G & EM, The Fundamental Forces of Everyday Life: I.

1: Introduction

We all know that gravitation and electromagnetism are important. Gravity holds us on the Earth as well as holding the air, water and earth and rocks of Earth, itself, together and keeps the Earth in orbit around the Sun, the source of most of our energy. Electromagnetism lights our homes, runs our TV's and computers and radios and phones and electric motors, etc., etc. and may, depending on your choices, heat your home and cook your food. But what about all the other forces we encounter in daily life?

What about the impenetrability of solid objects like rocks or pieces of metal? What about the buoyancy of fluids and the pressure within fluids? What about the friction of one surface sliding over another or the viscosity encountered in moving through oil or mud? What about the forces of hurricanes and tornados and volcanos or the forces that produce morning fog and mist? Do FORCES produce morning fog and mist?! What about the surface tension that allows insects to walk on water and children to blow bubbles?

It is one of the great successes of physical science that we have learned to understand all of these cases and many more as being consequences of the action and interplay of just gravitational and electromagnetic forces!

Physical scientists are frequently accused of the most strident forms of reductionism encountered in modern intellectual communities. We are regarded as believing that *everything*, literally *EVERYTHING*, can be, and one day will be, reduced to, i.e., understood in terms of physical processes and forces. That there is nothing else in Existence but the physical world. In fact, many of us do not embrace such extreme views. I certainly don't. But it can not be denied that the impulse to reduce as much as possible in Existence to the physical, and as much as possible in the physical to a small corner of the physical has enjoyed some impressive and valuable successes. This course is about one such success.

As with all big claims, there is a catch! Gravitation and electromagnetism couldn't do all I will claim for them without **quantum mechanics**. And this course is not about quantum mechanics. So the contributions of quantum mechanics to our understanding will have to take the dissatisfying form of my saying, here and there, " – and according to quantum mechanics such and such is the case!". The fourth course in this series "The Quantum World", addresses these issues and those who are interested in that aspect of our present topic can dig deeper. For now, let me just say that while quantum mechanics is usually thought of as bizarre and exotic, our everyday world would be unrecognizably different without it – and we wouldn't be in it!

2: A Brief Account of the Evolution of the Concept of Physical Force

Our conjectures are that primitive humans recognized only forces generated by beings and spirits and life forms. Even the astounding ancient Greeks, for the most part, divinized and enlivened the forces of nature. For Plato all of nature was a living organism and so the forces in nature were strivings for goals. Even the more practical minded Aristotle saw falling bodies and rising flames as objects seeking their natural place in the universe while collisions and projectiles were instances of violence done to the natural order of things. Only with Democritus and the atomists and their effort to reduce all phenomena to the motions and collisions of inert, lifeless atoms do we see some hint of the modern conception of a purely physical non-living force. But in those days even a higher fraction of the educated citizenry than today regarded this conception as intolerably bleak and alien. It did not catch on and by the time Archimedes presented a surprisingly modern treatment of the forces of levers and buoyancy he was ahead of his time in a dying civilization. Only the later civilizations of Byzantium and Islam would see the value of preserving his work.

Through the Middle Ages and the Renaissance a long arduous struggle with concepts of forces very gradually evolved. Slowly and erratically a branch moved in the direction of disentangling and identifying the possibility of pieces of non-living, material nature being sources of forces. If unobstructed, such forces would produce or change motion and, obstructed, could maintain stable structures. But, notwithstanding Europe's eventual recovery of Greek manuscripts, nothing as explicit as the preceding two sentences emerged from the voluminous writings of this thousand year long period!

Only in the 17th Century does the modern scientific conception of physical force begin to explicitly emerge. The German astronomer, Johannes Kepler, anticipates Newton in conceiving the possibility of precise quantitative and unified treatment of terrestrial and celestial attractions. He gets the details wrong, among other errors identifying the forces with celestial versions of the terrestrial magnetism discovered by Gilbert in the 16th Century. Nevertheless, his conception of physical force is a signal advance! Galileo discovers the connection of terrestrial gravity with his precisely defined acceleration. Huyghens successfully analyzes the acceleration involved in circular motion. Hooke correctly assesses the initial restorative forces of elastically distorted matter. Descartes corrects Galileo in identifying the kind of motion, so called inertial motion, which does not require force and, as the father of modern philosophy, sharply articulates the conceptual dualism between matter and spirit. And then, finally, Isaac Newton improves upon and pulls these threads together and formulates the conceptual, mathematical and empirically testable scheme of physical nature in his masterpiece, "The Principia", or "The Mathematical Principles of Natural Philosophy". The central concept throughout this work is the concept of physical force.

The clarification of the concepts and implications of the Principia occupied the physical scientists of the 18th Century and, while there were several aspects of the Principia which many found disturbing, no successful challenge to its quantitative accuracy could be found for over two hundred years! More than anything else, this continued success supported the confidence that humans could grasp the structure of the Universe and propelled the ever increasing advance of science! Finally, the *excessive* forms of this confidence (Newton had got it RIGHT and could not, fundamentally, be improved upon!) was shaken with the discovery of the limitations of the Newtonian scheme in the early 20th Century. Nevertheless, for almost all the physical forces of everyday life, and modulo a little help from quantum mechanics here and there, the pre 20th Century development of the Newtonian scheme is quite sufficient.

3: The Basic Rules for All Kinds of Everyday Forces

Forces always come in pairs!

(Rule 1) IF ANY PHYSICAL SYSTEM, A, EXERTS A FORCE ON ANY PHYSICAL SYSTEM, B, THEN (AT THE SAME TIME) THE SYSTEM, B, EXERTS AN EQUAL AND OPPOSITE FORCE ON THE SYSTEM, A.

This rule is traditionally known as **Newton’s third law**, and is often condensed into the rubric: *For every action there is an equal and opposite reaction.*

(Rule 2) SEVERAL FORCES APPLIED AT THE *SAME* POINT AT THE *SAME* TIME TO THE *SAME* SYSTEM HAVE EXACTLY THE SAME EFFECT AS IF ONLY ONE FORCE, CALLED THE **RESULTANT**, WAS APPLIED AT THAT POINT AT THAT TIME TO THAT SYSTEM.

This rule is rarely explicitly stated and has no traditional name. But how are **resultants** determined?

Well, first we have to have a way of representing the quantitative aspects of forces. There are two quantitative aspects of forces. How strong are they? In what direction do they push, or pull, or point ?

In this regard forces are just like arrows. Of arrows we can ask; how long are they and in what direction do they point? Consequently we can represent forces by arrows. The length of the arrows will be proportional to the strength of the forces represented and the direction of the arrows will either be the same as or will represent the direction of the forces.

Notice that other things can be represented by arrows. Changes in position, for instance, or velocities, or accelerations. The length of the arrow represents the magnitude of what is represented, how much of it there is, and the direction of the arrow represents the direction of what is being represented. Velocities, accelerations and changes in position all have directions. In the technical jargon, things that can be represented by arrows this way, as well as the representing arrows themselves, are called **vectors**. It turns out that resultants are important for sets of vectors of any kind.

Now suppose someone made a sequence of several position changes. There would be an arrow to represent each **displacement** in the sequence and an arrow to represent the *net* change in position resulting from the whole sequence, a *resultant* displacement arrow. That resultant arrow could be obtained from the arrows for the sequence by arranging the sequence arrows end to end and then connecting the beginning of the first arrow to the end of the last arrow. When we do this we notice that *the length and direction of the resultant arrow doesn't depend on the order in which we arrange the sequence arrows*. In other words, the resultant change in position doesn't depend on the *order* in which the position changes that comprise the sequence were made! It only depends on *what* those individual position changes were (See **Fig. I. 1**).

Well, forces combine in the same way! For a bunch of forces applied at the same point at the same time to the same system the arrow representing the resultant force is obtained by arranging the arrows representing the forces in the bunch end to end and then connecting the beginning of the first arrow to the end of the last arrow.

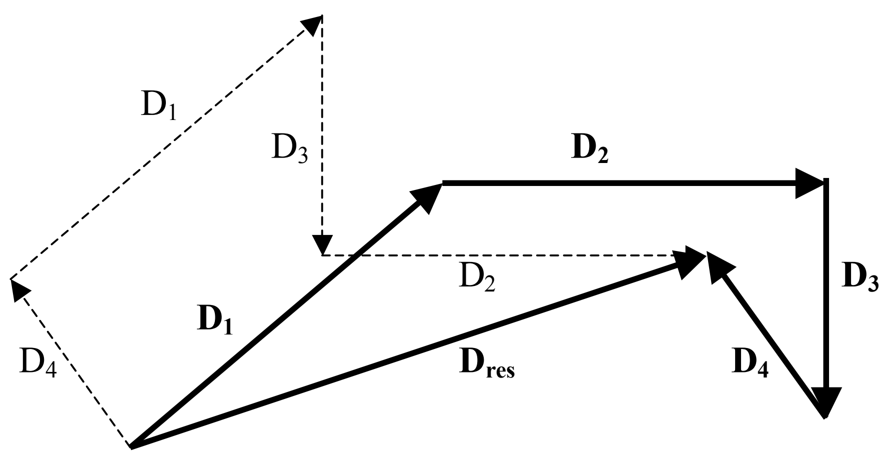


Fig. I. 1: Resultant of several displacements (or several forces). Note the independence of the resultant on the order of composition.

4: Components

Since any collection of forces applied at the same point and time to a system are equivalent to applying the single resultant of that collection, it follows that any single force is, in turn, equivalent to applying a collection of forces (at the same point, time and system) provided that the collection has the single force as its resultant. This conclusion gives rise to the useful concept of a *component* of a force in a given direction (**Fig. I. 2**).

For any force, **F**, and any direction, D, the force, $\mathbf{F}_D$, is called the **component** of **F** in the direction D if, and only if,

(1) $\mathbf{F}_D$ is parallel to the direction D and

(2) **F** is the resultant of $\mathbf{F}_D$ and another force perpendicular to $\mathbf{F}_D$.

As with resultants, any vector has components in any direction and they are always determined in the same way as for forces.

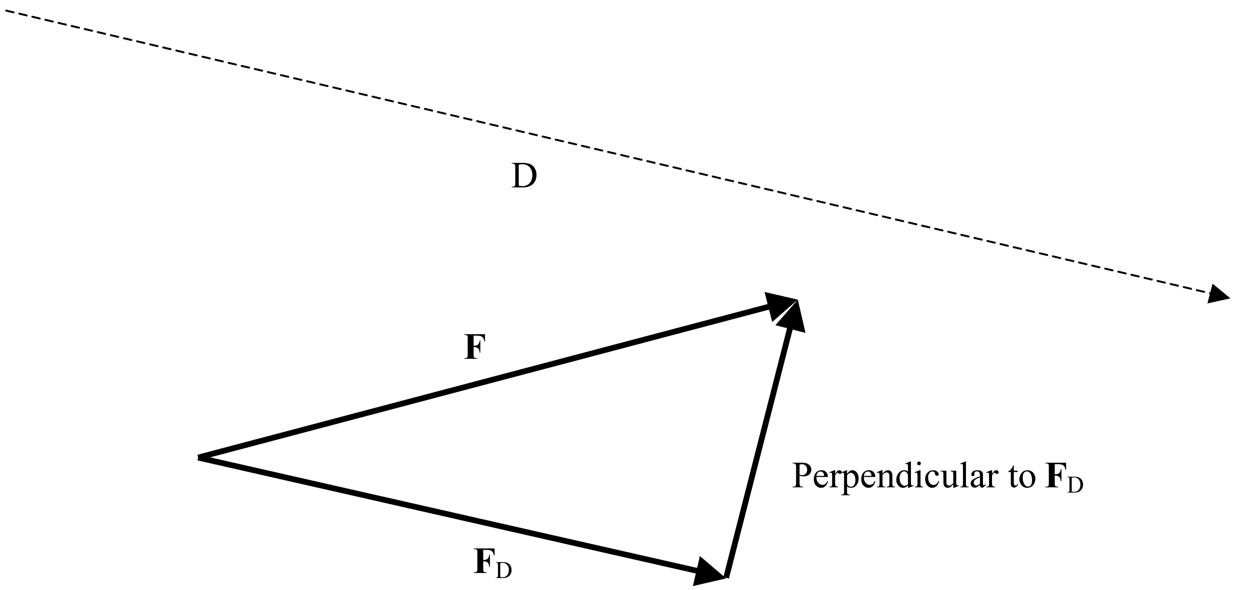


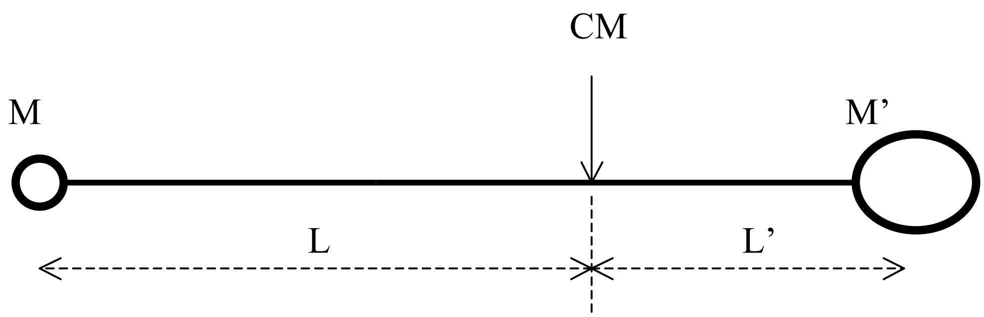
Fig. I. 2: Component, $\mathbf{F}_D$, of force, **F**, in the direction, D.

5: The general importance of resultants of forces

Even if a collection of forces are not coincident, i.e., even if the forces are not all applied at the same point, the resultant of the collection can still be important, not *equivalent* in effect to the collection, but *important*. The forces in the collection still have to be applied simultaneously and applied to the same system – but they don't have to be applied at the same point – for the resultant to be important.

The useful importance of such resultants depends on the concept of the **center of mass** of a material system (**Fig. I. 3**). For our present purposes, we'll not worry about how to locate, precisely, the center of mass (CM) of a system. We'll get into that later. For now it's enough to know that *the CM is a point, with respect to which the matter of the system is evenly distributed in some sense*. This always places the CM 'inside' the system, and, for geometrically symmetrical systems the CM is often coincident with what one might call the geometrical center of the system. For example, the CM of a sphere of uniform mass density, or a cube of uniform mass density, is at the geometrical center of the sphere or cube, respectively.

If the system is a 'rigid' solid, or even a ‘stiff’ elastic solid, the location of the CM, relative to the rest of the system, is highly stable and inflexible. On the other hand, the CM of highly flexible solids or fluids or gasses can move around extensively, within such systems, as the distribution of the matter in such systems changes.



$M < M'$ $L > L'$ $M L = M' L'$

Fig. I. 3: Center of mass of lopsided dumbbell

We're now ready to state the importance of the resultant of ALL the forces acting on a given system at the same time.

(Rule 3) IF THE RESULTANT OF ALL THE FORCES ACTING ON A SYSTEM SIMULTANEOUSLY IS ZERO, THEN THE MOTION OF THE CM OF THAT SYSTEM WILL NOT BE ACCELERATED. IF THE RESULTANT IS NON-ZERO, THEN THE ACCELERATION OF THE CM WILL BE IN THE DIRECTION OF THE RESULTANT AND PROPORTIONAL TO THE RESULTANT'S MAGNITUDE AND INVERSELY PROPORTIONAL TO THE SYSTEM'S MASS.

This rule is a slight variation of what is traditionally known as **Newton’s first law** (resultant = 0) and **Newton’s second law** (resultant =/= 0). It is expressed by the equation,

$$\mathbf{F} = M\,\mathbf{A}.$$

Here **F** is the resultant force (a vector, remember), **A** is the acceleration of the CM (also a vector since it has magnitude and direction) and M **A** is the product (as in multiplication) of the mass, M, of the system and **A**. The product (also a vector) has the same direction as **A** and a magnitude equal to the product of M and the magnitude of **A**.

6: Gravity

Newton's theory of gravity was the first enormously successful theory of a basic force of nature. It was superceded in the early twentieth century by Einstein's theory of gravity called "General Relativity" but in this course we will almost never need to consider Einstein's improvement. This is good since Einstein's theory is much more complicated in detail than Newton's.

Newton's theory is expressed in terms of the gravity of the tiniest particles of matter.

(Newton's Theory of Gravity) EVERY PARTICLE OF MATTER IN THE UNIVERSE ATTRACTS EVERY OTHER PARTICLE OF MATTER WITH A FORCE DIRECTLY PROPORTIONAL TO THE PRODUCT OF THE MASSES OF THE PARTICLES AND INVERSELY PROPORTIONAL TO THE SQUARE OF THE DISTANCE BETWEEN THE PARTICLES (**Fig. I. 4**).

For two particles with masses, m_1 and m_2, and separated by the distance, d_{12}, the magnitude of the force of mutual attraction is given by

$$F_{12} = F_{21} = G\, m_1 m_2 / (d_{12})^2 \ ,$$

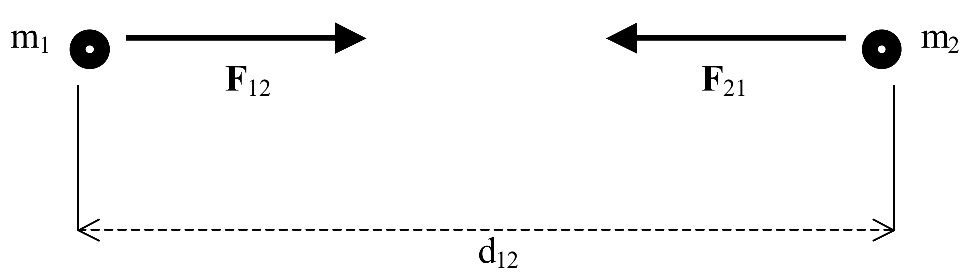


Fig. I. 4: Forces of mutual gravitational attraction between masses

Where G, a proportionality constant, is, historically, the first **universal constant of nature**, Newton's gravitational constant. It took awhile before people were able to determine its value.

Taking $\mathbf{F}_{12}$ as the gravitational force exerted *on* particle 1 *by* particle 2, and similarly for $\mathbf{F}_{21}$, we have $\mathbf{F}_{12}$ pointing *from* particle 1 *to* particle 2 and, from **(Rule 1)**,

$$\mathbf{F}_{21} = - \mathbf{F}_{12} .$$

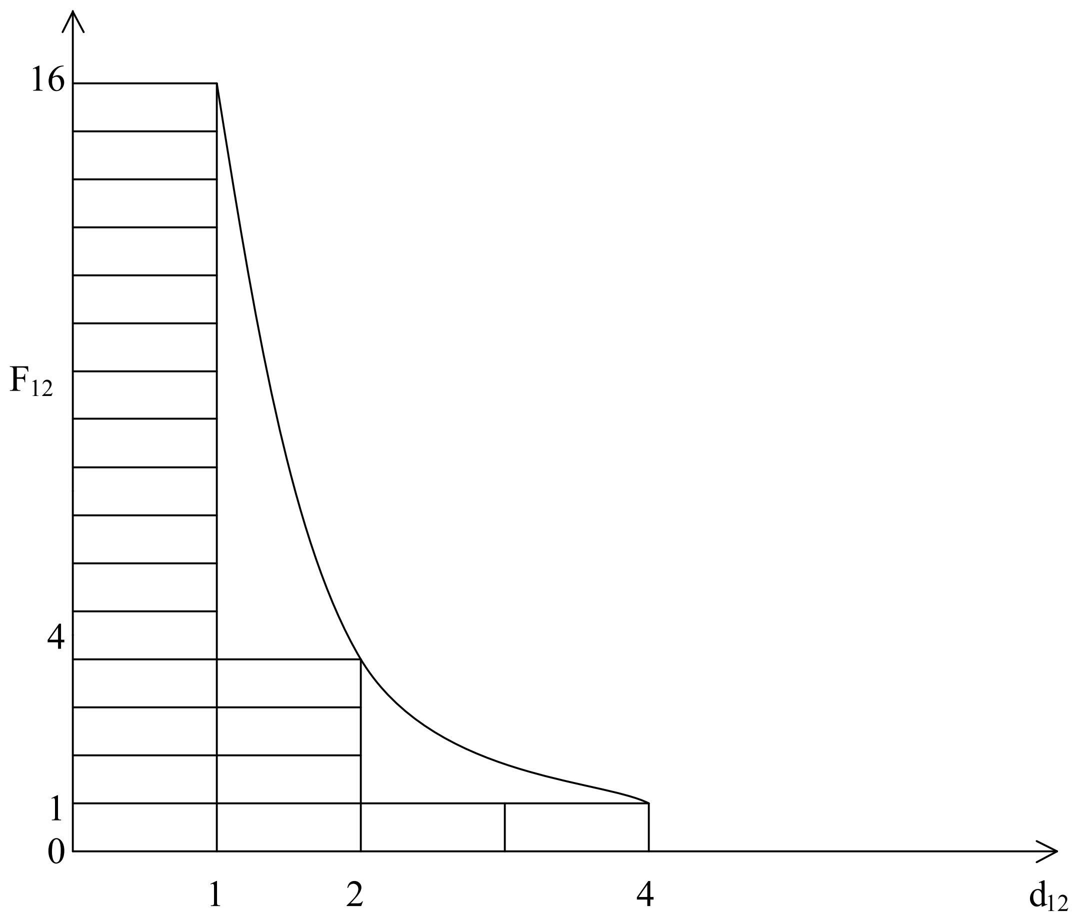


Fig. I. 5: Variation of the magnitude of gravitational force between particles with the distance between the particles

As we can see from the graph (**Fig. I. 5**) of the magnitude of the attraction varying with the distance between the particles, the force can get as large as you like if the particles get close enough together and the force gets as weak as you like if the particles are far enough apart.

The gravitational force on one ordinary object with size and shape due to another such object is obtained as the resultant of *all* the forces acting on *all* the particles of the one object due to *all* the particles of the other object. In general, determining such a resultant can be a complicated mathematical problem. But there are some interesting and surprising and important simple cases! One of the more interesting and surprising cases (**Fig. I. 6**) is that

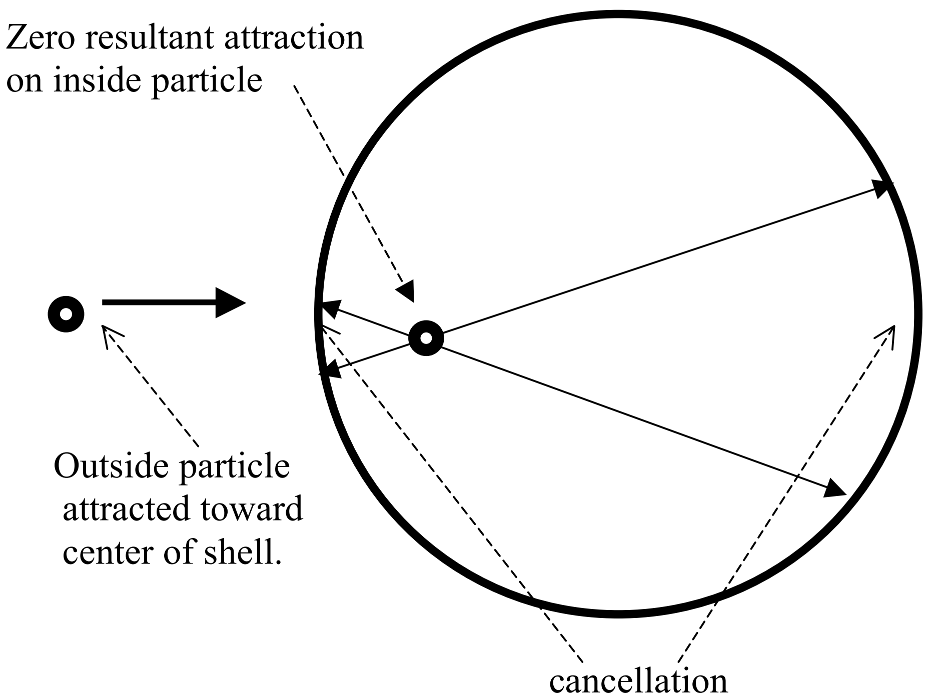


Fig. I. 6: Gravitational attraction of a uniform shell on a particle outside and inside the shell.

IF A THIN SPHERICAL SHELL HAS ITS MASS UNIFORMLY DISTRIBUTED OVER IT, THEN THE GRAVITATIONAL FORCE EXERTED BY THE SHELL ON A PARTICLE OUTSIDE THE SHELL IS THE SAME AS IF ALL THE MASS OF THE SPHERICAL SHELL WERE CONCENTRATED AT ITS CENTER. IF THE PARTICLE IS INSIDE THE SHELL THE FORCE IS ZERO!

The reason for this surprising result is that one gets exact cancellation of the gravitational forces pulling in opposite directions on any inside particle. The fact that some parts of the shell are closer to the inside particle than others is compensated by there being more shell matter attracting in the distant parts than in the near parts, and in just the right amounts to yield a zero resultant!

Imagine being inside a vast massive spherical shell. It doesn't attract you at all! You float free in zero gravity. But then you float through a tiny hole in the surface to the outside. Immediately, all the great mass of the shell attracts you towards its center, holding you fast to its surface if you don't slip back through the hole, as if all the mass of the shell were concentrated at its center.

It turns out that this erie fantasy is useful in determining how the Sun, the Earth, the Moon and the planets attract objects in their vicinity.

All of these objects are roughly spheres. Not exactly, but it's a decent first approximation. The mass inside these spheres is not uniformly distributed. We now know this, but even in the early days of modern science there was no good reason to assume otherwise. But it's more reasonable to assume that, roughly speaking, the mass near any given radius from the center of these objects is uniformly distributed over the spherical shell at that radius. This would be what is called a spherically symmetric mass distribution for the whole object. Distortions due to the rotation of the objects would be the main correction to this assumption. To the degree that this assumed symmetry is a good approximation, since each spherical shell attracts particles (outside itself) as if its mass is at its center, the whole spherical object, will gravitationally attract particles (outside itself) as if its entire mass were concentrated at its center!

In other words, to a good first approximation, the Sun, Earth and Moon attract each other and you and me as if they were tiny massive particles rather than enormous spheroids! Newton knew the data fit this conclusion very early in the game of building his theory (~ 1666). But he didn't know how to derive the conclusion from his theory until twenty years later (1686). The "Principia" was published in 1687 and included Newton's calculations of the major corrections to the conclusion (rotational deviations from sphericity and tidal effects).

Note also that, to the extent the spherically symmetric mass distribution is a good approximation, descending into deep mines will *decrease* the Earth's gravitational attraction in a definite way since only those spherical shells still below the mine will contribute to the attraction! This has been tested and found to corroborate the theory.

Next week we'll examine what led Newton to his theory.

Appendix: Speed versus Time Graphs*

Newton's inspiration for his theory of gravity came from several sources. (1) Galileo's studies of the motion of falling objects and projectiles, (2) Kepler's studies of planetary motions and (3) wide spread conjecture and theorizing about gravity which, before Newton, no one confronted with hard data.

So understanding Galileo's discoveries concerning the description of motion is our first serious order of business and that depends crucially on the graphical representation of motion. In fact I doubt Galileo could have made the discoveries without his use of graphs.

The graphs we're interested in are Speed versus Time graphs. They are supposed to represent how the speed of some moving object varies with time. To keep things simple, at first, we will assume our object is always moving in a constant direction.

The **speed** is represented on the vertical axis and is labeled by the letter, v. Yeah, I know, why not represent speed by the letter, s? Well, while *sometimes* s is used, v (for magnitude of *velocity*) is used much more frequently and so you should get used to v for speed. Anyhow, speed is on the vertical axis and represented by v. Time is represented on the horizontal axis and is sensibly labeled by the letter, t. The passage of time is represented by moving to the right along the horizontal axis.

In the discussion that follows, the first goal is to make it plausible to you that *for any time interval represented in a speed versus time graph, the* ***area*** *between the speed curve and the time axis is proportional to the* ***distance*** *covered during that time interval.*

In our first graph, **Fig. IA. 1**, the curve represents continuously varying speed with time between the times, t_1 and t_2. The magnitude of the speed is represented by the height of the curve from the horizontal axis. The higher the curve is from the horizontal axis at any particular time, the faster the object is moving at that time. If the curve were ever to touch the horizontal axis, that would indicate zero speed, or no motion of the object, at the time corresponding to where the curve touched the horizontal axis. In **Fig. IA. 1** that never happens. Between the times, t_1 and t_2 our object is always moving.

The arrows extending from the box labeled "Maximum speed" indicate the value, v_M, of the maximum speed on the vertical speed axis, the place on the curve where the maximum speed is reached and the moment of time, t_M, on the horizontal time axis at which the maximum speed occurs. Can you find the corresponding points for the minimum speed?

The motion represented in the graph of is almost always changing speed. The speed versus time curve is almost always climbing or dropping. Such motion with changing speed is called **accelerated motion**. Even when the speed is decreasing and we would ordinarily say it was *decelerating*, the proper technical expression is that it is *accelerating* - - *negatively*. We will discuss acceleration in these graphs more closely later.

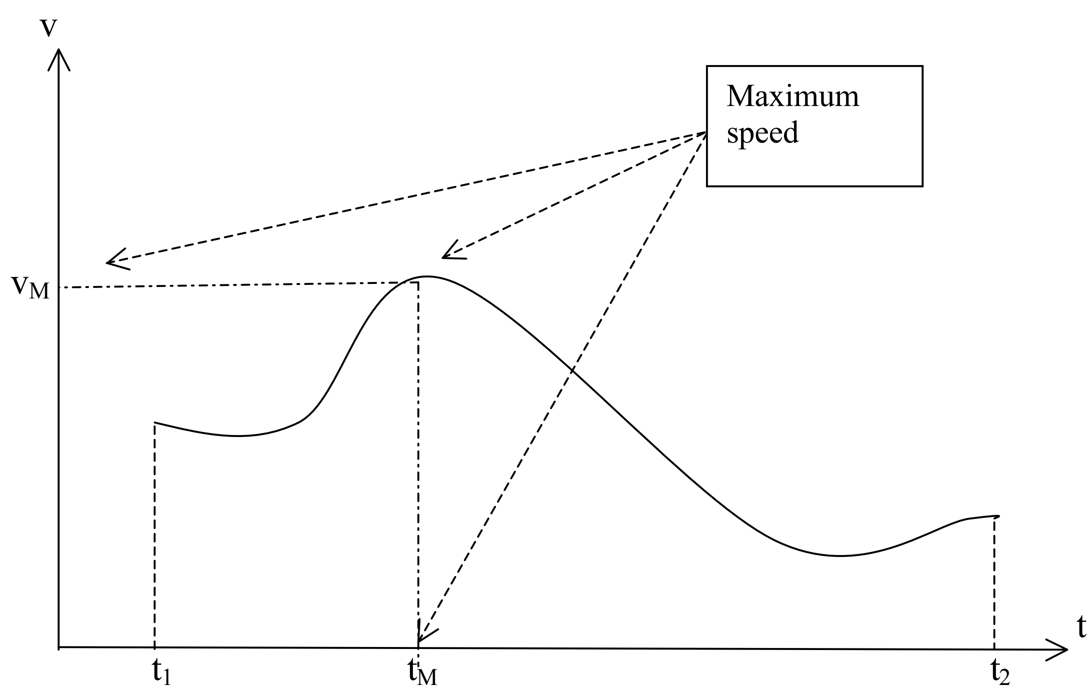


Fig. IA. 1: Example of speed versus time graph.

But to best learn how to read these graphs – to really plumb their depths – we need to consider graphs for motion with constant speed. **Fig. IA. 2** is a graph for motion with constant speed, v_1, during the time interval, Δt_1, constant speed, v_2, during the time interval, Δt_2, and then v_3 during Δt_3. Such a graph is called *stepwise constant*.

Now when the speed is constant it's very easy to determine the distance covered during the motion. If we drive down a straight road at a speed of 60 mi/hr for 20 min, the distance covered is given by multiplying the speed by the time, i.e.,

$$\text{Distance} = 60 \text{ mi/hr} \times 20 \text{ min} = 60 \text{ mi/hr} \times (1/3)\text{hr} = 20 \text{ mi.}$$

[In this simple example we get to see conversion of units (from minutes to hours) and cancellation of units (hours in the denominator of 60 mi/hr cancelled by hours in the numerator of (1/3)hr).]

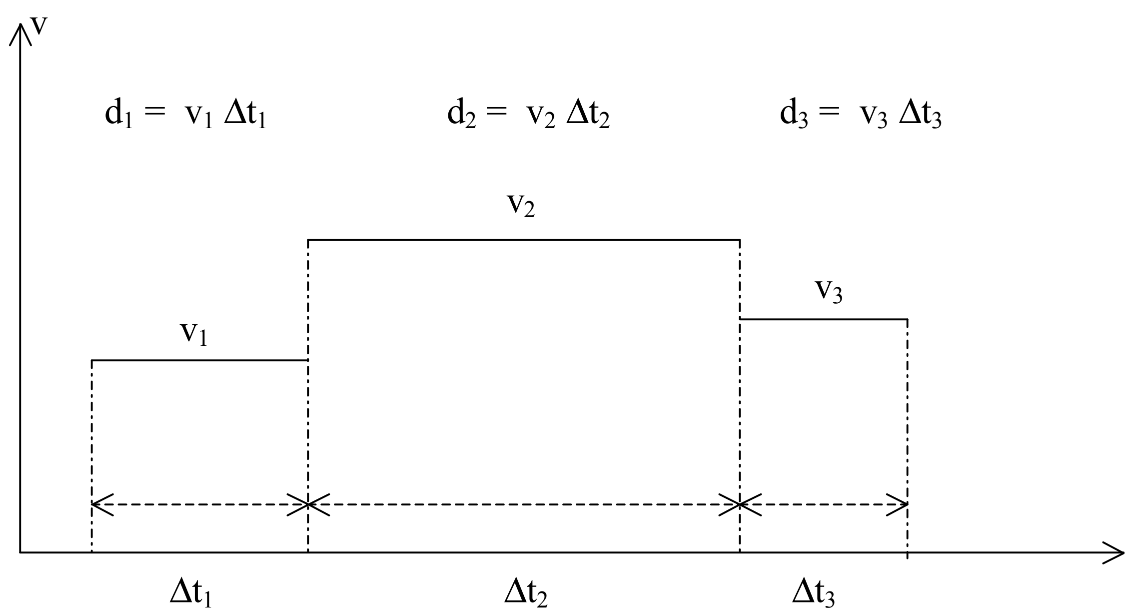


Fig. IA. 2: Stepwise constant speeds and the distances covered

The multiplication of constant speed by duration of travel always gives the distance covered and the formula or equation for this is,

$$d = v \Delta t .$$

But now look back at the figure! The value of v is represented by the *height* of the horizontal speed curve from the time axis and the duration of travel is represented by the interval of the time axis under the speed curve, i.e. by the *width* of the rectangle under the speed curve. So multiplying the speed v by the duration Δt corresponds to multiplying the height by the width of the rectangle under the speed curve. But that's just the *area* of the rectangle! *The distance covered is proportional to the area under the speed curve* **- -** *at least for constant speed*. The area of the rectangle under v_2 is obviously larger than that of the rectangles under v_1 or v_3. So more distance is covered during Δt_2 than is covered during Δt_1 or Δt_3. But how does the distance covered during Δt_3 compare with that covered during Δt_1? It's a bit harder to say, but the answer is determined by comparing the areas of the rectangles under v_3 and v_1.

Now - - back to accelerated motion. What, if anything, can the area under a speed curve for accelerated motion tell us? Well, first of all, notice that if our object behaved as in **Fig. IA. 2**, first moving with speed v_1 for Δt_1, then with speed v_2 for Δt_2, and finally with v_3 for Δt_3, we would get the *total* distance covered by adding up the areas of the three rectangles. Now, in **Fig. IA. 3** we compare the curve of **Fig. IA. 1** with a carefully chosen stepwise constant speed curve similar to that in **Fig. IA. 2**.

The stepwise constant speed sequence is chosen to never differ in speed from the continuous curve by very much. *If* the actual motion was that of the stepwise constant speed curve, *then* the total distance covered would be proportional to the sum of the areas of the rectangles under the constant speed segments. But notice that if we consider more and more numerous sequences of constant speed segments that are shorter and shorter in duration while deviating from the continuous curve by less and less, then (1) it would be more and more difficult to distinguish the continuously varying motion from the jumping sequence of very short constant speed segments and (2) the total area of the rectangles under the constant speed sequence would get closer and closer to equaling the area under the continuous curve. It thus appears as though ***we can assert the total distance covered to be proportional to the area under the speed curve in all cases*** and not just for the case of a stepwise constant speed sequence.

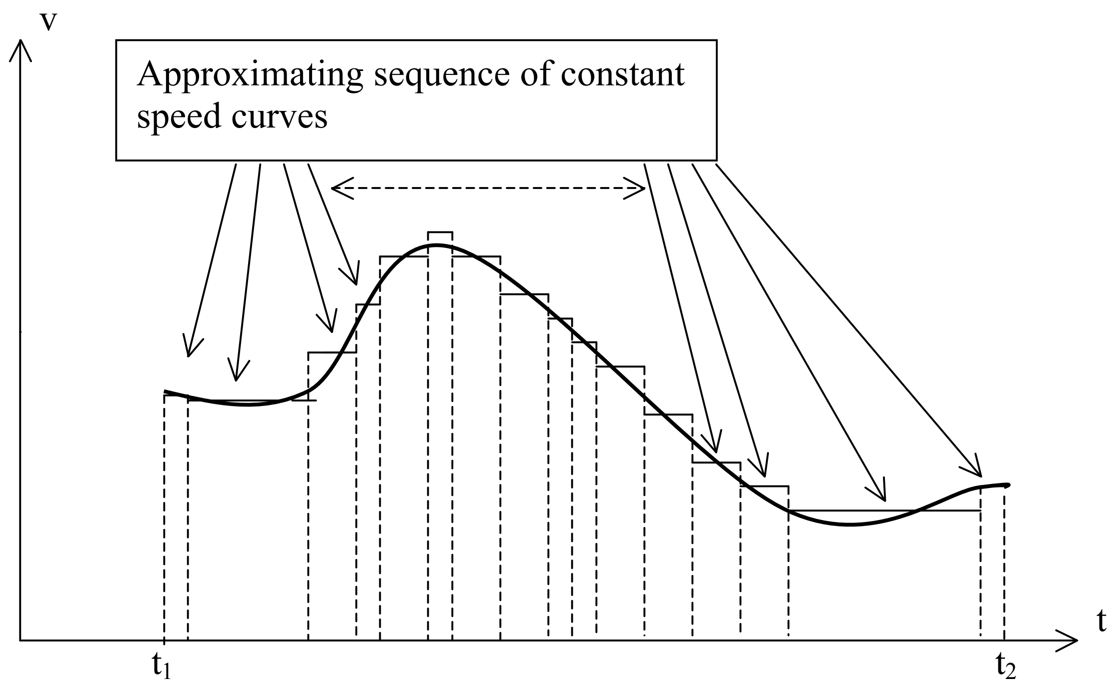


Fig. IA. 3: Stepwise constant speed curve approximating original continuous speed curve to establish relation between distance covered and area bounded by the curve.

Having considered the argument for the useful importance of the area in a speed versus time graph, we can extend that utility by pointing out the interpretation of a speed curve that dips *below* the time axis, as in **Fig. IA. 4**.

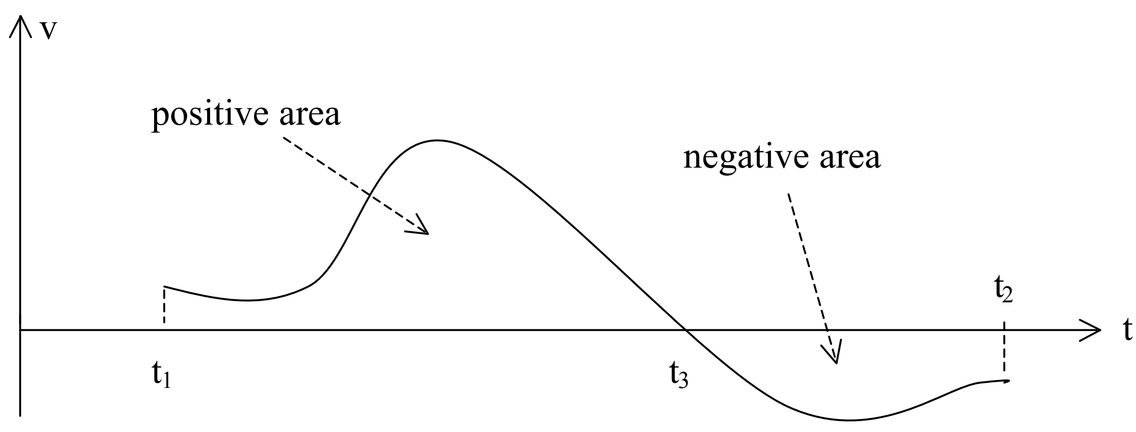


Fig. IA. 4: Speed curve for motion with direction reversal.

Remember that when the speed curve *touches* the time axis it means the speed is zero. Above the time axis a drop in the speed towards zero corresponds to subtracting speed from the initial value. But subtracting positive speed is the same as adding negative speed. So a natural interpretation of a speed curve that dropped below the time axis is that it describes motion for which a negative speed has been added to zero. But what does a negative speed mean?! *It means motion in the opposite direction.* If the curve drops below the time axis then the vertical speed axis must be extended below the time axis as well.

What happens to the interpretation of the area bounded by the speed versus time graph when the speed curve drops below the time axis? Well, if negative speed means motion in the opposite direction, then, presumably, the area between the negative speed curve and the time axis would be proportional to the distance covered while moving in the opposite direction. If we give that area a negative sign, then *the algebraic sum of the positive area under the positive speed curve and the negative area above the negative speed curve should be proportional to the net distance covered as a consequence of some motion in one direction and some motion in the opposite direction*. If, for instance, we consider motion to the right as positive and to the left as negative, then, if the negative area bounded by the speed curve overwhelms the positive area, that means the net motion leaves the moving object to the left of where it started.

The last feature of speed versus time graphs from which we can read off an important feature of the motion being described is the **slope,** or inclination of the speed curve (**Fig. IA. 5**). Where the speed curve is climbing to higher values of v as the time, t, increases, the motion is **accelerating** (positively) and the amount of acceleration is determined by the steepness of the climbing curve. Where the speed curve is falling to lower values of v as the time, t, increases, the motion is decelerating or, **accelerating** (negatively) and again the amount is determined by the steepness of the falling curve. Where the speed curve is flat, or horizontal, the speed is constant as time increases and the acceleration is zero.

The exact definition of the slope of a curve at a point on the curve is given in terms of a straight line that is just *tangent* to the curve at the point. First the change in vertical coordinate divided by the corresponding change in horizontal coordinate, which is constant for a straight line, is the **slope** of the straight line. Then the **slope** of a curve at a point on the curve is the **slope** of the tangent line to the curve at the point.

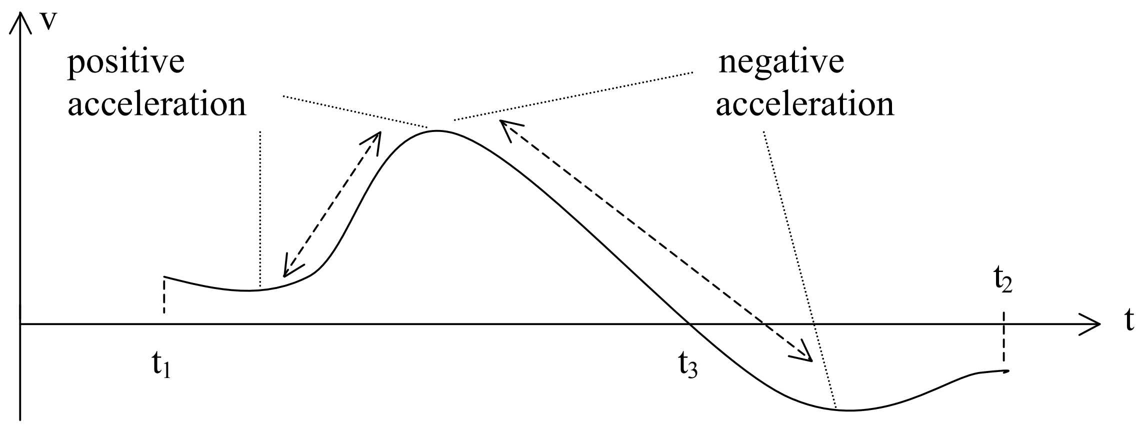


Fig. IA. 5: Reading acceleration from speed versus time curve.

This completes our examination of the incredible usefulness of speed versus time graphs in representing important features of motion. Such a graph gives visual representation of the distance covered (area bounded by the curve), the speed (vertical axis), the acceleration (slope of the curve) and the time (horizontal axis).

We are now ready to consider the first really important application of these ideas in the history of physical science; **Galileo's** study of the accelerated motion of falling bodies and projectiles due to Earth's gravity.