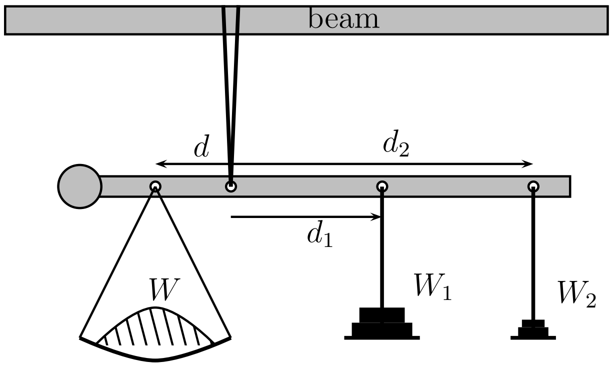


Figure 42

If you want to be more accurate you'll need also sets of smaller weights, perhaps going down from 500 gm to 50 gm, and a smaller plate to carry them, which you can hang from a different point on the bar – as shown in the Figure: this kind of balance used to be called a "steelyard" (the 'yard' being an English measure of length, a bit shorter than the metre, and 'steel' being a much stronger material for making the bar from). Often the bar has a number of holes at various points, so you can choose where to put the weights, according to the load you are weighing.

To use this kind of balance, you put the load on the scale pan (shown on the left in the Figure) and hang the weight you guess will be about right from the first hole to the right of the pivot. If the distance d_1 is *twice* the distance d, and the guessed weight is W_1 (e.g. 15 kg if you've put three of the 5 kg slabs on the plate), then the balance will go down on the left if the unknown W is *greater* than 2 ×15 = 30 kg; but down on the right if W is *less* than 30 kg. Suppose it goes down on the right. If you take away one of the 5 kg weights W_1 will be 10 kg; and if the balance then tips to the left you can then say $W > 2 \times 10 = 20$ kg – so you took away too much. Try, instead, with $W_1 = 14$ kg (adding two weights of 2 kg): if it still tips to the left, then $W > 2 \times 14 = 28$ kg. So you can say $28\,\text{kg} < W < 30\,\text{kg}$. Of course if your guess of 28 kg was correct the balance will tip neither left nor right – it will "stay in balance" on the knife-edge. But if this is not so, then you'll have to start with the smaller weights.

Suppose the smaller weight (W_2) hangs at a distance $d_2 = 4d$ from the pivot. Then the condition for staying in balance – the weights having zero total moment around the pivot – will be

$$d \times W - d_1 \times W_1 - d_2 \times W_2 = 0.$$

When this condition is satisfied (with $d_1 = 2d$ and $d_2 = 4d$) we can cancel the factor d and write $W = 2W_1 + 4W_2$. And if equilibium results when the smallest weight is $W_2 = 150$ g (=0.15 kg) then we can say $W = 28 + 4 \times 0.15 = 28.60$ kg.

Balances of this kind have been used for thousands of years in all parts of the world: you find pictures of them in the wall paintings in Egyptian tombs, in ancient Persian manuscripts, and in many other places. They can also be found in improved forms, which are easier to use. For example, the arm of the balance which carries the weights sometimes has a scale, with numbers showing the distance of points from the pivot, and a single fixed weight W_0 can slide along the scale. If equilibrium results when the weight (or rather its centre of mass!) is at a distance D then the unknown weight is $W = (D/d)W_0$ – which can be read off directly from the scale, once it has been calibrated.

7.3 The wheel

Wheels, in one form or another, have also been in use for many thousands of years. In this book, we met them first in thinking about pushing and pulling things, using some kind of cart: without the cart, and its wheels, you'd have to *drag* everything you wanted to move – so the invention of the wheel was an enormous step forward. Now we know about friction it's clear that the wheel makes it easier to move things by getting rid (almost completely) of the forces called into play when things rub together: if you try to drag a heavy box over rough ground it may be impossible, but if you put wheels on it it will run smoothly. Until a few hundred years ago this was one of the most important uses of the wheel; another one being that it made it easier also to *rotate* heavy objects, like a heap of wet clay, by putting them onto a horizontal wheel, or table, with the axis pointing vertically upwards. The "potter's wheel" has been in use for probably two or three thousand years in the production of urns and pots of all kinds.

The next big advance was in the use of wheels for moving *water* in countries where it hardly ever rains and water is very precious. Nothing can grow without water; and, even if there is a short rainy season, the water soon runs away unless you can get it out of the river and up onto the land, where it's needed for growing crops. How to do it is the problem of **irrigation**. And if you have to move water you need **energy**. One way of solving both problems came with the development of the **water wheel**.

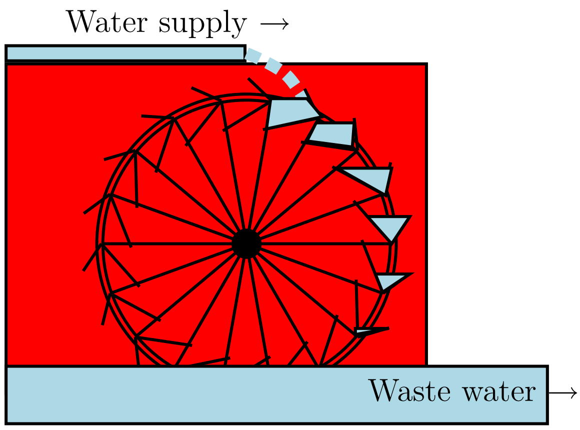


Figure 43

The first wheels of this kind probably came from Egypt, where they were in use over a thousand years ago: they came to be known as "noria" wheels and in some places they can still be found. At Hama, for example, in Syria there are some giant wheels (20 m or more in diameter!) which have been running continuously for hundreds of years, using water power to lift water from the River Euphrates and supply it to aqueducts, which carry it a long way to irrigate the fields.

First, let's look at a simple water wheel (Fig.43) of the kind that was widely used in some countries during the Industrial Revolution – when people started working with machines, in factories, instead of depending on their own muscles. Fig.43 shows a wheel of the type used in driving a heavy 'millstone' for milling grain to make flour for bread. They were also once used for driving mechanical hammers in the steel industry (but more of that later).

The 'water supply' comes from a point 'upstream' on the river, where the water level is higher; it comes to the mill wheel along an open pipe or channel (at the top in the Figure) and falls onto the specially made 'boxes' around the edge of the wheel. As the boxes fill with water, their weight produces the torque that turns the wheel. But as they go down the water spills out; and finally it goes back into the river as 'waste water' – having done its work.

The giant noria wheels at Hama are very similar in design, but do exactly the opposite job: they *take* water from the river (at low level), by scooping it up in the 'buckets' fixed around the rim of the wheel, and then emptying them into the aqueduct (at high level) when they reach the top. A wheel of this kind is shown Fig.44, where you see it from the edge, which lies in the vertical plane with the axle horizontal. The river, shown in blue, is flowing away from you and the top edge of the big wheel is coming down towards you.

Because a lot of energy is needed to lift all that water, the wheel needs *power* to drive it; but, if there's plenty of fast-flowing water in the river, some of it can be used to turn a much smaller wheel like the one in Fig.44 and this can provide the power. How to get the energy from one wheel to the other is a problem in **power transmission**, which we'll think about next.

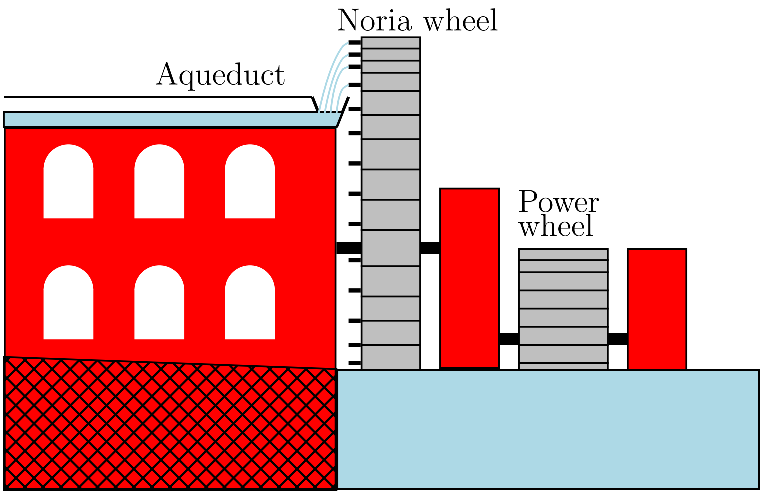


Figure 44

Suppose we have two wheels, one big and one small, and want to make one drive the other. The simplest way of doing it is to put them side by side in the same plane, each with its own axle, which supports the wheel and provides the axis around which it can turn; and then to tie them together with a loop of rope or a 'belt' – as in Fig.45 (below). Each axle must be carried by a pair of 'bearings', to hold it in the right place, and the belt must be kept *tight* so that it doesn't slip when the wheels are turning.

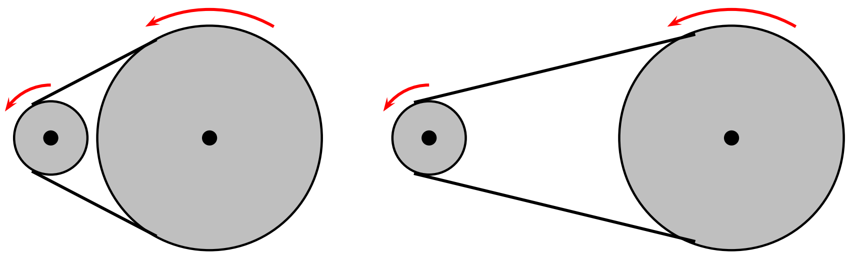


Figure 45

Note that the wheels don't have to be close together; but if they *are* (Fig 45a) then they mustn't *touch*. That's because (as the red arrows show) the two wheels turn in the same sense (clockwise or anticlockwise), so the parts that come closest are going in opposite directions; and if they were rubbing together the friction would slow them down – or even stop them.

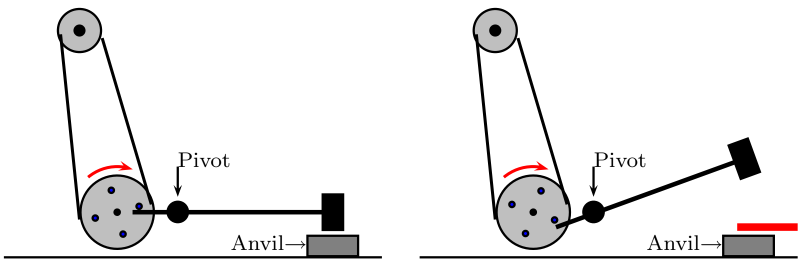
The big wheels in the Figure have diameters three times as big as the small wheels; so, if they are d and $D = 3d$, any point on the rim of the small one will travel a distance πd in one turn and that's the length of rope it will pull in. It will also be the distance moved by the other end of the rope, where it meets the big wheel; but it's only a *fraction* of the distance $\pi \times 3d$ that any point on the big wheel travels in one complete turn. So one turn of the big wheel takes 3 turns of the small one! **The small wheel has to turn D/d times as fast as the big one, where D is the diameter of the wheel it is driving and d is its own diameter.** This result doesn't depend on how far apart the wheels are, as long as the belt is tight and there is no slipping. The big wheel in both Fig.45a and Fig.45b goes just three times as slow as the wheel that is driving it.

Now we have a way of transmitting power from one rotating wheel to another we can look at ways of using the idea. In Fig.44, for example, the 'power wheel' has to transmit its energy (remember that power is the energy spent in unit time) to the noria wheel. If, in the Figure, the river is flowing away from you, then both wheels will turn the same way – their tops coming down towards you. So it's *possible* (though I don't know if this is the way it's done in Hama!) that the building between the two wheels holds two much smaller wheels, arranged as in Fig.45b, with the smaller of the two fixed on the axle of the power wheel and the bigger one fixed on the axle of the noria wheel. That way the flowing water, pushing agains the big flat boards of the power wheel, will drive the noria wheel.

Another example is the water-driven mechanical hammer, which was once used in the steel industry, in countries where it rains a lot and there are many small streams coming down from the hills. Some of this machinery, originally made from wood, can still be found today – in museums – and sometimes can be seen actually working!

Figure 46 shows how it works. The big hammer on the right has to be lifted and then dropped (it's very heavy) on a bar of red-hot metal which will be put on the heavy steel 'anvil' – where it will be beaten into shape. The power for doing this, again and again, all day long, comes from a water wheel like that in Fig.43. It is transmitted from a small wheel (overhead on the left) through a belt which drives a bigger wheel, which in its turn lifts and drops the hammer. How can it do that?

The big wheel has strong pegs sticking out of it, close to the rim, while the shaft of the hammer is supported on a pivot, which allows it to be turned when you press down on the free end. That's what is happening in Fig.46a, where one of the pegs on the wheel that drives it is pressing down on the end of the shaft – and lifting the hammer.



(a) (b)

Figure 46

In Fig.46b the hammer is shown in the lifted position and the end of its shaft is just about to slip off the peg as the wheel goes on turning. That gives the man who works the machine just enough time to put the red-hot bar on the anvil before the hammer comes crashing down and flattens the metal; and that's how knives (and swords) were made! Usually there would be many machines, side by side, all driven by the same water wheel.

Nowadays, of course, things have changed and the power needed in our factories hardly ever comes from water: it comes instead from burning fuel – wood or coal, gas or oil – and a lot of the energy it contains is wasted in heat and smoke. We'll study energy production in other Books of the Series; but the idea of the wheel is here to stay and plays an important part in our daily life.

7.4 Clocks and mechanisms

The next big advance in using the wheel came with the invention of the **gearwheel**, a wheel with 'teeth' or 'cogs' around its rim. Two such wheels are shown in Fig.47, which you can compare with Fig.45a. Instead of the belt, one wheel drives the other by making *contact* with it, the teeth of the first wheel fitting into the spaces between the teeth of the second; the gearwheels are said to 'engage'. So when the first wheel turns, the second one turns with it *but in the opposite sense*, as indicated by the red arrows in the Figure.

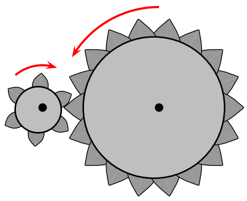


Figure 47

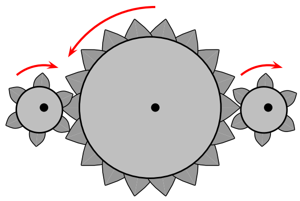


Figure 48

Of course you may want the second wheel to rotate in the *same* direction as the one that's driving it; and if the first wheel is going anticlockwise and you can't change it (perhaps it's fixed to a water wheel – and you can't change the way the water flows!) how can you make the driven wheel go the same way, anticlockwise?

The next Figure (Fig.48) shows how it can be done. With the driving wheel on the left, going anticlockwise, the next wheel must go clockwise; but, if you add a third wheel, then that will again go the opposite way – anticlockwise – which is what you wanted! In fact, if you have a 'train' of N wheels, the last one will turn the same way as the first if N is an odd number, but the opposite way whenever N is *even*. And this doesn't depend on the size of the wheels, or on how many teeth they have. So when you are transmitting power through a train of wheels you can always get the right sense of rotation by using the right number of wheels. You can also make the last wheel rotate *faster or slower* than the first, because when a small wheel drives a bigger one the speed will be reduced, while a big wheel driving a smaller one will make it go faster.

A train of wheels is an example of a **mechanism**, usually just a *part* of a machine, which carries out one special job. If you ever take an old clock apart you'll find it's full of strange mechanisms, each with its own job to do. In the rest of this Section we'll look at other examples.

In a clock the number 60 is very important: there are 60 seconds in a minute and 60 minutes in an hour. The simplest device for measuring time is the **pendulum** – just a weight (called a 'bob') on the end of a string or a light stick. The time taken for one 'double-swing' (back and forth) is the **period** and this depends (in good approximation) only on the length l of the pendulum and the acceleration due to gravity (g, which we met in Chapter 1). When $l \approx 1$ m the period is almost exactly one second; so we could use a simple pendulum as a clock – every 3600 double-swings would tell us that 1 hour had passed. But who is going to count them? That's the job of the clock.

Three things are needed: *power* to drive the clock; a mechanism to convert every 3600 swings into the turn of a pointer (the 'hour-hand') through one twelth of a complete revolution; and some way of giving the pendulum a little push, now and then, to keep it swinging. It can all be done by using wheels.

Let's start in the middle by looking for reducing the rate of rotation of a wheel by a factor of 60, which is $3 \times 4 \times 5$, so that the clock won't run down too fast. We already know (Fig.47) that we can get a factor of 3 from two wheels, by making the diameter of one three times that of the other; and clearly we can do the same for factor 4 and 5. In the next Figure (Fig.49a) we show wheels of diameters d and $4d$, with six teeth and 24 teeth, respectively, so the big one will go four times slower than the one that's driving it. We could even get that factor of 60 by making a wheel *sixty* times the diameter of the small one and cutting 360 teeth into it – though that would take a lot of material and a lot of patience!

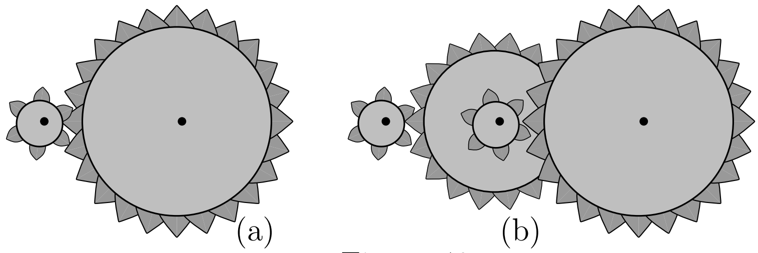
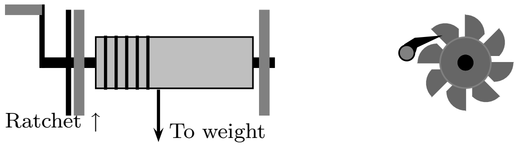


Figure 49

But now we do something clever! We *combine* this mechanism with the one in Fig.47 by putting the 6-tooth wheel on the same axle as the 18-tooth wheel. So now Wheel 2 will rotate 3 times slower than the driving wheel (Wheel 1) and Wheel 3 will rotate 4 times slower than Wheel 2 – and $4 \times 3\times$ times slower than the driving wheel! We only need one more pair of wheels, with diameters d and $5d$ (6 teeth and 30 teeth), to get in exactly the same way the remaining factor 5. And then we'll have a mechanism, with only six wheels and three axles, that will give us the magic factor of sixty! Once we have have made two of them, we can use the first one to go down from pendulum swings (seconds) to minutes; and the second to go down from minutes to hours. All we need now is a source of *power*, to keep everything moving, and some kind of *control* to make sure the wheels don't all turn too fast – running down the clock in almost no time.

In the simplest clocks the power is usually provided by a falling weight, the loss of potential energy being changed into rotational motion, which has to overcome the frictional forces that resist the motion. You have to 'wind up' the clock at night to give it enough energy to get through the next day; and to do this you can hang the weight on a string (or a wire cable) and wind the cable round a cylinder (or 'drum'), as in Fig.50a, by turning the handle. To prevent the weight dropping to the ground, as soon as you let go of the handle, another small device is needed. The axle of the drum must have a special kind of toothed wheel on it (called a 'ratchet') and something (called a 'pawl') to 'lock' the wheel if it tries to turn the wrong way. The 'ratchet and pawl' is shown in Fig.50b and you can see how it works: in the picture the wheel can only turn in the clockwise direction, the pawl being lifted by the force acting on it and letting the wheel turn – one tooth at a time; if you try to turn it the other way the pawl gets pushed to the bottom of the tooth – and stops it moving.



(a) Winding drum (b) Ratchet and pawl
– seen from left
Figure 50

Now we come to the most important thing, the 'brain' of the clock, which controls all its movements: what we need is some device to hold back the wheels, so they don't let the clock run down (wasting all the potential energy we gave it in winding it up) in the first few seconds. We want it to take 24 hours, so it need winding up only once a day.

The thing that takes care of this is called an 'escapement wheel', because it holds back the teeth but lets them 'escape' one at a time, once for every double-swing of the pendulum. Figure 51 shows just the top half of such a wheel, along with the 'escapement' itself, which has two legs and sits astride the wheel. In the first position (Fig.51a) the left-hand leg is digging its 'heel' (really called a "pallet") in between two teeth; and is pressing itself against the vertical side of a tooth so it can't turn in the direction of the arrow (clockwise). The power that drives the wheel will have to wait – but for what?

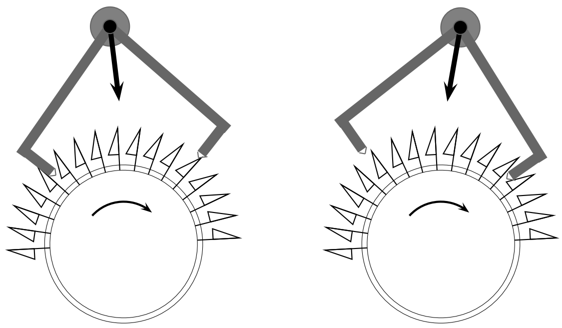


Figure 51

At this point, the pendulum (hanging down behind the escapement in the direction of the thick arrow) is just a bit to the right of vertical. But as the pendulum swings back to the left it moves the escapement so that its left leg comes up, letting the tooth 'escape' and move one place to the right. At the same time, its right leg goes down (Fig.51b), between two teeth further along the wheel, and again stops the wheel turning. But only until the pendulum completes its double-swing and everything returns *almost* to Position (a): this really means *just before* Position (a), where the tooth is just about to move. So the escapement wheel moves by one tooth at a time, once in every double-swing of the pendulum!

How does the power keep the pendulum swinging? The pendulum hangs down between the two prongs of a fork (not shown), which is fixed to the axle of the escapement. As the escapement rocks backwards and forwards, the fork gives the pendulum a little push, to the left or the right – just enough to keep it moving. To do this, the pallets must be carefully cut to shape so that, when the pallet slips off a tooth, its 'sloping' face (at the bottom) is given a sudden *impulse* by the tooth as it pushes its way past: that part of the pallet is called the "impulse face". Similarly, the tooth is 'stopped dead' when it meets the "dead face" of a pallet. The impulse goes to the pendulum through the fork that embraces it; and keeps it swinging, once every second, for as long as there is power to turn the wheels. The "tick...tock" of the clock is the noise made by the teeth of the escapement wheel alternately striking the 'impulse face' and the 'dead face' of the pallets.

What a marvellous invention! Such a small and simple thing – which has kept pendulum clocks going, all over the world, ever since it was first thought of 300 years ago.

Chapter 8

Turning mass into energy

8.1 A reminder of special relativity theory

In Book 2 we started from simple ideas about measuring distance and showed how the whole of Euclid's geometry could be built up from a **metric axiom** (Section 1.2): in three dimensional space this states simply that the distance s from some **origin** O to any point P can be obtained from the formula

$$s^2 = x^2 + y^2 + z^2 \tag{8.1}$$

where x, y, z, the **coordinates** of point P, are distances from O to P measured along three **perpendicular axes** (the x-axis, y-axis, and z-axis). (If you've forgotten all this, go back and look at Chapter 5 in Book 2.) And for two points whose coordinates differ by amounts $\mathrm{d}x, \mathrm{d}y, \mathrm{d}z$, however small, the distance between them is obtained in the same way

$$\mathrm{d}s^2 = \mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2. \tag{8.2}$$

This 'differential form' of (8.1) is called the 'fundamental metric form', $\mathrm{d}x, \mathrm{d}y, \mathrm{d}z$ being the **differentials** (Book 3, Section 2.4).

But we ended Chapter 7 of Book 2 by saying how much our ideas about space had changed over the last 100 years. Einstein showed that Euclidean geometry could not be perfectly correct and in his theories of **relativity** showed how and why it must be changed. The changes needed are so small that in everyday life they are completely negligible; but in Physics they can't be neglected. By taking them into account, the world has already been changed!

In coming to the end of Book 4, the first one on physics, you now know enough to understand what's been happening; so let's first remind ourselves of the relativity theory outlined in Section 7.2 of Book 2. The new thing is that the idea of a *point* in space, indicated by three distances (x, y, z), needs to be replaced by an *event* in which a fourth 'coordinate' t is also included: If I say "I'm here today but there tomorrow" then I'm referring to two events, the first being x_1, y_1, z_1, t_1 and the second being x_2, y_2, z_2, t_2. The sets of four 'coordinates' then indicate two points in **spacetime**; and if we want to define the 'separation' of two 'nearby' events we'll do it by writing

$$\mathrm{d}s^2 = -\mathrm{d}x^2 - \mathrm{d}y^2 - \mathrm{d}z^2 + c^2\mathrm{d}t^2, \tag{8.3}$$

where the constant c is put in to keep the physical dimensions right – it must have the dimensions of *velocity* LT^{-1} so that when multiplied by a time it gives a *distance*, like the other quantities $\mathrm{d}x, \mathrm{d}y, \mathrm{d}z$. But what about the + and – signs? Why don't we just add all the terms together?

Equation (8.3) is used as the fundamental metric form in '4-space'; and *ds* defined in this way is called the **interval** between the two events. We met a similar equation first in Section 7.2 of Book 2, where we noted that the condition

$$s^2 = c^2t^2 - x^2 - y^2 - z^2 = 0 \tag{8.4}$$

was one way of saying that some kind of signal, sent out from the origin of coordinates ($x = y = z = 0$) at time $t = 0$ and travelling with velocity c, would arrive after time t at points on the surface of a sphere of radius $R = \sqrt{x^2 + y^2 + z^2}$. That was why we started to think that separations in 'ordinary' space ($\mathrm{d}x, \mathrm{d}y, \mathrm{d}z$) and those in time, namely $c\,\mathrm{d}t$ (defined after multiplying by c to get the dimensions right), should be treated differently. By choosing (8.3) as a measure of the 'interval' it is clear from the start that a 'time coordinate' *ct* is not the same as a space coordinate: the interval between two events is said to be 'time-like' if the time term $c^2\mathrm{d}t^2$ is greater than the space term $\mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2$, or 'space-like' if it's the other way round.

The next important idea in Section 7.2 of Book 2 was that of the **invariance** of the interval as measured by two different people (the 'observers') in different reference frames, each moving with uniform velocity relative to the other. Such frames are **inertial frames**, in which Newton's laws about the motion of a particle, and its resistance to change ('inertia') when no force is acting, are satisfied for an observer in the same frame as the particle.

In the Figure below two such reference frames are indicated within an outer box, Frame 1 holding the first observer and Frame 2 (shaded in grey) holding the second; Frame 2 is moving with constant speed u along the x-direction and for Observer 2 it's 'his world'.

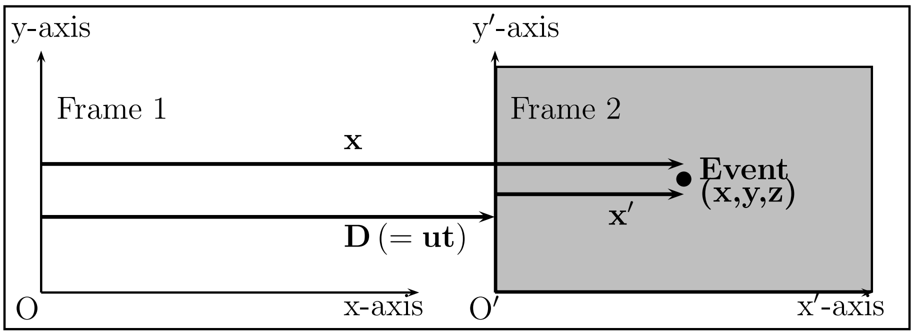


Figure R1

We first met the idea of invariance in Book 2 (Section 5.2), where we noted that certain changes of coordinates, in which x, y, z are replaced by x', y', z', left unchanged lengths and angles in 3-space. For example, the distance r from the origin O to Point P, with coordinates x, y, z is given by $r^2 = x^2 + y^2 + z^2$; and any **transformation** in which the line OP is simply rotated into OP′ leaves invariant the squared length –

$$x^2 + y^2 + z^2 \;\rightarrow\; x'^2 + y'^2 + z'^2 = r^2.$$

But now we're thinking about *events*, in which *four* coordinates are needed to specify a corresponding point in spacetime; and we already know from Book 2 that it's possible to find transformations in which x, y, z, t are replaced by x', y', z', t' in such a way that the form (8.4) stays invariant. The simplest transformation, is the one that corresponds to shifting the reference frame for Observer 1 along the x-axis by an amount D equal to an x-velocity (u) times the time (t) on his clock: this is the distance from the origin O to the origin O′ of the new reference frame (Frame 2) in which we're putting Observer 2. This is usually taken as the 'standard' **Lorentz transformation**: it relates the distances and time (x', y', z', t') at which Observer 2 records the event to those (x, y, z, t) recorded by Observer 1. There's only one event, taking place at the point shown by the bold dot in Frame 2, but both observers can see it. The transformation equations are

$$\begin{aligned} x' &= \gamma_u(x - ut), \\ y' &= y, \\ z' &= z, \\ t' &= \gamma_u\left(t - \frac{u}{c^2}x\right), \end{aligned} \tag{8.5}$$

where the quantity γ_u, which depends on the speed u with which Frame 2 is moving relative to Frame 1, is given by

$$\gamma_u = \frac{1}{\sqrt{(1 - u^2/c^2)}}. \tag{8.6}$$

γ_u is called the *Lorentz factor*. The transformation ensures that

$$s^2 = c^2t^2 - x^2 - y^2 - z^2 = c^2t'^2 - x'^2 - y'^2 - z'^2 \tag{8.7}$$

so that (8.4) is an invariant, even when the space-time interval is not a differential form, the separation between O and P being as big as we please. And now we see why the + and − signs are needed.

Some of the amazing results that follow from the Lorentz transformation equations were noted in Book 2 Section 7.2. Perhaps they seemed unbelievable at the time – especially as you hadn't studied any physics. But now you know something about mass and energy we can start to connect all these strange ideas together; and you'll get some even bigger surprises. First, however, you'll have to do a little bit more mathematics: after all we're going from 3-space to 4-space and that's a big jump. A hundred years ago the cleverest people in the world were only just beginning to think about it.

8.2 Vectors in 4-space

First let's remember again what (8.4), or its differential form (8.3) really means. A typical 'event' can be some kind of signal or *disturbance*, which travels through space with a certain speed, which we've called c; it might start at point O and move away in all directions, like the ripples on a pond when you throw a stone into it, arriving at point P (and many others) at some time t. Each observer has a clock and the two clocks are set to the same time (or 'synchronized') so that when Frame 2 is just passing Frame 1 (O′ coinciding with O) they show the same zero time, $t' = t = 0$. And each observer has, we suppose, reliable instruments for accurately measuring distances. So each observer can record the values of the coordinates and times, *relative to his own frame*, at which the event – the arrival of the signal at P – takes place.

Each observer thinks his own time is 'right' – after all they have 'perfect' clocks and they were set to agree at the start of the experiment; but what does that mean? The fundamental invariant is often written as

$$\mathrm{d}s^2 = -\mathrm{d}x^2 - \mathrm{d}y^2 - \mathrm{d}z^2 + c^2\mathrm{d}t^2 = c^2\mathrm{d}\tau^2 = \text{invariant.} \tag{8.8}$$

By writing the invariant as $c^2\mathrm{d}\tau^2$, we simply introduce a *proper time* interval ($\mathrm{d}\tau$) between the two events. So if a clock is fixed in any frame it will not move relative to the frame; and in any small interval we can therefore take $\mathrm{d}x = \mathrm{d}y = \mathrm{d}z = 0$, finding $\mathrm{d}t = \mathrm{d}\tau$. The 'proper time' for any observer is the time he reads on his own fixed clock. But this does not mean that the time t' at which an event is observed in Frame 2 (i.e. by Observer 2) is the same as the time t recorded by Observer 1; because both Observer 2 and his clock are *moving* with velocity u relative to Frame 1 and the times recorded will therefore be related by the Lorentz transformation (8.5). In particular

$$t' = \gamma_u\left(t - \frac{u}{c^2}x\right) = \gamma_u t\left(1 - \frac{u^2}{c^2}\right) = \gamma_u t/\gamma_u^2 = t/\gamma_u$$

– since the moving clock (fixed in Frame 2) is now at the point with $x = ut$. Thus, the times at which the event takes place are t for Observer 1 and t' for Observer 2, related by

$$t = \gamma_v t' \tag{8.9}$$

– the proper (or 'local') time for an observer in Frame 2 must be multiplied by γ_u to get the time for one in Frame 1. Since γ_u is always greater than 1, it will always appear to Observer 1 that things happen happen later (t larger) in the moving frame than they 'really' do (as shown on the local clock).

One outcome of the time relationship (8.9) will seem very strange. If one of two twins travels at enormous velocity in a spacecraft (Frame 2) and returns home after 10 years to the other twin, who never left Frame 1, they may find it hard to recognise each other. The 'travelling twin' will say he has been away only 10 years (by his clock); but, if the speed u is big enough to make $\gamma_u = 2$, the 'stay-at-home' twin will say it was 20 years – and he will have *aged* by 20 years, because everything that goes on in living material in Frame 1 will be going on at the same rate as the clock fixed in Frame 1. Of course, this is not an experiment you could actually *do* because $\gamma_u = 2$ would require the speed u to be unbelievably large (how large?– given that $c \approx 3 \times 10^8\,\mathrm{m\,s^{-1}}$); and the velocity would have to be uniform and rectilinear to satisfy Einstein's assumptions (Book 2, p.60). All the same, many experiments have been made, with smaller velocities and very accurate clocks, and all confirm the equations of this Section.

Now let's get back to the ideas of mass and motion. In Newton's second law, the *mass* of a particle enters as a proportionality constant relating the rate of change of its velocity to the *force* applied to it. The mass is a *property* of the particle and is taken to be a constant. In relativity theory, things are different, but it is still supposed that any particle has a property called its **proper mass** or **rest mass**, which 'belongs' to it and will be denoted by m_0. This rest mass, carried along with the particle, is taken to be an *invariant*, independent of the frame in which observations are made. How should we relate it to the coordinates (x, y, z) and velocity components (v_x, v_y, v_z) when we go from 3-space to 4-space? Note that the *particle velocity* may not be the same as the frame velocity so a different letter is used for it (v, not u).

First we have to learn how to use vectors in 4-space. With any displacement of a particle in 4-space we can associate a **4-vector**, whose 'length' squared now includes a time component as in (8.3):

$$\mathrm{d}s^2 = -\mathrm{d}x^2 - \mathrm{d}y^2 - \mathrm{d}z^2 + c^2\mathrm{d}t^2.$$

But what about the vector itself? In Book 2, and also in the present book, distances and lengths of vectors have always been expressed in terms of **cartesian components** along perpendicular axes: in 3-space, for example, the squared length of a vector r with components x, y, z is given by $r^2 = x^2 + y^2 + z^2$. Here, instead, there are some minus signs and if we tried taking $(-\mathrm{d}x, -\mathrm{d}y, -\mathrm{d}z, c\mathrm{d}t)$ as the components it just wouldn't work: the sum-of-squares form would contain only *positive* terms, giving $\mathrm{ds^2 = dx^2 + dy^2 + dz^2 + c^2dt^2}$ – which is not what we want.

The 4-vector components can, however, be chosen in various other ways to give us the correct invariant $\mathrm{d}s^2$. The simplest one is to introduce, along with the first three ('spatial') components, an extra factor i – the 'imaginary unit' with the property $i^2 = -1$, which you met long ago in Book 1. Remember that measurements always give *real* values and that the components only give a way of getting those values. So let's take as the components of the 4-vector with length $\mathrm{d}s^2$ the 'complex numbers' (which contain somewhere a factor i)

$$-i\mathrm{d}x_1 = -i\mathrm{d}x, \quad -i\mathrm{d}x_2 = -i\mathrm{d}y, \quad -i\mathrm{d}x_3 = -i\mathrm{d}z, \quad \mathrm{d}x_4 = c\mathrm{d}t \qquad (8.10)$$

and note that the sum of squares now gives the right value for $\mathrm{d}s^2$:

$$\mathrm{d}s^2 = \mathrm{d}x_1^2 + \mathrm{d}x_2^2 + \mathrm{d}x_3^2 + \mathrm{d}x_4^2 = -\mathrm{d}x^2 - \mathrm{d}y^2 - \mathrm{d}z^2 + c^2\mathrm{d}t^2.$$

Sometimes a 4-vector is indicated just by showing its four components in parentheses: thus the infinitesimal interval with squared length (8.3) would be

$$\mathsf{ds} \to (-i\mathrm{d}x_1 \quad -i\mathrm{d}x_2 \quad -i\mathrm{d}x_3 \quad \mathrm{d}x_4). \qquad (8.11)$$

This is the first of a number of important 4-vectors: all have a similar form, the first three components behave like those of a displacement vector in 3-space (e.g. on rotating the frame by changing the directions of the x- y- and z-axes), but the fourth is a scalar (not depending on axial directions).

We can get other 4-vectors, all with invariant lengths, by multiplying the components in (8.10) by any other invariant quantities, for example the proper time interval ($\mathrm{d}\tau$) or the proper mass (m_0). First think of the velocity of a particle: it will have three components $v_x = \mathrm{d}x/\mathrm{d}t$, $v_y = \mathrm{d}y/dt$, $v_z = \mathrm{d}z/\mathrm{d}t$ for an observer in Frame 1, and has been called by the letter v, because it has nothing to do with the u used for the speed along the x-axis of Frame 2 relative to Frame 1. It is a *local* velocity whose x-component, for example, is the limit of the ratio of two small quantities, the displacement ($\mathrm{d}x$) of the particle and the time taken ($\mathrm{d}t$) – all measured by Observer 1 in Frame 1. How will this particle velocity look to Observer 2 in Frame 2?

To answer this question we must start from the *invariant* interval, whose components are shown in (8.10) and form the 4-vector (8.11). If we divide every component by the invariant time $d\tau$ corresponding to the time interval dt measured on the clock in Frame 1, we shall get a 4-vector with components (leaving out, for the moment, the factors $-i$ in the first three)

$$\frac{\mathrm{d}x}{\mathrm{d}\tau}, \frac{\mathrm{d}y}{\mathrm{d}\tau}, \frac{dz}{d\tau}, \frac{cdt}{d\tau}.$$

– and these will behave, on going from Frame 1 to Frame 2, just like those in the basic 4-vector (8.10); they will undergo a Lorentz transformation. But how can we express them in terms of velocity components like $v_x = \mathrm{d}x/\mathrm{d}t$, as defined above? Clearly, we need an expression for $\mathrm{d}\tau$ in terms of Frame 1 quantities. This comes from the definition (8.8), namely $c^2\mathrm{d}\tau^2 = c^2\mathrm{d}t^2 - \mathrm{d}x^2 - \mathrm{d}y^2 - \mathrm{d}z^2$, which gives (note that this is just the ratio of two squares – nothing has been differentiated!)

$$\frac{\mathrm{d}\tau^2}{\mathrm{d}t^2} = 1 - \frac{1}{c^2}\left(\frac{\mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2}{\mathrm{d}t^2}\right) = (1 - v^2/c^2),$$

where $(\mathrm{d}x/\mathrm{d}t)^2 = v_x^2$ and the sum of three similar terms gives the squared magnitude of the particle velocity, v^2.

Now we have $\mathrm{d}\tau$ as a function of $\mathrm{d}t$, namely $\mathrm{d}\tau = \sqrt{(1 - (v^2/c^2)}\mathrm{d}t$, we can find the *relativistic* velocity components that will appear in the velocity 4-vector. A common convention is to name the 4-vector components with a capital (upper-case) letter, so they won't get mixed up with the ordinary 3-vector components (shown in lower-case letters as v_x etc.). With this notation, the first three 4-vector components (still without the $-i$ factors) will be $V_x = \mathrm{d}x/\mathrm{d}\tau$, $\quad V_y = \mathrm{d}y/\mathrm{d}\tau$, $\quad V_z = \mathrm{d}z/\mathrm{d}\tau$.

Let's get V_x, knowing the others will be similar: we'll do it by relating v_x to V_x, which is easier. Thus,

$$v_x = \frac{\mathrm{d}x}{\mathrm{d}t} = \frac{\mathrm{d}x}{\mathrm{d}\tau}\frac{\mathrm{d}\tau}{\mathrm{d}t} = V_x\frac{\mathrm{d}\tau}{\mathrm{d}t}. \tag{8.12}$$

Here v_x is a function of $\mathrm{d}t$ but is also a function of $\mathrm{d}\tau$, since $\mathrm{d}\tau$ is related to $\mathrm{d}t$ by $\mathrm{d}\tau = \sqrt{(1 - (v^2/c^2)}\mathrm{d}t$; and we are using what we know from calculus (Book 3, Chapter 3) to 'change the variable'. Thus

$$\frac{\mathrm{d}\tau}{\mathrm{d}t} = \sqrt{(1 - (v^2/c^2)} = 1/\gamma_v, \tag{8.13}$$

where γ_v is defined exactly like the Lorentz factor (8.6) except that now it contains the *particle* speed v instead of the the speed u of Frame 2 relative to Frame 1.

On putting this last result into (8.12) we find

$$v_x = \frac{\mathrm{d}x}{\mathrm{d}\tau}\left(\frac{1}{\gamma_v}\right) = V_x/\gamma_u. \tag{8.14}$$

Similar results follow for the y- and z-components of velocity, while the fourth (time) component in (8.10) will give

$$\frac{c\mathrm{d}t}{\mathrm{d}\tau} = c\gamma_v,$$

where we've remembered (Book 3, Section 2.4) that when $y = f(x)$ the derivative of x as a function of y (the 'inverse' function) is simply $\mathrm{d}x/\mathrm{d}y = (\mathrm{d}y/\mathrm{d}x)^{-1}$.

On adding the $-i$ factors to the first three (spatial) 4-vector components we finally get the velocity 4-vector

$$(-iV_1 \ \ -iV_2 \ \ -iV_3 \ V_4) = \gamma_v(-iv_x \ \ -iv_y \ \ -iv_z \quad c). \tag{8.15}$$

This particle velocity will behave, under a change of reference frame, just like the basic 4-vector (8.11) for the interval: its components will follow the standard Lorentz transformation.

8.3 Momentum and energy: $E = mc^2$ – a hope for the future?

In earlier chapters we soon discovered that, in pre-relativistic dynamics, the linear momentum vector p and the kinetic energy $E = \frac{1}{2}mv^2$ were very important quantities. The components of linear momentum of a particle moving with velocity v were simply

$$p_x = mv_x, p_y = mv_y, p_z = mv_z, E = \tfrac{1}{2}mv^2,$$

where of course $v^2 = v_x^2 + v_y^2 + v_z^2$. We now want to know what are the corresponding quantities for a very fast moving particle.

In the last section we saw how a new 4-vector could be obtained from a given 4-vector simply by multiplying its four components by any invariant quantity: in that way we got the *velocity* 4-vector (8.15) from the displacement 4-vector (8.11) on multiplying it by the reciprocal $(1/d\tau)$ of the proper time interval. The next invariant quantity we'll use is the proper *mass* m_0; and, since linear momentum is particle mass × velocity, we might expect that m_0 times the velocity 4-vector (8.15) will give us something interesting. Let's try it. The result is

$$\begin{aligned} & m_0(-iV_1 \ \ -iV_2 \ \ -iV_3 \ V_4) = \\ & \quad m_0\gamma_v(-iv_x \ \ -iv_y \ \ -iv_z \quad c) \\ = & \quad (-i\gamma_v m_0 v_x \ \ -i\gamma_v m_0 v_y \ \ -i\gamma_v m_0 v_z \quad \gamma_v m_0 c). \end{aligned} \tag{8.16}$$

The first three components are the ordinary (pre-relativistic) momentum components p_x, p_y, p_z, multiplied by the Lorentz factor γ_v (along with the usual imaginary factor $-i$, for spatial components); the fourth component is $m_0V_4 = \gamma_u m_0 c$). What does all this mean?

What Newton called m, the mass, now seems to be replaced by $\gamma_v m_0$ – the *rest* mass, multiplied by a factor depending on the particle speed v. So let's go on using m for this 'apparent mass', noting that when the speed is very small compared with c, m will become the same as m_0 (quite independent of the speed, just as Newton had supposed and experiments seemed to confirm). The relativistic momentum components can thus be written

$$-iP_1 = -imV_1, -iP_2 = -imV_2, -iP_3 = -imV_3, \quad P_4 = mc,$$

capital letters again being used for the components of the relativistic 4-vector. The first three components are all of 'mass × velocity' form, as we expected.

The fourth component, however, doesn't look like anything we've met before: it's simply mc. To interpret it, remember that in Newton's dynamics a moving particle had a **kinetic energy** (KE) of the form $\frac{1}{2}mv^2$. Can you guess what form it will take in the relativistic theory? Perhaps, like the momentum components, it will change only because Newton's mass m must be replaced by the apparent mass $m = \gamma_v m_0$, which depends on how fast it's going? It's easy to test this idea. Putting in the value of γ_v, from (8.13), the apparent mass becomes

$$m = m_0\gamma_v = \frac{m_0}{\sqrt{(1 - v^2/c^2)}} = m_0\left(1 - \frac{v^2}{c^2}\right) = m_0\left(1 + \tfrac{1}{2}\frac{v^2}{c^2} + ...\right),$$

where we've expanded the square root, using the binomial theorem (Book 3 Section 3.1). On throwing away the negligible terms (represented by the dots) this can be written $m = m_0 + (\frac{1}{2}(m_0v^2)/c^2$ or, multiplying by c^2,

$$mc^2 = m_0c^2 + \tfrac{1}{2}m_0v^2. \tag{8.17}$$

Now $\frac{1}{2}m_0v^2$ is the KE of a particle of mass m_0 moving with speed v. So what we've discovered is that the quantity m_0c^2 (mass times velocity squared – which has the dimensions of *energy*, $\mathrm{ML^2T^{-2}}$) is increased by the amount $\frac{1}{2}m_0v^2$ when the particle is moving. When the particle is *not* moving, relative to the observer, the KE term in (8.17) is zero and $m \rightarrow m_0$; but the energy term m_0c^2 never disappears – it is called the **rest energy** of a particle of rest mass m_0 and was discovered by Einstein, who first wrote down the equation

$$E = m_0c^2 \tag{8.18}$$

– perhaps the most famous equation of the last century. What it tells us is that any bit of mass is exactly equivalent to a certain amount of *energy*; and because c is so large ($\approx 3 \times 10^8$ m s^{-2}) that energy will be enormous. One teaspoonful, for example, holds perhaps 10 grammes of water (mass units) – but $10 \times (3 \times 10^8)^2 = 9 \times 10^{14}$ kg m^2 s^{-2}=$9{\times}10^{14}$ Joules of *energy*; and that's enough to boil more than 200 million kg of ice-cold water (roughly 200 million litres)! And that's where we are today: if only we could get the energy out of matter, where it's locked away in the form of mass, there would be enough for everyone in the world. You'll come back to such problems when you know what 'matter' *is* – what's it made of? This is one of the great questions we meet in Book 5.

Before stopping, however, note that the third 4-vector we have found holds nearly all we need to know about the dynamics of a moving particle: it is

$$(-iP_1 \ -iP_2 \ -iP_3 \ E/c) \tag{8.19}$$

– where the first three components are just like the x- y- z-components of momentum in pre-relativity times, except that the particle mass is $m = \gamma_v m_0$. The fourth component is shown in energy units; and now we know that $E = mc^2$ it corresponds to $P_4 = E/c = mc$ – agreeing with what we found above. The important vector (8.19) is called the **energy-momentum 4-vector**. Many of the principles we discovered in 'classical' (pre-relativistic) dynamics still apply to very fast moving particles, as long as you remember that the mass m is not the same as for a particle at rest. So we find momentum and energy conservation laws need very little change. These results haven't been *proved* here, but they *can* be proved and you can take them on trust. The very small changes are not noticible unless the particle speed v is enormous; but we've already noted that the constant c will turn out (in Book 10) to be the speed of light. There is a natural limit to how fast anything can go; and now we can see what will happen when the speed of a particle gets closer and closer to that limit. When $v \to c$ in the Lorentz factor γ_v, the mass $m = \gamma_v m_0$ gets bigger and bigger, going towards the limit $m_0/0$, infinity! The faster it goes the heavier it gets, until finally nothing can move it faster.

Looking back –

You started this book knowing nothing about Physics. Where do you stand now?

Building only on the ideas of number and space (Books 1 and 2) and simple mathematical relationships (Book 3), you've come a long way:

- In Chapers 1 and 2 you've learnt about building physical *concepts* from your own experience of pulling and pushing, working and using you energy. You know about force, mass, weight, and how things move; and about Newton's famous laws. You've learnt that energy is *conserved*, it doesn't just disappear – it can only change from one kind to another.

- Chapter 3 extended these ideas to the motion of a **particle** (a 'point mass'), acted on by a force and moving along any path. Energy is still conserved. You learnt how to calculate the path of the Earth as it goes round the Sun, using the same simple laws that worked for a small particle. Amazing that it came out right, predicting a year of about 360 days!

- In Chapter 4 you found out how that could happen, by thinking of a big body as a collection of millions of particles, and using Newton's laws. You learnt about the **centre of mass**, which moves as if all the mass were concentrated at that one point; and about **momentum** and **collisions**.

- Chapter 5 showed how you could deal with **rotational motion**. You found new laws, very much like Newton's laws, and met new concepts – 'turning force', or **torque**, and **angular momentum**. And from the new laws you were able to calculate the orbits of the planets.

- In Chapters 6 and 7 you've begun to study the **Dynamics** and **Statics** of a **rigid body**; and the construction of simple machines. You're well on the way to the Engineering Sciences!

- The final Chapter 8 brought you to the present day and to the big problems of the future. You found that **mass was a form of energy** and that *in theory* a bottle of seawater, for example, could give enough energy to run a big city for a week! – if only we could get the energy out! This is the promise of **nuclear energy**.

We all need energy in one way or another: for transporting goods (and people), for digging and building, for running our factories, for keeping warm, for almost everything we do. At present most of that energy comes from burning fuel (wood, coal, oil, gas, or anything that will burn); but what would happen if we used it all? And *should* we go on simply burning these precious things (which can be used in many other ways) until they're finished. If we do, what will our children use? Another thing: burning all that stuff produces tons of smoke, which goes into the atmosphere and even changes the world's climate – always for the worse!

We probably have to solve such problems before the end of this century: how can we do it? Do we go back to water-mills and wind-mills, or do we turn to new things, like trying to trap the energy that comes to us as sunlight – or getting the energy out of the atom? Nuclear power is being used already in many countries; but it brings new problems and many dangers. To understand them you'll have to go beyond Book 4.

In Book 5 you'll take the first steps into Chemistry, learning something about atoms and molecules and what everything is made of.

Index