

Chapter 1

Newton's Laws of Motion

1.1 Translating words into equations

Before we get started, it is important to understand what we are trying to achieve. This is generally good advice, and when it comes to mathematical modelling it is essential. Our understanding of reality is inevitably somewhat limited and we need to keep this in mind as we try to model what is going on. However, it may also be that our understanding is too precise and that we do not need all the details in order to describe a particular phenomenon. In essence, we have to make choices.

The main purpose of applied mathematics is to make quantitative statements and predictions about the world that surrounds us. We achieve this by writing down and solving equations expressing relationships between quantities, such as distance, time, volume, money, density, value etc. This typically involves using **vectors** to describe the position of object in space and time (relative to some suitable frame of reference) and differential equations, expressing relationships between the **rates of change** of quantities with respect to others¹.

The ultimate aim of the process is to provide quantitative (or descriptive) predictions concerning “real world” situations. This is not a purely academic exercise. Without the study of these kinds of problems, life as we know it would not exist, your mobile phone would not work, you could not drive a car, planes would not fly, stock markets would be a lot poorer and people could not make predictions about climate and global warming. Non-mathematical models tend to be unable to make quantitative predictions that can be verified against experiments. Good mathematical models should be able to do this.

In order to model something, it is important to be able to write down your assumptions and translate them into mathematical statements. Once you have a well-formed mathematical system you can begin to try solve the problem and make predictions. At this point it is fairly obvious that you need the mathematical problem to be simple enough that you can solve it², but at the same time the model must retain the key elements of the process we are trying to describe. At the end of the day, the result must be matched against observations

¹Ultimately we will need derivatives of vectors, meaning that we need to master vector calculus and partial differential equations.

²As you gain experience, you will be equipped to deal with more complex problems, but there will always be a limit to what you are able to do. That's life.

and experiments. If our model provides a satisfactory description of reality, then we are done. If we need a more accurate model, then we need to go back and re-think our assumptions. You can think of this as an iterative process, outlined by the flow-chart in figure 1.1.

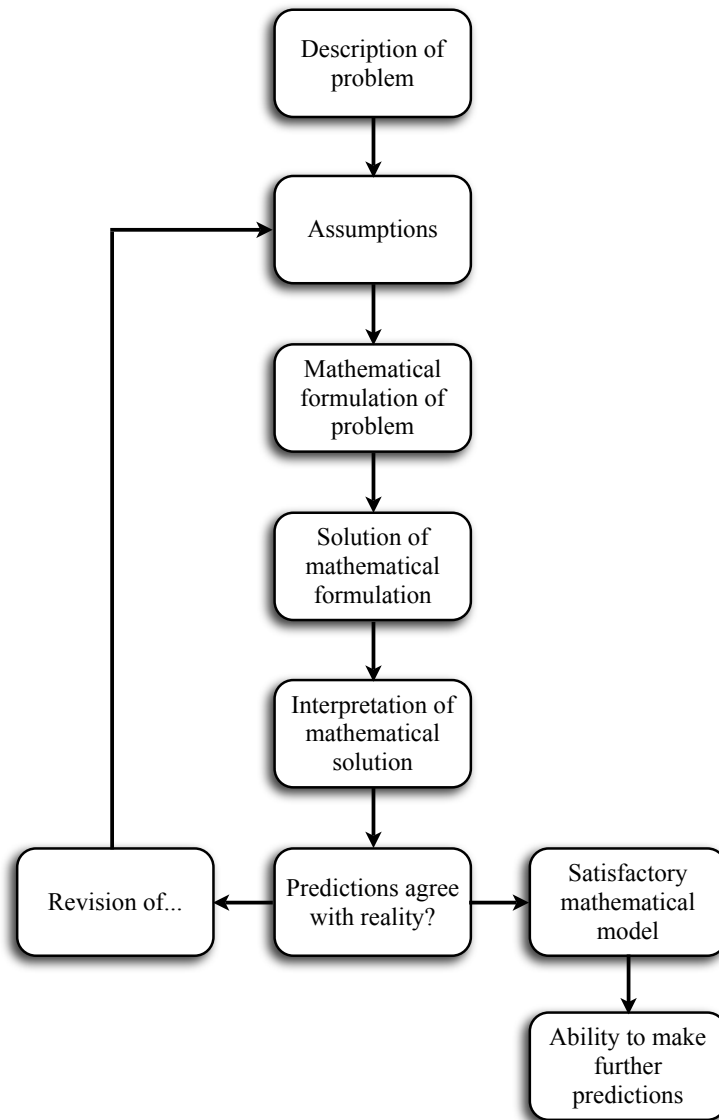


Figure 1.1: The process of mathematical modelling.

This may seem somewhat abstract, but it is easy to find examples of how

our understanding has developed in just this way. Consider for example our understanding of moving objects. We can immediately think of two major paradigm shifts. First we have Aristotle's notion that there can be no effect without a cause (essentially, there is no motion without a force), which gave way to Newton's laws (where the role of the force is to alter the motion, or cause it in the first place). Secondly, as it became apparent that there was a contradiction between Newtonian mechanics and electromagnetism in the late 1800s, developments were set in motion³ that led to the development of Einstein's special theory of relativity. Along the way, our understanding space and time changed. We moved from the fixed clockwork universe of Newton to the flexible space-time of Einstein.

As we go on, we will consider these world views in some detail. We will start with Newton's laws, mainly as a reminder but also in order to take the opportunity to revise (some of) the mathematical tools we need in order to solve relevant problems.

1.2 Case study: Newton's Laws of Motion

One of the first, and fundamental, applications of differential equations was to the laws of motion of objects, as written down by Newton in his *Philosophiae naturalis principia mathematica*, or *Principia* for short, from 1687. The fundamentals of classical mechanics (as it is now called) were presented in precise and complete detail for the first time. They were based on experimental observations and written down in verbal form.

Suppose we want to describe the motion of a single body. In order to do this, we need two definitions. First of all:

- the **mass** of a body m (measured in kilogrammes, kg) is a physical quantity associate with the body that acts to resist any change in motion of it⁴. Secondly,
- the **momentum** of the body (measured in kilogrammes per metre per second, kgms^{-1}) is defined to be the mass times the rate of change of position of the body (relative to some point). A speeding bullet of low mass can have similar momentum to a slow moving but heavy truck.

With these two definitions, we can state the experimentally deduced laws of classical motion. These are Newton's three laws:

- **N1:** A body remains at rest or in uniform motion with respect to an inertial frame unless acted upon by a net force.
- **N2:** In an inertial frame a body acted upon by a net force moves in such a way that the rate of change with time of momentum equals the net force applied.

³If that's not a terrible pun, then I don't know what it is...

⁴Strictly speaking, this is the **inertial** mass of the body.

- **N3:** If two bodies exert net forces upon each other, these net forces are equal in magnitude and opposite in direction.

Clearly, these statements introduce several additional concepts. First of all, we have the concept of a **force**. This is a derived quantity, in that its definition depends on that of momentum. The force is the amount of push or pull it takes to alter the motion of the body (measured in Newton, or kgms^{-2}). In essence, the first law (**N1**) defines what is meant by a zero net force, which in turn can be used to define an *inertial frame* as one on which no net force acts. The three laws are only valid in such inertial (non-accelerating) frames. Finally, the third law (**N3**), is not general. It holds only for particles in contact, or acted upon by forces parallel to the straight line connecting them.

Newton originally used classical geometry to try to make predictions concerning the motion of bodies. However, the differential calculus of Newton/Leibnitz (developed at the same time) was quickly applied to the subject and this allowed a systematic approach to be developed, kick-starting a scientific revolution.

1.3 Reminder: Scalars and vectors

Some quantities, like temperature (and pressure) are identified by their magnitude. These are called **scalar** quantities, and they are simply described by numbers:

$$\text{scalar} = \text{number (real or complex)}.$$

Other quantities, like wind velocity, are identified by both **magnitude** (size or length) *and* **direction**. Such quantities are represented geometrically by arrows.

$$\text{vector} = \text{number (its size) and direction}.$$

It is easy to think of examples of either kind of quantity:

- The position of one point relative to another in space is a vector.
- The distance between the two points is a scalar.
- The velocity of a particle in space is a vector.
- The speed of a particle is a scalar.
- The temperature of a body is a scalar.

As we will need to make a distinction⁵ we need to introduce suitable notation. There are a variety of ways to represent vectors. For example, the vector that joins point A with point B can be equivalently represented by indicated by \vec{r}_{AB} , AB, \overrightarrow{AB} or by a bold letter, **a**, say. Throughout these notes we represent vectors by arrows, e.g. \vec{a} , as it is not practical to use bold symbols in calculations done by hand or on the blackboard. In most of the illustrations vectors are indicated by bold symbols.

⁵This is **very** important. We always need to indicate if a quantity is a scalar or a vector. Sloppy notation is the first step towards confusion!

The magnitude (modulus) of a vector \vec{v} is the length of the arrow and is denoted by the symbol $|\vec{v}|$. Vectors with unit length are called **unit vectors** and are often denoted by a hat. So for example, $\hat{a} \equiv \vec{a}/|\vec{a}|$. Two vectors are equal if they have the same length and the same direction.

A point in three dimensions can in general be reached by three non-coplanar vectors. One particularly important set of coordinates is the Cartesian. These coordinates can be represented by an orthonormal (the unit vectors are orthogonal) set of vectors called the **standard basis vectors** relative to an origin \mathcal{O} of coordinates. These basis vectors are:

- The unit vector from \mathcal{O} parallel to the x axis, denoted by \hat{e}_x or \hat{i} ;
- The unit vector from \mathcal{O} parallel to the y axis, denoted by \hat{e}_y or \hat{j} ;
- The unit vector from \mathcal{O} parallel to the z axis, denoted by \hat{e}_z or \hat{k} .

These vectors form a **right**-handed set. The Cartesian system is but one possible set of coordinates. Later it will be convenient to introduce different orthogonal sets of basis vectors to deal with (for example) rotating systems.

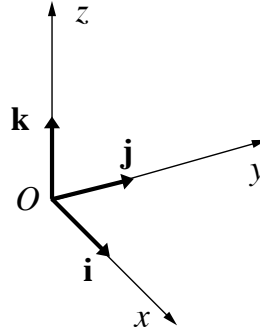


Figure 1.2: Triad of unit basis vectors in Cartesian coordinates.

Any point or vector in Cartesian 3D can be represented in component form by a linear combination of the standard basis vectors. For example

$$\vec{a} = a_1 \hat{i} + a_2 \hat{j} + a_3 \hat{k}$$

where (a_1, a_2, a_3) are the Cartesian (x, y, z) coordinates of the point \vec{a} with respect to the origin or the vector. These three numbers are called the **components** of the vector with respect to the $\hat{i}, \hat{j}, \hat{k}$ basis.

1.4 A few comments on derivatives

It is important to understand that there are different ways to define vector derivatives. In general, we have to use the tools from vector calculus involving

grad, div and curl, but here we will focus on the simplest case where we take derivatives with respect to a single variable (typically time).

As we have already seen, vectors can be used to indicate the position, velocity and acceleration of a moving particle. Consider such a particle at a point \vec{r} with coordinates (x, y, z) at time t . Its position is indicated by a vector \vec{r} , called the position vector, whose tail is at the origin and whose head lies at the point (x, y, z) .

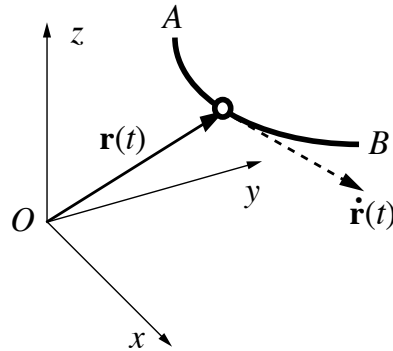


Figure 1.3: The instantaneous position and velocity vectors of a particle at a typical time as it moves along a trajectory between points A and B .

Its Cartesian components are x , y and z , which themselves are functions of t , say

$$x = x(t), \quad y = y(t), \quad z = z(t).$$

The position vector can thus be written as

$$\vec{r} = \vec{r}(t) = x(t)\hat{i} + y(t)\hat{j} + z(t)\hat{k}.$$

This is an example of a **vector function**, a vector that depends on one (or more) parameters (here time). As t varies $\vec{r}(t)$ describes a curve in 3D space. The position along that curve is parameterised by t .

Let us now consider the positions of the particle at time t and $t + \delta t$ where $\delta t \rightarrow 0$. The distance between these points is given by

$$\vec{r}(t + \delta t) - \vec{r}(t).$$

The rate of change of the position as time $\delta t \rightarrow 0$ is nothing other than instantaneous velocity \vec{v} of the particle:

$$\vec{v}(t) = \lim_{\delta t \rightarrow 0} \frac{\vec{r}(t + \delta t) - \vec{r}(t)}{\delta t}.$$

We recognise this as the limit definition of a derivative.

If we consider Cartesian coordinates, the basis vectors \hat{i} , \hat{j} and \hat{k} are unit vectors that remain constant relative to the world. Hence any derivative of them

with respect to time must be zero⁶:

$$\frac{d\hat{i}}{dt} = \frac{d\hat{j}}{dt} = \frac{d\hat{k}}{dt} = 0.$$

Consequently in Cartesian coordinates the velocity of the particle is given by

$$\vec{v}(t) \equiv \frac{d\vec{r}}{dt} = \dot{\vec{r}}(t) = \frac{dx(t)}{dt}\hat{i} + \frac{dy(t)}{dt}\hat{j} + \frac{dz(t)}{dt}\hat{k}$$

In the limit as $\delta t \rightarrow 0$ the rate of change of position vector with time becomes tangential to the trajectory (see figure 1.3). Hence the velocity vector is tangential to the trajectory of the particle at each time t .

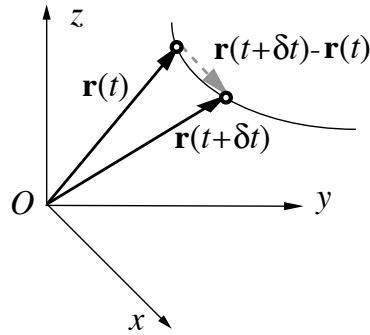


Figure 1.4: The change in position vector $\vec{r}(t + \delta t) - \vec{r}(t)$ for $\delta t \ll t$.

From now on we shall often refer to the velocity \vec{v} as $\dot{\vec{r}}(t)$ where the dot “ $\dot{}$ ” refers to differentiation with respect to time, d/dt .

The definition of the acceleration is similar – it is the rate of change of velocity with time. Mathematically we have:

$$\frac{d\vec{v}}{dt} \equiv \frac{d^2\vec{r}}{dt^2} = \ddot{\vec{r}}(t) = \frac{d^2x(t)}{dt^2}\hat{i} + \frac{d^2y(t)}{dt^2}\hat{j} + \frac{d^2z(t)}{dt^2}\hat{k}$$

If $\ddot{\vec{r}}(t) \neq \vec{0}$ then the velocity $\dot{\vec{r}}(t) \neq \vec{0}$ changes with time. If the $\dot{\vec{r}}(t) \neq \vec{0}$ changes with time, then so does the tangent vector to the curve. Hence if a particle travels along a trajectory that is **not** straight, then it must be experiencing an acceleration.

Finally, we may also consider the acceleration as the limit

$$\ddot{\vec{r}}(t) = \lim_{\delta t \rightarrow 0} \frac{\dot{\vec{r}}(t + \delta t) - \dot{\vec{r}}(t)}{\delta t}.$$

The acceleration is therefore (the limit of) the scaled difference between the velocities at time t and $t + \delta t$ as the particle moves along a trajectory. Since the velocities are tangential to the trajectory, it is not too difficult to see that the acceleration vector points in the direction that the curve is bending.

⁶Note that this is not true for other coordinate systems, like spherical polar coordinates, for example. So we have to be careful.

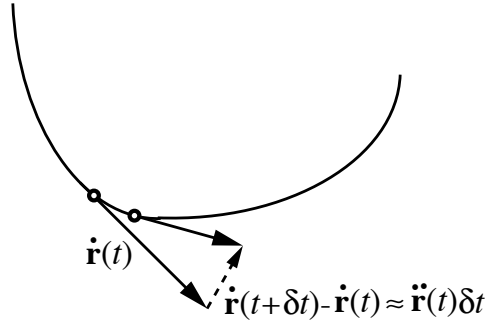


Figure 1.5: The change in velocity vector $\dot{\mathbf{r}}(t + \delta t) - \dot{\mathbf{r}}(t)$ for $\delta t \ll t$.

1.5 Mathematical expression of Newton's Laws

As we are typically trying to work out how an object moves through space as time progresses, we need to write down Newton's laws in three space dimensions. We need to work with vectors⁷, as in figure 1.6. In general, this means that we need some understanding of vector calculus. However, we are going to assume that we have not yet added the relevant tools to our arsenal, and mostly consider problems in one space dimension. Then we have

$$\begin{array}{ll}
 x(t) & \text{the position of a particle} \\
 v(t) = \frac{dx}{dt} & \text{its velocity} \\
 \left| \frac{dx}{dt} \right| & \text{the speed, and} \\
 p(t) = mv = m \frac{dx}{dt} & \text{the momentum}
 \end{array}$$

In three dimensions we have to use vectors to represent these quantities. If the position of a body relative to some origin is

$$\vec{r}(t) = x(t)\hat{i} + y(t)\hat{j} + z(t)\hat{k}$$

at time t , then its velocity $\vec{v}(t)$ is its rate of change of position with respect to time⁸, i.e.,

$$\vec{v}(t) = \frac{d\vec{r}(t)}{dt} = \frac{dx}{dt}\hat{i} + \frac{dy}{dt}\hat{j} + \frac{dz}{dt}\hat{k}$$

Qualitatively Newton's laws state that force is the amount of push or pull that it takes to change the motion of a body. Quantitatively, **N2** states that the *sum of all forces* \vec{F} acting on a body is given by

⁷Some revision of basic vector operations may be required as we need to be comfortable applying them.

⁸Note that this is easy to work out since the Cartesian basis vectors are fixed. Later, we will discuss coordinate systems for which this is not true.

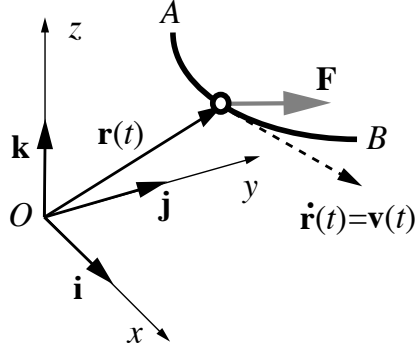


Figure 1.6: A Cartesian coordinate system (x, y, z) for a body moving under a force $\vec{F}(t)$ under Newton's laws along a trajectory from A to B . The position of the particle at time t is given by the vector $\vec{r}(t) = x(t)\hat{i} + y(t)\hat{j} + z(t)\hat{k}$. The velocity of the particle is the rate of change of position vector with respect to time, $\dot{\vec{r}}(t) = \vec{v}(t)$ and is, by definition, tangential everywhere to the trajectory. Note that the force is not necessarily in the same direction as the velocity. From N2, the force is parallel to the acceleration $\ddot{\vec{r}}(t)$.

$$\vec{F} \equiv \frac{d\vec{p}}{dt}$$

where $\vec{p} = m\vec{v}$ is the momentum.

If we insert the definition of momentum into the expression for force we obtain

$$\vec{F} = \frac{d(m\vec{v})}{dt} = \frac{dm}{dt}\vec{v} + m\frac{d\vec{v}}{dt} = \frac{dm}{dt}\dot{\vec{r}} + m\ddot{\vec{r}}$$

N2 is often referred to as “force = mass \times acceleration”, or “ $F = ma$ ”. However we can see that this is in general incorrect. Only if the mass of the body is constant do we have the simplification that

$$\vec{F} = m\ddot{\vec{r}}.$$

If a force only acts in a single direction, say x , then we have

$$F = m\frac{d^2x}{dt^2}.$$

This is a second order differential equation. Given the force, the solution of the differential equation⁹ will give the position of the particle x at any future time

⁹We clearly need to be on top of the solution of second order differential equations. Such problems can be complicated, but at the same time there are many useful situations that are easy to deal with.

t .

N2 can be written in terms of a first order differential equation involving just the speed of the particle, $v(t)$ as

$$F = m \frac{du}{dt}.$$

Here the solution will give the speed of the particle at any future time, $u(t) = u_s(t)$, say. We will see below that we will need to provide an initial speed of the particle to specify unambiguously the future speed $u_s(t)$. Once this has been done we can predict the future position of the particle by remembering that the speed is just the rate of change of distance with time, i.e.,

$$\frac{dx}{dt} = u_s(t).$$

This is again a differential equation that, in principle, could be integrated to obtain $x(t)$, after specifying an initial position $x(0)$, say.

1.6 Reminder: Multiplying vectors

A vector \vec{a} can be multiplied by a scalar t . The resulting vector $t\vec{a}$ has magnitude $|t||\vec{a}|$ which may be greater or less than $|\vec{a}|$ depending on whether $|t| > 1$ or < 1 , respectively. If $t > 0$, the direction of the vector remains unchanged. If $t < 0$, the direction reverses. If $t = 0$, then the result is the zero vector, denoted by $\vec{0}$.

It is also easy to add vectors. In order to obtain the sum of two vectors \vec{u} and \vec{v} draw a parallelogram that has sides \vec{u} and \vec{v} . The sum $\vec{u} + \vec{v}$ is the vector that joins the tail of \vec{u} with the vertex of \vec{v} . This is equivalent to the triangle rule shown in figure 1.7

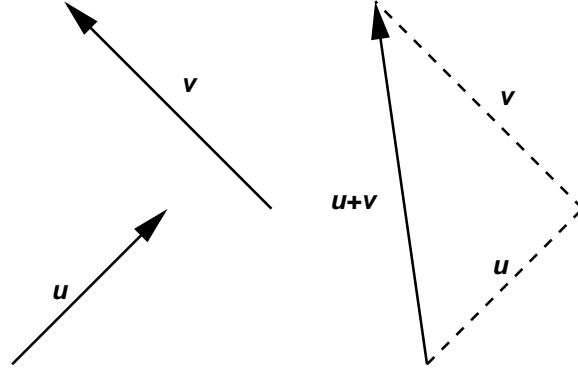


Figure 1.7: Triangle law of vector addition.

The addition of vectors is *commutative* and *associative*.

$$\vec{a} + \vec{b} = \vec{b} + \vec{a} \quad \vec{a} + (\vec{b} + \vec{c}) = (\vec{a} + \vec{b}) + \vec{c}$$

In component form vector addition is easy. If $\vec{a} = a_1\hat{i} + a_2\hat{j} + a_3\hat{k}$ and $\vec{b} = b_1\hat{i} + b_2\hat{j} + b_3\hat{k}$ then

$$\vec{a} + \vec{b} = (a_1 + b_1)\hat{i} + (a_2 + b_2)\hat{j} + (a_3 + b_3)\hat{k}$$

1.6.1 Scalar Product (Dot Product)

There are two different ways multiplying vectors with vectors. First of all, the **scalar product** combines the two vectors into a single number. The scalar product of the vectors \vec{a} and \vec{b} is defined as

$$\vec{a} \cdot \vec{b} = |\vec{a}| |\vec{b}| \cos \theta$$

where θ is the angle between the directions of \vec{a} and \vec{b} .

Hence, if $\vec{a} = a_1\hat{i} + a_2\hat{j} + a_3\hat{k}$, and $\vec{b} = b_1\hat{i} + b_2\hat{j} + b_3\hat{k}$, then the scalar product is the scalar quantity

$$\vec{a} \cdot \vec{b} = a_1b_1 + a_2b_2 + a_3b_3 .$$

In the special case where $\vec{a} \cdot \vec{b} = 0$ then either $\vec{a} = \vec{0}$ or $\vec{b} = \vec{0}$ or \vec{a} and \vec{b} are orthogonal.

We can use the scalar product to work out the length of a vector. Simply note that

$$\vec{a} \cdot \vec{a} = |\vec{a}|^2$$

or, in component form,

$$|\vec{a}| = \sqrt{\vec{a} \cdot \vec{a}} = \sqrt{a_1^2 + a_2^2 + a_3^2}$$

Alternatively, we can work out the angle between two vectors. The cosine of the angle θ between the directions of \vec{a} and \vec{b} can be determined from the definition of the scalar product.

$$\cos \theta = \frac{\vec{a} \cdot \vec{b}}{|\vec{a}| |\vec{b}|}$$

which if written in (x, y, z) components is

$$\cos \theta = \frac{a_1b_1 + a_2b_2 + a_3b_3}{\sqrt{(a_1^2 + a_2^2 + a_3^2)(b_1^2 + b_2^2 + b_3^2)}}$$

We can also use the scalar product to work out the component of one vector in the direction of another. This is often called a “projection”. For example, the component of \vec{b} in the direction of the vector \vec{a} is

$$\frac{\vec{b} \cdot \vec{a}}{|\vec{a}|} = \vec{b} \cdot \hat{a}$$

Finally, it is important to note that the order of the scalar product is irrelevant (the procedure is commutative): $\vec{a} \cdot \vec{b} = \vec{b} \cdot \vec{a}$.

1.6.2 Vector Product (Cross Product)

We can also multiply two vectors in such a way that we obtain a new vector. This is the so-called **vector product** (or **cross product**), which is defined in such a way that the vector product of the vectors \vec{a} and \vec{b} is defined as

$$\vec{a} \times \vec{b} = |\vec{a}| |\vec{b}| \sin \theta \hat{n}$$

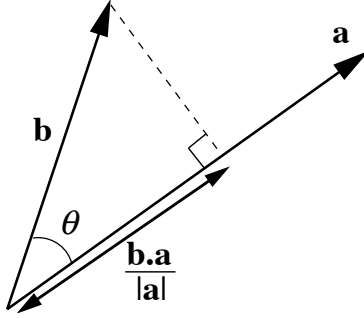


Figure 1.8: Resolved component of a vector

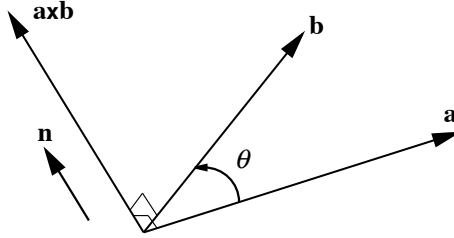


Figure 1.9: Definition of vector product

where θ is the angle between the directions of \vec{a} and \vec{b} and \hat{n} is the *unit* vector perpendicular to both them using the right-hand rule, travelling from \vec{a} to \vec{b} .

Geometrically $|\vec{a} \times \vec{b}|$ is the area of the parallelogram with sides \vec{a} and \vec{b} .

Note that the vector product is non-commutative, i.e.,

$$\vec{b} \times \vec{a} = -(\vec{a} \times \vec{b})$$

Written in components the vector product can be calculated using the 3×3 determinant

$$\vec{a} \times \vec{b} = \begin{vmatrix} \hat{i} & \hat{j} & \hat{k} \\ a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{vmatrix} = (a_2b_3 - a_3b_2)\hat{i} + (a_3b_1 - a_1b_3)\hat{j} + (a_1b_2 - a_2b_1)\hat{k}.$$

When we are dealing with basis vectors, it is worth noting the following special cases:

$$\hat{i} \times \hat{i} = \hat{j} \times \hat{j} = \hat{k} \times \hat{k} = \vec{0}$$

and

$$\hat{i} \times \hat{j} = -\hat{j} \times \hat{i} = \hat{k} \quad \hat{j} \times \hat{k} = -\hat{k} \times \hat{j} = \hat{i} \quad \hat{k} \times \hat{i} = -\hat{i} \times \hat{k} = \hat{j}$$

1.7 Simple examples

We start with a couple of examples where the path followed by the particle is known and we want to find the force acting on the particle.

1.7.1 Motion in a circle

A particle of unit mass (in suitable units) moves on the circular path given by

$$\vec{r}(t) = \cos t \hat{i} + \sin t \hat{j} + 0 \hat{k}.$$

Calculate the total force (again in suitable units) on the particle and show that it is always orthogonal to the trajectory.

The particle is of constant mass and so **N2** becomes

$$\vec{F} = m \ddot{\vec{r}}(t) = \frac{d^2 \vec{r}(t)}{dt^2}$$

The velocity of the particle is given by

$$\dot{\vec{r}}(t) = \frac{d}{dt} (\cos t \hat{i} + \sin t \hat{j} + 0 \hat{k}) = -\sin t \hat{i} + \cos t \hat{j} + 0 \hat{k}.$$

while the acceleration is

$$\ddot{\vec{r}}(t) = \frac{d}{dt} (-\sin t \hat{i} + \cos t \hat{j} + 0 \hat{k}) = -\cos t \hat{i} - \sin t \hat{j} + 0 \hat{k}.$$

Hence we have:

$$\vec{F}(t) = -\cos t \hat{i} - \sin t \hat{j} + 0 \hat{k} = -\vec{r}.$$

The particle is moving in a circle of radius 1, centred on the origin of co-ordinates and in the (x, y) -plane. This follows from elimination of t from the Cartesian components of $\vec{r}(t)$:

$$x(t) = \cos t, \quad y(t) = \sin t, \quad z(t) = 0 \quad \Rightarrow \quad x^2 + y^2 = 1.$$

The velocity of this particle is always tangential to this circle (you can show $\dot{\vec{r}}(t) \cdot \vec{r}(t) = 0$). The speed of the particle is constant since

$$|\dot{\vec{r}}(t)| = \sqrt{\dot{x}(t)^2 + \dot{y}(t)^2} = \sqrt{\sin^2(t) + \cos^2(t)} = 1$$

The force acting on the particle is of constant magnitude and in this case it is always orthogonal to the velocity since

$$\vec{F}(t) \cdot \dot{\vec{r}}(t) = \cos t \sin t - \cos t \sin t = 0$$

This force is a “radial” force as it always points radially inwards to the origin. Such forces are also sometimes called “central” or “centripetal” forces, since they “seek” the centre. Note that if the circular path of this particle is caused by it being constrained to move on a smooth wire, then the particle experiences a reactionary force **N3** from the wire to keep it in place: this is the “centrifugal” force. We will return to the problem of circular motion later.

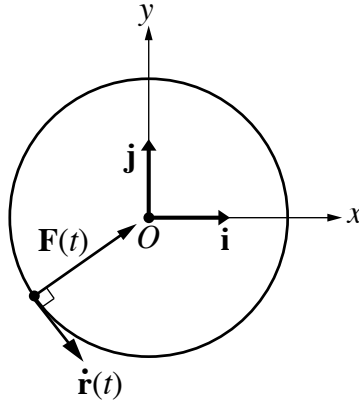


Figure 1.10: Velocity and forces experienced during motion in a circle.

1.7.2 Motion under gravity

Let us now turn to a situation where the forces are known and the motion of a particle must be calculated. Consider a body of constant mass moving under the influence of gravity near the surface of the earth. Later we will see that gravity varies from place to place, but near the surface of the earth it can be taken as approximately constant. Under these conditions, experiment shows that, if gravity is the only force acting then all bodies accelerate in the same way. If we take a Cartesian coordinate frame with \hat{k} pointing away from the earth, then¹⁰

$$\ddot{\vec{r}} = -g \hat{k} \approx -9.81 \text{ms}^{-2} \hat{k}.$$

where

$$\ddot{\vec{r}} = \ddot{x} \hat{i} + \ddot{y} \hat{j} + \ddot{z} \hat{k}$$

The force acting on a particle of constant mass moving under the influence of gravity is thus

$$\vec{F} = -mg \hat{k}.$$

The magnitude of this force that is said to be the “weight” of the particle¹¹

Suppose a projectile, the particle, of constant mass is launched from the surface of the earth with initial velocity $\vec{u} = U\hat{u}$ (with U the initial speed of the particle) at an angle θ to the horizontal. In order to find the path that the particle takes as it is acted on by this force, we choose a Cartesian frame with \hat{i} along the horizontal projection of the direction of launch and \hat{k} pointing vertically. **N2** then gives the equation of motion as¹²

$$m\ddot{\vec{r}}(t) = -mg \hat{k}.$$

¹⁰To get an idea of the magnitude of the answer, it is often sufficient to approximate g as 10 ms^{-2} .

¹¹Strictly speaking, this is the body’s **gravitational mass**. Actually, if we want be precise it is the **passive gravitational mass**. There is also an **active gravitational mass**, which is the one that leads to the gravitational force in the first place.

¹²Technically, we are making an assumption by taking the inertial mass (on the left) as equal to the gravitational mass (the right).

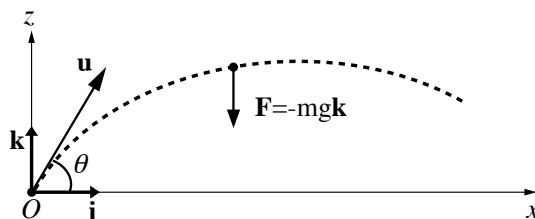


Figure 1.11: Trajectory of a projectile.

We need to solve a second order linear (vector) differential equation so require two pieces of boundary or initial data to determine the solution uniquely. We have these. They are:

$$\vec{r}(0) = 0, \quad \dot{\vec{r}}(0) = \vec{u}$$

We solve this problem to find the path given by the vector $\vec{r}(t)$. The most elegant strategy is to solve the vector equation directly. In the present case this is easy, since the force we consider is constant. We have

$$\ddot{\vec{r}}(t) = -g \hat{k},$$

$$\dot{\vec{r}}(t) = A\hat{i} + C\hat{j} + (E - gt) \hat{k},$$

and

$$\vec{r}(t) = (At + B)\hat{i} + (Ct + D)\hat{j} - \left(\frac{1}{2}gt^2 - Et - F\right) \hat{k},$$

The initial conditions then tell us that

$$\vec{r}(0) = 0 \implies B = D = F = 0$$

and

$$\dot{\vec{r}}(0) = \vec{u} \implies A = U \cos \theta, \quad C = 0, \quad E = U \sin \theta,$$

leading to the final result

$$\vec{r}(t) = Ut \cos \theta \hat{i} + \left(Ut \sin \theta - \frac{1}{2}gt^2\right) \hat{k}.$$

Note that we would have arrived at the same answer by writing out the problem in terms of the vector components, starting with

$$\vec{r} = x\hat{i} + y\hat{j} + z\hat{k}$$

and integrating them separately. It is a matter of choice—each strategy has its merits. Either way, we have equations for the displacements in terms of time¹³ t :

$$\begin{aligned} x(t) &= Ut \cos \theta \\ y(t) &= 0 \\ z(t) &= Ut \sin \theta - \frac{1}{2}gt^2 \end{aligned}$$

¹³As a bit of a caution... you may have memorised this result during your previous life (and perhaps know the “suvat” equations). Now it is time to forget all about that! It is good practice to derive things from scratch, especially as the simple formulas may not apply under the more general circumstances we are going to consider.

In this case we can simplify these and get an algebraic expression for the path of the particle. We can eliminate t between the components by using the first equation to write

$$t = \frac{x}{U \cos \theta},$$

and substituting for t in the third equation. We then arrive at the Cartesian form of the trajectory

$$z = x \tan \theta - \frac{gx^2 \sec^2 \theta}{2U^2},$$

This is an equation of the form $z = \alpha x - \beta x^2$, which is a parabola.

Example 1: A projectile

Assuming that the projectile is launched over flat ground determine the range of the projectile and hence show that the maximum range for a given launch speed u is achieved when $\theta = \pi/4$.

The particle hits the horizontal plane when $z = 0$, i.e., when

$$x \tan \theta - \frac{gx^2 \sec^2 \theta}{2U^2} = 0$$

This is a quadratic equation for x . The first solution $x = 0$ corresponds to the launch position while the second solution is

$$x = \frac{2U^2 \tan \theta}{g \sec^2 \theta} = \frac{2U^2}{g} \sin \theta \cos \theta = \frac{U^2}{g} \sin 2\theta.$$

This is the range of the projectile.

For fixed launch speed u and gravitational acceleration g , this can be maximised by maximising $\sin 2\theta$. This leads to $\sin 2\theta = 1 \Rightarrow \theta = \pi/4$.

In a similar way you can calculate the maximum height achieved in the trajectory as a function of θ for fixed launch speed. This is known as the safety parabola, since the projectile can never pass outside this curve.

You may also repeat the above example, to find the trajectory of a projectile being launched on a surface that has constant slope and is inclined at an angle α to the horizontal (α could be positive or negative).

1.7.3 Springs

A spring is a piece of material that extends its length when a force is applied to it. It has a natural length l , say when it is unstressed. If a force is applied along the spring and is not too large (a strong force might break the spring) then the change in length of the spring is proportional to the magnitude of the force. That is, we have

$$\text{magnitude of force along spring} = k[(\text{actual length of spring}) - (\text{natural length of spring})].$$

The constant of proportionality k is called the spring constant and varies between springs. It has units of N m^{-1} . In addition the spring always tries to

return to its natural length (this gives the direction of the force). Such a spring, where the force depends linearly on the displacement, is commonly called a **Hookean spring**.

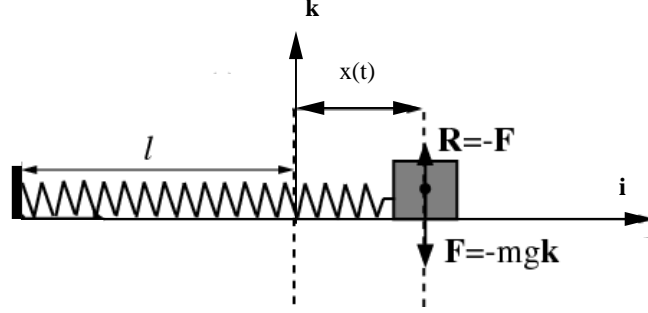


Figure 1.12: Oscillating spring system.

Here we consider the motion of a light (=negligible mass) spring with spring constant k that is attached at one end to a wall and at the other end to a mass m moving on a smooth table (i.e., we will neglect friction). The mass is displaced along the length of the spring to a distance of L metres, held stationary and then released. We want to calculate the subsequent motion of the mass.

For simplicity, we take the coordinate system to be such that the origin is at the location of the end of the unstretched string with \hat{i} pointing along the string. Hence, the position of the mass will be given by

$$\vec{r}(t) = x(t) \hat{i} + 0 \hat{j} + 0 \hat{k}.$$

where the actual length of the spring at any time is $x(t) + l$. The force acting on the mass has three parts, one due to the spring, one due to gravity, and the other due to the smooth table it is sitting on.

$$\vec{F} = -k[(x(t) + l) - l] \hat{i} - mg \hat{k} + \vec{R}$$

Note that we need to find $x(t)$ and that we also do not currently know \vec{R} . (The minus sign in front of k is because an extended spring creates a force that will seek to draw it back to its natural unextended length).

From **N2** we obtain:

$$\vec{F} = m \ddot{\vec{r}}(t)$$

and if we equate the components of this equation then in the \hat{k} direction we find that we must have

$$\vec{R} = mg \hat{k}.$$

in order for the mass to stay on the smooth surface. This is the reaction force applied by the table to the mass. In addition, we find that in the \hat{i} direction we get the equation

$$m \ddot{x}(t) = -k x(t)$$

We expect that both m and k are positive by definition. The equation is then a second order linear ordinary differential equation with constant coefficients and we may use standard methods (we will return to this later) to find the solution. This leads to

$$x(t) = A \cos \omega t + B \sin \omega t, \quad \text{where} \quad \omega = \sqrt{\frac{k}{m}},$$

with constants A and B . So we require two pieces of information to solve the problem uniquely. From the problem description we know that the initial displacement is L and that the mass is initially stationary. Hence we can write

$$x(0) = L, \quad \dot{x}(0) = 0.$$

Using these to find A and B we arrive at

$$\vec{r}(t) = L \cos \sqrt{\frac{k}{m}} t \hat{i}$$

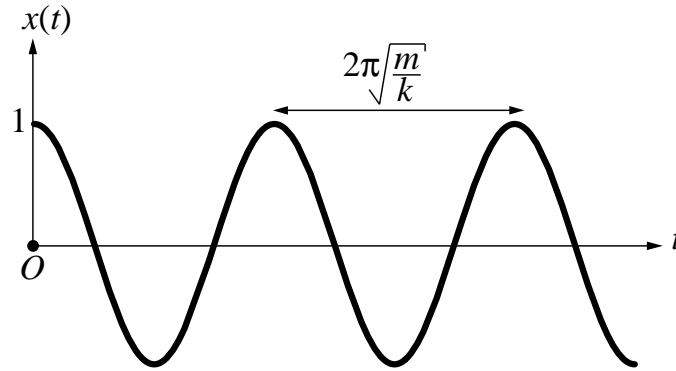


Figure 1.13: Simple harmonic motion for $L = 1$.

The motion is oscillatory, with amplitude L , and repeats after a time period of $2\pi\sqrt{m/k}$. Motion like this – that involves linear combinations of the harmonic functions cosines and sines – is called **simple harmonic motion** and is often referred to as a **harmonic oscillator**. If you actually did this experiment with a mass and a spring on a table then you would find that friction between the mass and the table would damp the oscillation away (we will work this out later).

1.7.4 Motion in a magnetic field

As a final example, let us consider a problem from a different part of physics. Newton's laws of motion don't just apply to situations that arise out of simple mechanical situations. As an illustration of this (and an example of a more general use of vectors), we turn to the motion of an electrically charged particle in a magnetic field, for example an electron or proton. A magnetic field can be

represented by a vector field denoted by \vec{B} . We shall assume that the strength of the magnetic field can be increased to the point where the effect of gravity on the charged particle can be neglected.

The magnetic force \vec{F} acting on a particle, which has an electrical charge q and is moving at velocity $\dot{\vec{r}}(t) = \vec{v}(t)$ in a field \vec{B} , is given (experimentally) by the vector product¹⁴

$$\vec{F} = q \left(\vec{v} \times \vec{B} \right).$$

We are interested in determining the path the particle takes in a particular magnetic field. We shall suppose the particle is of constant mass m and that the magnetic field (constant throughout the region of interest) is of the form $\vec{B} = B\hat{k}$ relative to the reference frame. Let us assume that initially the particle is located at $\vec{r}(0) = \vec{0}$ with velocity $\dot{\vec{r}}(0) = \hat{i} + \hat{k}$. We now use **N2** and the given form of the force to find the trajectory of the particle.

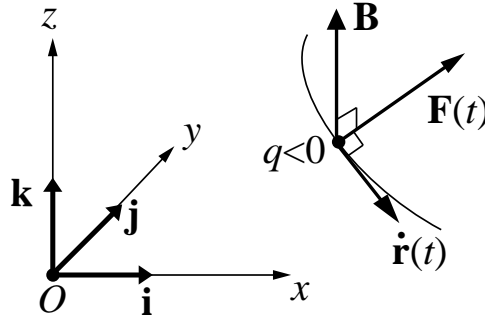


Figure 1.14: Forces acting on a negatively charged particle in a magnetic field.

N2 gives the equation of motion of the particle as

$$\begin{aligned} \vec{F} = m\ddot{\vec{r}}(t) &= q \left(\vec{v} \times \vec{B} \right) \\ &= q \left(\dot{\vec{r}} \times \vec{B} \right) \\ &= q \left(\dot{\vec{r}} \times B\hat{k} \right) \\ m\ddot{\vec{r}}(t) &= qB \left(\dot{\vec{r}} \times \hat{k} \right). \end{aligned}$$

This looks awful! Our approach to solving this problem will be to write everything out in component form and then try to solve the resulting equations. However, to show that it is possible to find some properties of the solution using vector manipulations we note that: If we take the scalar product of both sides of the **N2** equation with the velocity vector $\dot{\vec{r}}$ we get

$$\dot{\vec{r}} \cdot \ddot{\vec{r}} = \frac{qB}{m} \dot{\vec{r}} \cdot \left(\dot{\vec{r}} \times \hat{k} \right) = 0.$$

The right hand side of the equation vanishes because a vector product of any two vectors is, by definition, orthogonal to both the vectors and the dot product

¹⁴You do not have to remember this!

of any vector with a vector that is orthogonal to it is zero. Thus, we have

$$\dot{\vec{r}} \cdot \ddot{\vec{r}} = \frac{1}{2} \frac{d(\dot{\vec{r}} \cdot \dot{\vec{r}})}{dt} = \frac{1}{2} \frac{d|\dot{\vec{r}}|^2}{dt} = 0 \quad \Rightarrow \quad |\dot{\vec{r}}|^2 = \text{constant}.$$

We have found that, for this problem, the speed of the particle will be constant. However, this does not mean that there is no force acting on the particle – the direction of the velocity could still be changing.

To find the path of the particle we now write everything in the **N2** equation in component form—in this case it is a bit trickier to “integrate” the vector equation—using

$$\vec{r}(t) = x(t) \hat{i} + y(t) \hat{j} + z(t) \hat{k}.$$

Hence **N2** becomes

$$\begin{aligned} \ddot{x}(t) \hat{i} + \ddot{y}(t) \hat{j} + \ddot{z}(t) \hat{k} &= \frac{qB}{m} \begin{vmatrix} \hat{i} & \hat{j} & \hat{k} \\ \dot{x} & \dot{y} & \dot{z} \\ 0 & 0 & 1 \end{vmatrix} \\ &= \frac{qB}{m} (\dot{y} \hat{i} - \dot{x} \hat{j}). \end{aligned}$$

Rewriting the initial data in component form we have

$$\begin{aligned} \ddot{x}(t) &= \frac{qB}{m} \dot{y}, & x(0) &= 0, & \dot{x}(0) &= 1 \\ \ddot{y}(t) &= -\frac{qB}{m} \dot{x}, & y(0) &= 0, & \dot{y}(0) &= 0 \\ \ddot{z}(t) &= 0, & z(0) &= 0, & \dot{z}(0) &= 1. \end{aligned}$$

We must now solve these three ordinary differential equations. The z equation is easy, and using the initial conditions gives $z = t$. In other words, in the z direction, the particle moves at constant speed 1 ms^{-1} in the direction parallel to the magnetic field, passing through $z = 0$ when $t = 0$.

The other two equations are coupled: the x depends on y and vice-versa and we cannot solve one without the other. However, we can solve this system by eliminating variables to make a single higher-order differential equation. To eliminate $y(t)$ we differentiate the \ddot{x} equation once and then observe that the right-hand side only involves $\ddot{y}(t)$. Using the second component equation we can then write

$$\ddot{\ddot{x}} = \frac{qB}{m} \ddot{y} = \frac{qB}{m} \left(-\frac{qB}{m} \dot{x} \right) = -\left(\frac{qB}{m} \right)^2 \dot{x}.$$

This is a third order linear differential equation with constant coefficients. To solve this, let us introduce a new variable $u(t)$ which is the velocity in the x -direction and hence take $\dot{x} = u(t)$. Using this in the equation we have

$$\ddot{u} + \left(\frac{qB}{m} \right)^2 u = 0.$$

This is a second order linear harmonic oscillator-type equation, and we know how to deal with this. The standard approach leads to

$$u(t) = C \cos \lambda t + D \sin \lambda t, \quad \lambda = \frac{qB}{m},$$

for constants C and D . Since $u = \dot{x}$, we now just have to integrate this expression one more time with respect to t to obtain x .

$$x(t) = \frac{C}{\lambda} \sin \lambda t - \frac{D}{\lambda} \cos \lambda t + E, \quad \lambda = \frac{qB}{m}.$$

We can then immediately deduce $y(t)$ from the coupled equation above which gives

$$\begin{aligned} \dot{y} = \frac{1}{\lambda} \ddot{x} &\Rightarrow y(t) = \frac{1}{\lambda} \dot{x}(t) + F \\ y(t) = \frac{C}{\lambda} \cos \lambda t + \frac{D}{\lambda} \sin \lambda t + F, &\quad \lambda = \frac{qB}{m}. \end{aligned}$$

We now apply the initial data. Recall that we had:

$$\begin{aligned} x(0) &= 0 &\Rightarrow E &= \frac{D}{\lambda}, \\ \dot{x}(0) &= 1 &\Rightarrow C &= 1, \\ y(0) &= 0 &\Rightarrow F &= -\frac{C}{\lambda}, \\ \dot{y}(0) &= 0 &\Rightarrow D &= 0. \\ &&\Rightarrow C &= 1, D = 0, E = 0, F = -\frac{1}{\lambda} \end{aligned}$$

from which we deduce that

$$\vec{r}(t) = \frac{1}{\lambda} \sin \lambda t \, \hat{i} + \frac{1}{\lambda} (\cos \lambda t - 1) \, \hat{j} + t \, \hat{k},$$

Clearly the (x, y) components are related by the formula

$$x^2 + \left(y + \frac{1}{\lambda}\right)^2 = \frac{1}{\lambda^2}, \quad z = t$$

and we see that the path is a circular helix of radius $1/\lambda$, centred on the point $-\lambda^{-1} \hat{j}$ with vertical pitch (repeat) every $2\pi/\lambda$ seconds, see figure 1.15. We also observe that the total speed of the particle is given by

$$\begin{aligned} |\dot{\vec{r}}(t)| &= \sqrt{\dot{\vec{r}}(t) \cdot \dot{\vec{r}}(t)} \\ &= \sqrt{\left\{ \dot{x}(t) \hat{i} + \dot{y}(t) \hat{j} + \dot{z}(t) \hat{k} \right\} \cdot \left\{ \dot{x}(t) \hat{i} + \dot{y}(t) \hat{j} + \dot{z}(t) \hat{k} \right\}} \\ &= \sqrt{\cos^2 \lambda t + \sin^2 \lambda t + 1} \\ &= \sqrt{2}. \end{aligned}$$

As expected, this is constant.

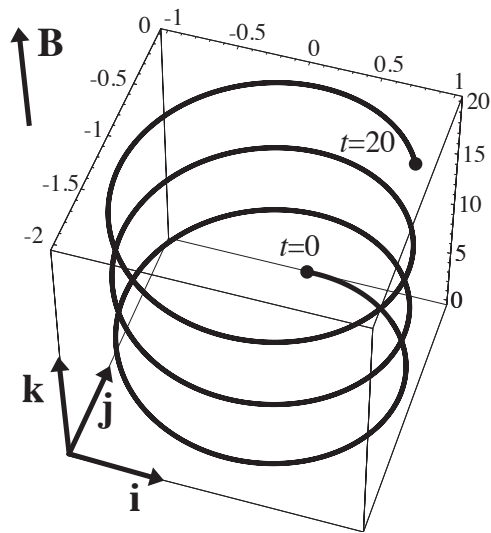


Figure 1.15: Motion of a charged particle in a magnetic field. As an exercise you may want to use the trajectory to deduce the direction of the velocity of the the particle and draw in the force acting on the particle at a chosen time.

Chapter 2

Air resistance and friction

So far we have explored forces that do not *per se* resist the motion—we have assumed that there is no friction or air resistance. This is not our everyday reality. While energy is conserved at a fundamental level, it does not appear to be conserved in many things we do. At a microscopic level, kinetic energy is transferred to the atoms in the floor you may be trying to slide across, manifesting itself as heat. But if we only want to know how far our socks will slide, the details of the atomic processes are of little interest. Instead, we often summarise everything in a single, macroscopic force that we call friction. This is a messy business, but the effective description is quite simple.

In this chapter we consider the effect of such forces. Newton's laws formally remain unchanged, but the presence of resistive forces complicates the problem we have to solve.

2.1 Air Resistance

A particle moving through a fluid (or a gas, such as air!) is subject to a drag force. This arises because in moving through the liquid, the particle has to push the fluid molecules out of the way.

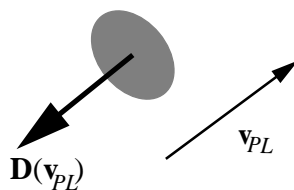


Figure 2.1: Resistive force $\vec{D}(\vec{v}_{PL})$ due to motion of body relative to liquid \vec{v}_{PL} .

Experimentally it has been found that a good model for the drag force \vec{D} is one that depends on the velocity of the particle P relative¹ to the fluid L , \vec{v}_{PL} . Such a model takes the form

$$\vec{D} = -K\vec{v}_{PL}$$

¹We will return to the issue of relative velocities later.

where K is a positive quantity that depends on the speed and shape of the particle as well as the density and viscosity of the fluid. Note that the drag force \vec{D} is, at all times, in the opposite direction to the velocity. A common form for this law is

$$\vec{D} = -k|\vec{v}_{PL}|^{(\alpha-1)}\vec{v}_{PL} = -k|\vec{v}_{PL}|^\alpha \hat{v}_{PL}$$

where the exponent α depends on the physical situation. For “high” speed flows, such as cars or planes moving through air, $\alpha \approx 2$. This is natural when the friction is proportional to the surface area of the moving object. For “low” speed flows, such as a pebble falling through thick oil, $\alpha \approx 1$. For “intermediate” speeds α may itself be a function of the speed. A special case, valid when the particle is moving very slowly, is $\alpha = 1$ in which case

$$\vec{D} = -k\vec{v}_{PL}$$

2.2 Reminder: Separable first order equations

The simplest differential equations are those that can be solved by direct integration. As an example, consider

$$y'(x) = f(y)$$

which can be solved in the following way: First we divide both sides by $f(y)$ (assuming initially that $f(y) \neq 0$ on the range of interest)

$$\frac{1}{f(y)} \frac{dy}{dx} = 1 \quad \Rightarrow \quad \int \frac{1}{f(y)} \frac{dy}{dx} dx = \int 1 dx$$

Now suppose that $H(y)$ is the integral of $1/f(y)$, i.e.,

$$H'(y) = \frac{1}{f(y)}$$

Then, using the chain rule we have

$$\frac{d}{dx} H(y(x)) = H'(y(x)) \frac{dy}{dx} = \frac{1}{f(y)} \frac{dy}{dx}.$$

Hence, by the fundamental theorem of calculus, the integral on the left-hand side above becomes

$$\int \frac{1}{f(y)} \frac{dy}{dx} dx = \int \left(\frac{d}{dx} H(y(x)) \right) dx = H(y) + c \quad (2.1)$$

If you look carefully at (2.1), effectively all we did was evaluate the integral of $1/f(y)$ with respect to y : the answer being the integral $H(y)$.

Note that you could have obtained the same result just by “cancelling” the dx top and bottom in the integral on the left-hand side and integrating with respect to dy . Hence, a short cut for the integration is as follows:

$$\frac{1}{f(y)} \frac{dy}{dx} = 1 \quad \Rightarrow \quad \int \frac{1}{f(y)} \frac{dy}{dx} dx = \int 1 dx \quad \Rightarrow \quad \int \frac{1}{f(y)} dy = x + C$$

The last integral with respect to y just has to be evaluated depending on what $f(y)$ is and C is an arbitrary constant.

However, this cancellation trick is (formally) quite naughty, since the derivative dy/dx is not a fraction that can be cancelled! Still, this notational short cut is justified due to the argument outlined above. Hence everybody does it. Just remember why it is allowed!

A related kind of first order equation contains functions of x and y such that it is possible to separate them into two terms, one on each side of the equal sign, one a function of x only and the other only a function of y . In general these equations take the form.

$$\frac{dy}{dx} = g(x)f(y).$$

Separating the terms we may integrate with respect to x only:

$$\int \frac{1}{f(y)} \frac{dy}{dx} dx = \int g(x) dx \Rightarrow \int \frac{1}{f(y)} dy = \int g(x) dx$$

Note that we are using the naughty cancellation trick between derivatives again. This is justified by the same procedure as outlined above.

2.3 Projectile under gravity with air resistance

Suppose a particle of constant mass m is launched from the surface of the earth at $t = 0$ with initial speed U at an angle θ up from the horizontal. We shall assume that the air is stationary. Suppose the position of the particle at time t is $\vec{r}(t)$ and that the motion is very slow (this makes the mathematics easier!) so that the resistive force due to the air is $\vec{D} = -\gamma\dot{\vec{r}}(t)$. Calculate the subsequent motion of the particle as it moves under the influence of gravity $-g\hat{k}$ and air resistance.

As before we use a coordinate system with \hat{i} horizontal and \hat{k} vertical with the origin at the point of release of the particle.

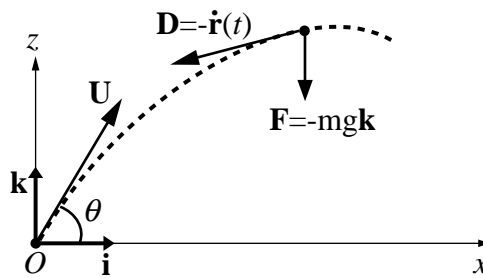


Figure 2.2: Projectile subject to drag force.

N2 then gives the equation of motion as

$$m \ddot{\vec{r}}(t) = -mg \hat{k} - \gamma \dot{\vec{r}}(t),$$

with

$$\vec{r}(t) = x(t) \hat{i} + y(t) \hat{j} + z(t) \hat{k}.$$

As in many cases, this problem can be solved using the second order equation above written in component form. However it is easier to introduce variables to indicate the velocity of the particle so we write

$$\dot{\vec{r}}(t) = u(t) \hat{i} + v(t) \hat{j} + w(t) \hat{k}$$

so that **N2** is now the system of equations

$$\begin{aligned} m\dot{u}(t) &= -\gamma u, & u(0) &= U \cos \theta \\ m\dot{v}(t) &= -\gamma v, & v(0) &= 0 \\ m\dot{w}(t) &= -\gamma w - mg, & w(0) &= U \sin \theta \\ \dot{x}(t) &= u(t) & x(0) &= 0 \\ \dot{y}(t) &= v(t) & y(0) &= 0 \\ \dot{z}(t) &= w(t) & z(0) &= 0 \end{aligned}$$

Now we solve this system of differential equations.

The first equation is separable and can be integrated:

$$\int_{U \cos \theta}^u \frac{du}{u} = -\frac{\gamma}{m} \int_0^t dt \Rightarrow [\ln u]_{U \cos \theta}^u = -\frac{\gamma t}{m} \Rightarrow u(t) = U \cos \theta \exp\left(-\frac{\gamma}{m}t\right).$$

Using the equation $u(t) = \dot{x}(t)$ we can integrate this expression once more and use the fact that $x(0) = 0$ to obtain $x(t)$

$$\dot{x}(t) = U \cos \theta \exp\left(-\frac{\gamma}{m}t\right), \quad x(0) = 0 \Rightarrow x(t) = \frac{mU \cos \theta}{\gamma} \left\{1 - \exp\left(-\frac{\gamma}{m}t\right)\right\}$$

The second equation can be solved in a similar manner

$$\begin{aligned} m \frac{dv}{dt} &= -\gamma v \Rightarrow \int \frac{dv}{v} = -\frac{\gamma}{m} \int dt \\ \Rightarrow \ln v &= -\frac{\gamma t}{m} + \ln A \Rightarrow v(t) = A \exp\left(-\frac{\gamma}{m}t\right) \\ v(0) &= 0, \quad \Rightarrow \quad A = 0 \quad \Rightarrow \quad v(t) = 0. \end{aligned}$$

Hence we have

$$\dot{y}(t) = 0, \quad \Rightarrow \quad y(t) = B, \quad \text{but } y(0) = 0 \quad \Rightarrow \quad B = 0 \quad \therefore y(t) = 0.$$

So the projectile always stays in the $y = 0$ plane and never has a non-zero \hat{j} component (this should have been obvious from the start!).

The third equation is also separable and can be solved as

$$\int_{U \sin \theta}^w \frac{dw}{(mg/\gamma + w)} = -\frac{\gamma}{m} \int_0^t dt \Rightarrow w(t) = -\frac{mg}{\gamma} + \left(U \sin \theta + \frac{mg}{\gamma}\right) \exp\left(-\frac{\gamma}{m}t\right)$$

This can be integrated to give the z -component as

$$\dot{z}(t) = -\frac{mg}{\gamma} + \left(U \sin \theta + \frac{mg}{\gamma}\right) \exp\left(-\frac{\gamma}{m}t\right)$$

$$\Rightarrow z(t) = -\frac{mgt}{\gamma} + \frac{m}{\gamma} \left(U \sin \theta + \frac{mg}{\gamma} \right) \left\{ 1 - \exp \left(-\frac{\gamma}{m} t \right) \right\}$$

The vector equation of the trajectory is thus

$$\vec{r}(t) = \frac{mU \cos \theta}{\gamma} \left\{ 1 - \exp \left(-\frac{\gamma}{m} t \right) \right\} \hat{i} + \left[-\frac{mgt}{\gamma} + \frac{m}{\gamma} \left(U \sin \theta + \frac{mg}{\gamma} \right) \left\{ 1 - \exp \left(-\frac{\gamma}{m} t \right) \right\} \right] \hat{k}.$$

which is much more complicated than the parabolic motion with no air resistance.

By considering the components $x(t)$ and $z(t)$ and eliminating t between them the Cartesian equation of the trajectory can be found to be

$$z = \left(\tan \theta + \frac{mg}{\gamma U \cos \theta} \right) x + \frac{m^2 g}{\gamma^2} \log \left(1 - \frac{\gamma x}{mU \cos \theta} \right).$$

Exercise: Take the solution above and put $\epsilon = \gamma U / mg$ and then expand the resulting expression in a Taylor series (see later) as $\epsilon \rightarrow 0$. For the first term in the series you should recover the resistance-free parabola and the next terms give you the correction to the path due to air resistance.

2.4 Reminder: Linear Equations and integrating factors

Many first order ordinary differential equations can be solved by multiplying both sides by a suitable factor, called an *integrating factor* $R(x, y)$. Finding such an integrating factor can be difficult but there is a simple rule for the case of a linear first-order ordinary differential equations.

An equation

$$\frac{dy}{dx} + p(x)y = q(x)$$

is called a *linear first order differential equation* and can be solved by use of an integrating factor. The integrating factor is given by

$$R(x) = \exp \left[\int^x p(s) ds \right].$$

Note that the integrating factor is usually taken not to contain an arbitrary constant (since this will be a multiplicative factor that can be cancelled on both sides of the equation).

We solve the equation by multiplying through by $R(x)$,

$$\exp \left(\int^x p(s) ds \right) \frac{dy}{dx} + \exp \left(\int^x p(s) ds \right) p(x)y = \exp \left(\int^x p(s) ds \right) q(x).$$

The left-hand side can then be recognised as an exact derivative:

$$\exp \left(\int^x p(s) ds \right) \frac{dy}{dx} + \exp \left(\int^x p(s) ds \right) p(x)y = \frac{d}{dx} \left[y \exp \left(\int^x p(s) ds \right) \right]$$

where we have used the fundamental theorem of calculus:

$$p(x) = \frac{d}{dx} \left(\int^x p(s) ds \right).$$

Hence, we can write the linear equation above as

$$\frac{d}{dx} \left(y \exp \left(\int^x p(s) ds \right) \right) = q(x) \exp \left(\int^x p(s) ds \right)$$

or more simply as

$$\frac{d}{dx} (y R(x)) = q(x) R(x).$$

Both sides can now be integrated directly with respect to x to obtain a solution for $y(x)$:

$$\begin{aligned} y(x)R(x) &= \int^x q(s)R(s)ds + C \\ \Rightarrow y(x) &= \frac{1}{R(x)} \int^x q(s)R(s)ds + \frac{C}{R(x)}. \end{aligned}$$

Note that if you chose not to use the initial conditions in the integration limits here, so that the arbitrary constant C is automatically found as part of the definite integration progress, it is vital that you don't forget to put C on the right hand side **before** you ultimately divide through by $R(x)$ to obtain $y(x)$. Otherwise you miss out a whole term in the solution.

Note: When seeking the integrating factor be careful. For example for the equation

$$x \frac{dy}{dx} - x^2 y = e^{3x},$$

first put the equation in the **standard** form with the coefficient of dy/dx being unity and a function of x multiplying a single linear term in y .

$$\frac{dy}{dx} - x y = \frac{e^{3x}}{x},$$

and then find the integrating factor (taking care to include the correct sign of $p(x)$)

$$R(x) = \exp \left(\int p(x) dx \right) = \exp \left(\int -x dx \right) = \exp (-x^2/2).$$

Also, you will have noticed that for a first order equation, there is one arbitrary constant in the general solution. Thus to find a specific solution you need to provide one additional bit of information, or boundary condition, to determine the specific value of the arbitrary constant.

2.5 Alternative solution to the projectile problem

While it is often natural to solve the problem in component form (as we did above), we have already seen that simple ballistic motion can also be solve directly as a vector equation. As before, we have

$$m \frac{d\vec{v}}{dt} = m\vec{g} - \gamma\vec{v}$$

This is a linear equation, so we can solve it using an integrating factor. We have

$$\frac{d}{dt} \left(e^{\frac{\gamma t}{m}} \vec{v} \right) = e^{\frac{\gamma t}{m}} \vec{g}$$

which leads to

$$\vec{v} = \frac{m}{\gamma} \vec{g} + \vec{c} e^{-\gamma t/m}$$

where the integration constant, \vec{c} , is fixed by the initial speed. Suppose that we have $\vec{v} = \vec{u}$ at $t = 0$. Then it follows that

$$\vec{v} = \frac{m}{\gamma} \vec{g} + \left(\vec{u} - \frac{m}{\gamma} \vec{g} \right) e^{-\gamma t/m}$$

If we want the position, we can integrate again. This leads to

$$\vec{r} = \frac{m}{\gamma} \vec{g} t - \frac{m}{\gamma} \left(\vec{u} - \frac{m}{\gamma} \vec{g} \right) e^{-\gamma t/m} + \vec{b}$$

If we let $\vec{r} = 0$ at $t = 0$, then

$$\vec{r} = \frac{m}{\gamma} \vec{g} t + \frac{m}{\gamma} \left(\vec{u} - \frac{m}{\gamma} \vec{g} \right) \left(1 - e^{-\gamma t/m} \right)$$

We can make contact with the previous calculation by expanding our result in components. Let $\vec{r} = (x, y, z)$ and send the projectile off with initial velocity $\vec{u} = (u \cos \theta, 0, u \sin \theta)$. With gravity acting downwards, so $\vec{g} = (0, 0, -g)$, our vector equation becomes three equations. One is trivial: $y = 0$. The other two are

$$x = \frac{m}{\gamma} u \cos \theta \left(1 - e^{-\gamma t/m} \right)$$

and

$$z = -\frac{mgt}{\gamma} + \frac{m}{\gamma} \left(u \sin \theta + \frac{mg}{\gamma} \right) \left(1 - e^{-\gamma t/m} \right)$$

The time scale² m/γ is important. For $t \gg m/\gamma$, the horizontal position is essentially constant. By this time, the particle is dropping more or less vertically. The trajectory is not symmetric.

We may also consider what happens when friction is small. A couple of terms look as if they are going to become singular (as we would appear to divide by zero) in this limit, but that seems unphysical. To resolve this, we should ask what γ is small relative to. In the present case, the answer lies in the exponential terms. To say that γ is small, really means $\gamma \ll m/t$ or, in other words, it means that we are looking at short times, $t \ll m/\gamma$. Then we can expand the exponential in a Taylor series. Reverting to the vector form of the equation, we have

$$e^{-\gamma t/m} \approx 1 - \frac{\gamma t}{m} + \dots$$

so we find

$$\begin{aligned} \vec{r} &= \frac{m}{\gamma} \vec{g} t + \frac{m}{\gamma} \left(\vec{u} - \frac{m}{\gamma} \vec{g} \right) \left[1 - 1 + \gamma \frac{t}{m} - \frac{1}{2} \left(\frac{\gamma t}{m} \right)^2 + \dots \right] \\ &= \vec{u} t + \frac{1}{2} \vec{g} t^2 + \mathcal{O} \left(\frac{\gamma t}{m} \right) \end{aligned}$$

This is the usual projectile result.

²Note that $\gamma t/m$ must be dimensionless.

2.6 Terminal velocity

Let us now consider a particle of mass m moving in a constant gravitational field, subject to quadratic friction. We measure the height z to be in the upwards direction, meaning that if $v = dz/dt > 0$, the particle is going up. We will look at the cases where the particle goes up and goes down separately.

First suppose that we drop the particle from some height. The equation of motion is given by

$$m \frac{dv}{dt} = -mg + \gamma v^2$$

It is worth commenting on the minus signs on the right-hand side. Gravity acts downwards, so comes with a minus sign. Since the particle is falling down, friction is acting upwards so comes with a plus sign. Dividing through by m , we have

$$\frac{dv}{dt} = -g + \frac{\gamma}{m} v^2$$

Integrating this, we get

$$t = - \int_0^v \frac{dv'}{g - \gamma v'^2/m}$$

This can be worked out—either through the substitution $v = \sqrt{mg/\gamma} \tanh x$ or partial fractions (which may be the easiest route)—leading to

$$t = -\sqrt{\frac{m}{\gamma g}} \tanh^{-1} \left(\sqrt{\frac{\gamma}{mg}} v \right)$$

which we can invert to get the speed as a function of time

$$v = -\sqrt{\frac{mg}{\gamma}} \tanh \left(\sqrt{\frac{\gamma g}{m}} t \right)$$

This illustrates the effect of the air resistance. As time increases, the velocity does not increase without bound. Instead, the particle reaches a maximum speed (the terminal velocity)

$$v \longrightarrow -\sqrt{\frac{mg}{\gamma}}, \quad \text{as } t \rightarrow \infty$$

The sign is negative because the particle is falling downwards. Note that if all we want is the terminal velocity, then we do not need to go through the whole calculation. We can simply look for solutions with constant speed, so $dv/dt = 0$. This gives us the same answer. The advantage of going through the full calculation is that we learn how the velocity approaches its terminal value.

This example teaches us an important lesson. A small animal (like a mouse) may survive a fall where a large animal (say, an elephant) will not³. The point is that, if we compare objects of equal density the masses scale as the volume, meaning $m \sim L^3$ where L is the linear size of the object. In contrast, the coefficient of friction usually scales as surface area, $\gamma \sim L^2$. This means that the terminal velocity depends on size. For objects of equal density, we expect the terminal velocity to scale as $v \sim \sqrt{L}$.

³While it would be interesting to demonstrate this experimentally, health and safety regulations prevent us from doing so at this time.

Let us now throw the object upwards. Since both gravity and friction are then acting downwards, we get a flip of a minus sign in the equation of motion. It is now

$$\frac{dv}{dt} = -g - \frac{\gamma}{m}v^2$$

Suppose that we throw the object up with initial speed u and we want to figure out the maximum height, h , that it reaches. We could follow our earlier calculation and integrate the equation to determine $v = v(t)$. But since we are not asking about time, it is much better to instead consider velocity as a function of distance: $v = v(z)$. We write⁴

$$\frac{dv}{dt} = \frac{dv}{dz} \frac{dz}{dt} = v \frac{dv}{dz} = -g - \frac{\gamma v^2}{m}$$

or

$$\frac{1}{2} \frac{dv^2}{dz} = -g - \frac{\gamma v^2}{m}$$

Using $y = v^2$ we can integrate this to get

$$\int_{u^2}^0 \frac{dy}{g + \gamma y/m} = -2 \int_0^h dz \quad \longrightarrow \quad \frac{m}{\gamma} \left[\log \left(g + \frac{\gamma y}{m} \right) \right]_{y=u^2}^{y=0} = -2h$$

That is,

$$h = \frac{m}{2\gamma} \log \left(1 + \frac{\gamma u^2}{mg} \right)$$

Here, it is worth looking at what happens when the effect of friction is small. Naively, it looks like we are in trouble because as $\gamma \rightarrow 0$, the term in front gets large. But surely the height should not go to infinity just because the friction is small. The resolution to this is that the log is also getting small in this limit. Expanding as a Taylor series, we have

$$h = \frac{u^2}{2g} \left(1 - \frac{\gamma u^2}{2mg} + \dots \right)$$

The leading term is the answer we would get in the absence of friction; the subleading terms tell us how much the friction, γ , affects the attained height.

2.7 Friction

Friction is the resistive force that occurs when two solid surfaces pressed together attempt to slide over each other. Even the apparently macroscopically smoothest of surfaces will, in practice, give rise to frictional forces, since on the macroscopic scale the surfaces are imperfect and so are rough. The surfaces then snag on and scrap against each other's microscopic mountain ranges as they move past each other.⁵

⁴Strictly speaking, we are using $v(z(t))$ here.

⁵The effect of lubricating the surfaces reduces the friction by separating the rough surfaces with a thin layer of fluid (for example, oil). With the solid surfaces are no longer in microscopic contact the problem is effectively reduced to one of sliding a fluid under pressure over a surface, which requires less effort.

Consider two solid, reasonably flat surfaces in contact. Assume they are both stationary and are being pushed together by a normal force \vec{N} . This normal force is the component orthogonal to the surfaces of any force holding the surfaces together. For example you might consider the upper surface to be a book and the lower surface to be a horizontal table in which case the normal force would be due to the weight of the book.



Figure 2.3: Cartoon close-up of the interface between two apparently flat surfaces.

As there is no acceleration we can apply **N2** to infer that there will be a reactive force \vec{R} due to the lower surface acting on the upper surface (the table pushing on the book) of equal magnitude to \vec{N} , but in the opposite direction.

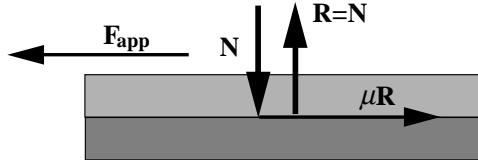


Figure 2.4: Effect of applying forces to the frictional interface between two surface forced together.

The rough surface creates a frictional force \vec{F}_F that acts tangentially to each surface to oppose any translational force applied in an orthogonal direction to \vec{R} (see diagram). If a translational force, \vec{F}_{app} , is applied to the upper surface (the book) then the surfaces will not slide over each other and hence will remain stationary only as long as

$$|\vec{F}_{\text{app}}| \leq \mu_s |\vec{R}|,$$

where μ_s is an experimentally determined positive number called the **coefficient of static friction**. If \vec{F}_{app} is greater in magnitude than the static frictional force then the surfaces will slide over each other. Once the surfaces start to slide there will still be a frictional force tangential to the surfaces but it will have a magnitude

$$|\vec{F}_F| = \mu_k |\vec{R}|,$$

and act in a direction to oppose the sliding motion. The experimentally determined positive number μ_k is called the **coefficient of kinetic friction** (also sometimes known as the coefficient of dynamic friction). Suppose the instantaneous motion of one surface relative to another in contact with it is in the direction $\hat{v} \neq \vec{0}$. In vector notation the friction force resisting the motion can then be written as

$$\vec{F}_F = -\mu_k |\vec{R}| \hat{v}.$$

This force continues to act until the surfaces become stationary at which time the static friction returns.

Several points are worth making:

- the coefficients of friction are properties of the surfaces in contact. Hence the coefficient of friction between two wooden surfaces can be expected to be different from that between a steel and wood surface.
- Frictional forces are independent of the area of the surface in contact, they depend only on the magnitude of the normal force with which the surfaces are pushed together.
- Usually $\mu_k \leq \mu_s$, however to simplify algebra they are often taken as being equal and take a value less than 1, i.e.,

$$\mu_k = \mu_s < 1.$$

We shall follow this procedure here. Note that coefficients of friction greater than unity only occur if the surfaces interlock and that a book sliding on a table might have values around 0.2.

- Note that there is also similarly defined **coefficient of rolling friction** μ_s which is applicable to situations where two surfaces are only in minimal contact, for example a cart wheel in contact with the road. However we shall not deal so much with this situation.
- We can attack problems involving friction using Newton's laws of motion.

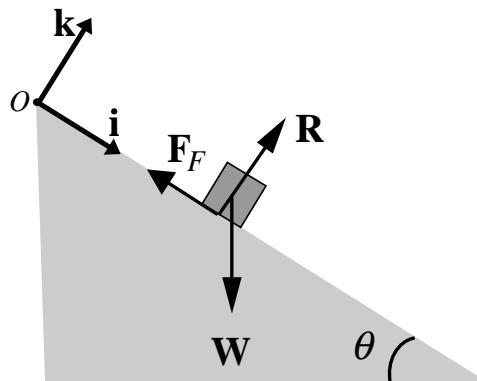


Figure 2.5: A, somewhat square, child sliding down a slide.

Example 2: A child on a slide

Consider a straight slide inclined at an angle θ to the horizontal. Suppose a child of mass m attempts to slide down the slide under gravity of magnitude g . The forces acting on the child are

- their weight \vec{W} which has magnitude mg and acts vertically downwards;
- a force of reaction \vec{R} that acts normal to the slide (that this exists is due to **N3** as without it the child would fall through the slide);
- the frictional force \vec{F}_F acting tangentially to the slide and in a direction (up or down the slide) to oppose any motion. The coefficient of friction is supposed to be μ .

These forces are summarised in the diagram below.

We can use **N2** to analyse the situation. We fix a Cartesian coordinate frame at the base of the slide and orient the axes such that \hat{i} points down the tangent to the slide, \hat{j} points into the page and \hat{k} points away from the ground but normal to the slide.

Note that there is no fixed rule that says that a Cartesian coordinate system has to be oriented so that \hat{k} has to point directly upwards! In fact for this problem this orientation saves a lot of algebraic effort. If you don't believe this, try orienting the coordinates with \hat{i} horizontal and \hat{k} pointing upwards and repeating the analysis.

Then we can write the normal and frictional forces as

$$\vec{F}_F = -F_F \hat{i}, \quad \vec{R} = R \hat{k}.$$

If the position of the child on the slide is given by $\vec{r}(t) = x(t)\hat{i} + y(t)\hat{j} + z(t)\hat{k}$ with respect to this coordinate frame then **N2** says that

$$m \ddot{\vec{r}}(t) = \vec{F}_F + \vec{R} + \vec{W}.$$

where

$$\vec{W} = mg \sin \theta \hat{i} + 0\hat{j} - mg \cos \theta \hat{k}$$

To solve this problem we do exactly as before and rewrite the problem in component form. By symmetry, $y(t) = 0$ for all time, so we will not need to solve for $y(t)$. We find:
 \hat{i} component:

$$m \ddot{x}(t) = mg \sin \theta - F_F.$$

\hat{k} component:

$$m \ddot{z}(t) = R - mg \cos \theta.$$

Assuming that the child does not spontaneously fly off the slide (i.e., we assume no motion in the \hat{k} direction) we must have no acceleration in this direction $\Rightarrow \ddot{z}(t)=0$. Hence this last equation reduces to

$$R = mg \cos \theta,$$

and gives us an expression relating the normal reaction R to the weight mg .

Two possible situations can arise.

1. The child is just stationary, on the point of sliding. In this case we must have $\ddot{x}(t) = 0$:

$$m \ddot{x}(t) = 0 \quad \Rightarrow \quad F_F = mg \sin \theta.$$

Using the fact that, in this case, $F_F = \mu R$ we combine this result with the z -component to arrive at:

$$R = mg \cos \theta, \quad F_F = mg \sin \theta, \quad F_F = \mu R \quad \Rightarrow \quad \tan \theta = \mu,$$

$$\theta = \arctan \mu.$$

This is a (rather scary for the child) way to determine μ_s . The angle of the slide θ is altered until the child first just starts to move. μ_s is then just the tangent of the angle of inclination at which slipping first occurs.

2. The child slides downwards. The equations of motion must be solved subject to some initial data, say

$$\vec{r}(0) = -L\hat{i}, \quad \dot{\vec{r}}(0) = \vec{0}.$$

Once again it is best to rewrite the problem using $\dot{\vec{r}}(t) = u(t)\hat{i} + v(t)\hat{j} + w(t)\hat{k}$ as it makes the algebra easier and leads to the following, where we have again used $F_F = \mu R = \mu mg \cos \theta$:

- \hat{i} component:

$$m\dot{u}(t) = mg \sin \theta - \mu mg \cos \theta, \quad u(0) = 0 \quad \Rightarrow \quad u(t) = g(\sin \theta - \mu \cos \theta)t.$$

- \hat{k} component:

$$m\dot{w}(t) = R - mg \cos \theta = 0, \quad w(0) = 0 \Rightarrow w(t) = 0.$$

Integrating once more with respect to t we obtain:

- \hat{i} component:

$$\dot{x}(t) = g(\sin \theta - \mu \cos \theta)t, \quad x(0) = -L \rightarrow x(t) = -L + g(\sin \theta - \mu \cos \theta)\frac{t^2}{2}$$

- \hat{k} component:

$$m\dot{z}(t) = 0, \quad z(0) = 0 \rightarrow z(t) = 0.$$

Hence the position as a function of time is

$$\vec{r}(t) = \left[-L + g(\sin \theta - \mu \cos \theta)\frac{t^2}{2} \right] \hat{i}.$$

Note: If the angle of the slope were increased to $\theta = \pi/2$ the slide would be vertical and the child would no longer be held in contact with the slide by component of its weight. In the equation above, the terms involving μ vanish and the equation of motion becomes

$$\vec{r}(t) = \left(-L + g\frac{t^2}{2} \right) \hat{i}$$

which is just the usual equation of a body falling under gravity (with the height measured downwards!) If the slide is horizontal with $\theta = 0$, then the above dynamic analysis does not hold.

2.8 Addendum: Taylor expansions

The Taylor series is one of the most useful mathematical tools applied to real life problems. As in the example of potential motion near equilibrium, we can often use a Taylor series to approximate a complicated function by a (simpler) polynomial function.

The starting point is Taylor's theorem (which we state without proof, as we are mainly interested in the applications):

Let $f(x)$ be continuous and differentiable an infinite number of times at point $x = a$ (i.e., f and all of its derivatives f', f'', \dots exist at $x = a$). Then $f(x)$ can be expanded about the point $x = a$ in terms of powers of $(x - a)$, and this expansion is given by the formula

$$f(x) = f(a) + (x - a)f'(a) + \frac{(x - a)^2}{2!}f''(a) + \dots + \frac{(x - a)^n}{n!}f^{(n)}(a) + O(x - a)^{n+1}$$

This is a **Taylor expansion**. It has the form

$$f(x) = P_n(x) + O(x - a)^{n+1}$$

where $P_n(x)$ is a polynomial of degree n and $O(x - a)^{n+1}$ is a remainder term. The remainder term reminds us that in writing the Taylor expansion explicitly up to the term $\propto (x - a)^n$ ("up to order n ") we are omitting terms "of order $(x - a)^{n+1}$ and higher".

For x close to a , we can expect the remainder term to be small (since $x - a$ is small), so that $f(x) \simeq P_n(x)$. Thus Taylor's theorem enables us to approximate functions by polynomials.

In many cases the remainder term goes to zero as $n \rightarrow \infty$ and the Taylor series expansion of $f(x)$ takes the form of a possibly infinite, but convergent, series.

Example 3: The exponential

As an example, let us determine the Taylor series expansion of the function $f(x) = e^x$ about $x = 0$.

We first establish the necessary data. We know that we need the derivatives:

$$\begin{aligned} f(x) &= e^x & f'(x) &= e^x & f''(x) &= e^x & \dots \\ f(0) &= e^0 = 1 & f'(0) &= e^0 = 1 & f''(0) &= e^0 = 1 & \dots \end{aligned}$$

Then we use this in the Taylor series expansion with $a = 0$

$$f(x) = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!} + \dots = \sum_{n=0}^{\infty} \frac{1}{n!} x^n$$

Similarly, we can work out the Taylor series expansion of the function $f(x) = \cos x$ about $x = 0$.

In this case, we need:

$$\begin{aligned}f(x) &= \cos x, & f(0) &= 1 \\f'(x) &= -\sin x, & f'(0) &= 0 \\f''(x) &= -\cos x, & f''(0) &= -1 \\f'''(x) &= \sin x, & f'''(0) &= 0 \\f^{(4)}(x) &= \cos x, & f^{(4)}(0) &= 1\end{aligned}$$

and we arrive at

$$f(x) = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \cdots = \sum_{k=0}^{\infty} \frac{(-1)^k}{(2k)!} x^{2k}$$

Chapter 3

Oscillations and resonances

We have already seen the simple example of a harmonic oscillator. With a better understanding of simple differential equations, we can turn to more concrete applications and consider more complicated oscillating systems.

3.1 Reminder: Homogeneous second order constant coefficient equations

Second order differential equations are more complicated—we can no longer get away with “integrating”—but we can readily deal with a particular class of equations: linear equations with constant coefficients.

In general we have

$$a \frac{d^2 x}{dt^2} + b \frac{dx}{dt} + cx = 0$$

First we take a homogeneous equation (no driving force) and divide through by a (this is allowed as a – the coefficient multiplying the second derivative – is non-zero which must be true, otherwise the equation would not be second order!)

$$\frac{d^2 x}{dt^2} + \frac{b}{a} \frac{dx}{dt} + \frac{c}{a} x = 0 \quad \Rightarrow \quad \frac{d^2 x}{dt^2} + m \frac{dx}{dt} + nx = 0$$

where m and n are constants.

To find the general solution we make the **ansatz** that there is a solution of the form:

$$x = Ae^{kt}$$

where A and k are constants that we will try to find. Now

$$\frac{dx}{dt} = Ake^{kt}, \quad \frac{d^2 x}{dt^2} = Ak^2 e^{kt}$$

Thus if $x = A \exp(kt)$ is a solution of the homogeneous equation by substituting it into the equation we see that we must have

$$Ak^2 e^{kt} + mAke^{kt} + nAe^{kt} = 0$$

The $\exp(kt)$ can be divided through as it is always non-zero for finite t . We can also divide through by A , since this must be non-zero for a **non-trivial**

solution (a solution that is not $x(t) \equiv 0$). This leaves what is called the **auxiliary** equation of the differential equation (occasionally called the characteristic equation):

$$k^2 + mk + n = 0$$

This is a quadratic in k which has two roots k_1, k_2 .

$$k_1 = \frac{-m + \sqrt{m^2 - 4n}}{2}, \quad k_2 = \frac{-m - \sqrt{m^2 - 4n}}{2}$$

Thus we typically have two possible solutions to the original equation $x_1 = A \exp(k_1 t)$ and $x_2 = A \exp(k_2 t)$.

We can check that the solutions are linearly independent¹ by computing the Wronskian²:

$$W[e^{k_1 t}, e^{k_2 t}] = \begin{vmatrix} e^{k_1 t} & e^{k_2 t} \\ k_1 e^{k_1 t} & k_2 e^{k_2 t} \end{vmatrix} = (k_2 - k_1)e^{(k_1 + k_2)t} \neq 0 \text{ provided } k_1 \neq k_2.$$

Hence as long as the roots of the auxiliary equation are not identical, we have a set of linearly independent solutions.

It is a simple exercise of substitution to deduce that a linear combination of these two solutions is also a solution. Hence we have found that

$$x = Ae^{k_1 t} + Be^{k_2 t}$$

is a solution of the differential equation where A and B are **any** two (very loosely: integration) constants. This is the **general** solution for most equations of this type.

The two arbitrary constants A and B can only be determined by the **boundary** or **initial conditions**, as we shall see below.

From the formula for the solution of the auxiliary equation we can see that the nature of the solution of the equation will depend on whether $m^2 > 4n$, $m^2 = 4n$, or $m^2 < 4n$ (in exactly the same way as the roots of a quadratic equation depend on this quantity). We now consider each of these cases in turn.

3.1.1 $m^2 > 4n$: real and distinct roots

When $m^2 > 4n$ the auxiliary equation gives two unequal real solutions for k and hence the general solution can be written as above:

$$x = Ae^{k_1 t} + Be^{k_2 t}.$$

3.1.2 $m^2 = 4n$: real and equal roots

When $m^2 = 4n$, the roots k of the auxiliary equation are real and equal. Let these roots be called k_1 . In this particular case, it is not possible to form a general solution with two arbitrary constants, since we could rewrite it in the following form:

$$x = Ae^{k_1 t} + Be^{k_1 t} = (A + B)e^{k_1 t} = Ce^{k_1 t}$$

¹This simply means that they are not proportional to one another.

²This is a new concept. We will return to it later.

So that effectively there would be only one arbitrary constant. Put another way, in this case the Wronskian vanishes everywhere, since $k_1 = k_2$. Thus we do not have the most general form (which needs two linearly independent, solutions).

We thus modify the guessed form of the solution to a new form³:

$$x = (A + Bt) e^{k_1 t}$$

Substitution into the differential equation shows that this is in fact a general solution provided that

$$k_1^2 + mk_1 + n = 0 \quad \text{with} \quad m^2 = 4n,$$

i.e., the auxiliary equation is satisfied.

The Wronskian of the two parts of the general solution is given by

$$W[e^{k_1 t}, te^{k_1 t}] = \begin{vmatrix} e^{k_1 t} & te^{k_1 t} \\ k_1 e^{k_1 t} & (1 + k_1 t)e^{k_1 t} \end{vmatrix} = e^{2k_1 t} \neq 0$$

for finite t , and so $e^{k_1 t}$ and $te^{k_1 t}$ are linearly independent. Hence this is a general solution of the case with equal roots.

3.1.3 $m^2 < 4n$, complex roots

When $m^2 < 4n$ the auxiliary equation has no real roots. However for m and n real the complex roots k_1 and k_2 appear as a complex conjugate pair. The roots are of the form

$$k_1 = \frac{-m + i\sqrt{4n - m^2}}{2}, \quad k_2 = \frac{-m - i\sqrt{4n - m^2}}{2}$$

where $i = \sqrt{-1}$. Hence the general solution is

$$\begin{aligned} x &= Ae^{t(-m + i\sqrt{4n - m^2})/2} + Be^{t(-m - i\sqrt{4n - m^2})/2} \\ &= e^{-mt/2} \left(Ae^{it\sqrt{4n - m^2}/2} + Be^{-it\sqrt{4n - m^2}/2} \right) \end{aligned}$$

Note that this looks like a complex (and complicated!) solution, even though we started off from a real equation, that only involved real numbers, but this is not necessarily the case. Now recall the following facts:

$$\sin z = \frac{e^{iz} - e^{-iz}}{2i}, \quad \cos z = \frac{e^{iz} + e^{-iz}}{2}$$

Thus we can write the above solution as

$$\begin{aligned} x &= e^{-mt/2} \left((A + B) \cos \left(\frac{\sqrt{4n - m^2}}{2} t \right) + i(A - B) \sin \left(\frac{\sqrt{4n - m^2}}{2} t \right) \right) \\ &= e^{-mt/2} \left(C \cos \left(\frac{\sqrt{4n - m^2}}{2} t \right) + D \sin \left(\frac{\sqrt{4n - m^2}}{2} t \right) \right). \end{aligned}$$

where C and D are arbitrary constants, formed from A and B , $C = A + B$, $D = i(A - B)$. This equation “looks” real. In fact real boundary data will generate real values of C and D .

³You can derive this, but we will not do that here.

3.2 Damped Oscillations

Damped oscillations play an important role in mechanical systems. A prime example is the shock absorber in a car. This can be modelled fairly accurately by combination of a spring and a “dashpot”.

A dashpot is a plunger in a pot of viscous liquid. A linear dashpot provides a resistive force proportional to the velocity of plunger, the constant of proportionality being λ . A Hookean spring with spring constant k and natural length l provides a restoring force proportional to a displacement.

As the car of mass m moves along, it encounters a bump, or variation in z in the road, the spring component compresses, and the plunger pushes into the dashpot. We want to know how the car responds to the bump so we can try to design the car to give the passengers a smooth ride.

We will take each of the car’s four wheels to support a mass $m/4$. This mass will have three forces acting on it namely: gravity, the spring and the damper. We will only consider movement and forces in the vertical direction. Our coordinate system will take \hat{k} to be in the vertical direction with the point $z = l$ representing the position of the car when the spring is extended to its natural length.

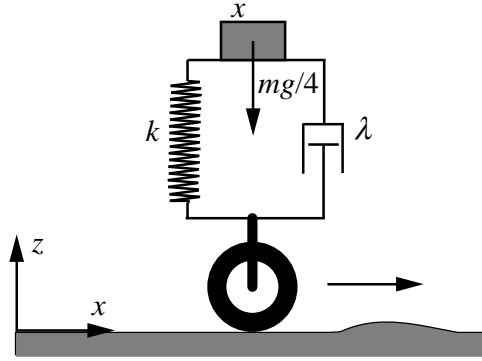


Figure 3.1: A schematic representation of one of the shock absorbers on a 4-wheeled car.

In the upwards vertical direction, **N2** then gives

$$\frac{m}{4}\ddot{z}(t) = -k(z - l) - \lambda\dot{z} - \frac{mg}{4}$$

Rearranging this we get

$$\ddot{z}(t) + \frac{4\lambda}{m}\dot{z} + \frac{4k}{m}z = \frac{4kl}{m} - g = \text{constant}$$

This is a second order linear constant coefficient ordinary differential equation. It can be represented schematically in terms of a linear operator \mathcal{L} and a forcing function f .

$$\mathcal{L}z = f$$

where

$$\mathcal{L} = \left(\frac{d^2}{dt^2} + \frac{4\lambda}{m} \frac{d}{dt} + \frac{4k}{m} \right)$$

$$f = \frac{4kl}{m} - g$$

We know how to solve this type of equation. We can break the solution down into two parts.

$$z = z_{\text{CF}} + z_{\text{PI}}$$

where

- “CF” stands for “complementary function” and satisfies

$$\mathcal{L}z_{\text{CF}} = 0$$

- “PI” stands for “particular integral” and satisfies

$$\mathcal{L}z_{\text{PI}} = f$$

Clearly we still have

$$\mathcal{L}z = \mathcal{L}(z_{\text{CF}} + z_{\text{PI}}) = 0 + f = f.$$

We solve for z_{PI} by an “educated guess”. Any derivative of a constant vanishes so we can guess that a possible solution to the equation is a constant. Putting this guess into the ODE we see:

$$\frac{4kz_{\text{PI}}}{m} = \frac{4kl}{m} - g \quad \Rightarrow \quad z_{\text{PI}} = l - \frac{mg}{4k}$$

This part of the solution corresponds to the position of the car if everything is stationary and the spring is compressed from its natural length due to the weight of the car.

Next we solve for z_{CF} by assuming there is a solution of the form $z = e^{\alpha t}$ where α is a constant. Substitution into the homogeneous equation

$$\ddot{z}(t) + \frac{4\lambda}{m}\dot{z} + \frac{4k}{m}z = 0$$

generates the auxiliary equation

$$\alpha^2 + \frac{4\lambda}{m}\alpha + \frac{4k}{m} = 0,$$

and hence α must be given by the solution to this quadratic

$$\alpha = -\frac{2\lambda}{m} \pm \frac{2\lambda}{m} \sqrt{1 - \frac{mk}{\lambda^2}}.$$

The complementary function is therefore of the form

$$z_{\text{CF}} = e^{-\frac{2\lambda}{m}t} \left(Ae^{\frac{2\lambda}{m}\sqrt{1 - \frac{mk}{\lambda^2}}t} + Be^{-\frac{2\lambda}{m}\sqrt{1 - \frac{mk}{\lambda^2}}t} \right).$$

where A and B depend on initial data $z(0), \dot{z}(0)$.

Note that the precise form of this solution will depend on whether $1 - (mk)/\lambda^2$ is positive or negative. Hence the size of the positive quantity mk/λ^2 is crucial and we now discuss the different possible cases.

If $mk/\lambda^2 > 1$ the values of α are complex and so the solution can be rewritten as a decaying oscillatory one.

$$z_{CF} = e^{-\frac{2\lambda}{m}t} (C \cos \omega t + D \sin \omega t), \quad \omega = \frac{2\lambda}{m} \sqrt{\frac{mk}{\lambda^2} - 1} > 0$$

This situation is said to be **underdamped**: the spring is dominating the response of the shock system. Any bump in the road initiates a long tail of oscillations in the shock system. It would be best not to attempt to try to drink anything while travelling in a car with this type of shock absorbing system. Indeed such a car would fail its annual MOT test of roadworthiness.

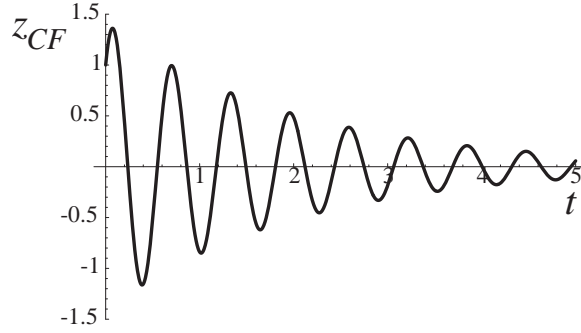


Figure 3.2: Schematic plot of response of underdamped shock system.

If $mk/\lambda^2 < 1$ then both values of α are real and negative. Hence the the solution is the sum of two exponential decaying functions.

$$z_{CF} = e^{-\frac{2\lambda}{m}t} \left(A e^{\frac{2\lambda}{m} \sqrt{1 - \frac{mk}{\lambda^2}} t} + B e^{-\frac{2\lambda}{m} \sqrt{1 - \frac{mk}{\lambda^2}} t} \right),$$

where A and B are constants that depend on the initial data $z(0), \dot{z}(0)$. This situation is said to be **overdamped**: the dashpot is dominating the response of the shock system. The cushioning effect of the spring is effectively absent and the effect of bumps in the road will be felt over a long timescale.

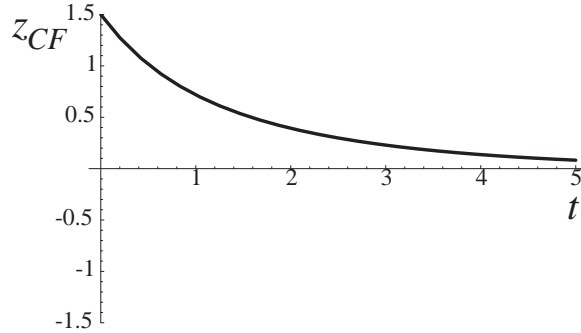


Figure 3.3: Schematic plot of response of overdamped shock system.

If $mk/\lambda^2 = 1$ mechanically the effect of the spring and dashpot are matched. Mathematically, the solution of the auxiliary equation is degenerate and only

one exponential effectively exists. From our discussion of the methods used to solve the equation we know that the form of the solution will then be slightly different and takes the form

$$z_{CF} = e^{-\frac{2\lambda}{m}t} (At + B),$$

where A and B depend on initial data $z(0), \dot{z}(0)$. This situation is said to be **critical damping**: the linear term ensures that there is some cushioning, with effectively one and only one oscillation in response to a bump, but no long-term oscillatory behaviour⁴. This one-oscillation behaviour is what MOT inspectors look for when they test the car.

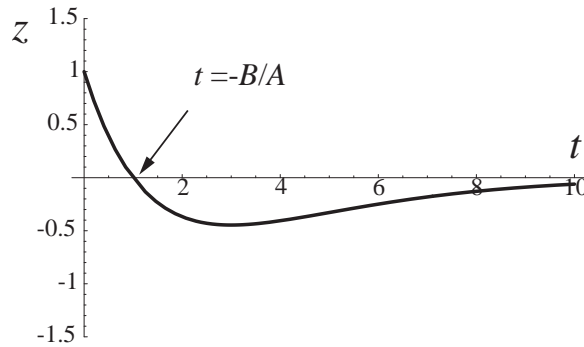


Figure 3.4: Schematic plot of response of critically damped shock system.

The ideas of underdamped, overdamped and critically damped oscillating systems governed by the second order linear constant coefficient differential equation arise in a large number of practical situations including motion of pendulums, radio waves, and water waves.

3.3 Reminder: Particular integrals and complementary functions

Let us suppose that we have found a **particular integral**, $x_P(t)$, i.e., a solution that satisfies

$$\mathcal{L}[x_P(t)] = f(t).$$

Due to the second order nature of the equation, we expect to have two arbitrary constants in the solution. Hence we might look for a second solution $x_Q(t)$ that also satisfies

$$\mathcal{L}[x_Q(t)] = f(t).$$

Then, following the ideas of the solutions of the homogeneous equation, we could construct a general solution of the form

$$x(t) = \alpha x_P(t) + \beta x_Q(t).$$

The only problem is, this is **wrong!**

⁴The exponential always wins over the linear term if you wait a while.

To see why, we substitute to the proposed solution into the left-hand side of the equation:

$$\mathcal{L}[\alpha x_P(t) + \beta x_Q(t)] = \mathcal{L}[\alpha x_P(t)] + \mathcal{L}[\beta x_Q(t)] = \alpha \mathcal{L}[x_P(t)] + \beta \mathcal{L}[x_Q(t)] = (\alpha + \beta)f(t).$$

This clearly does not balance the right-hand side of the equation unless the (assumed) arbitrary $\alpha + \beta$ is always set to be unity (which they can't be because they are arbitrary).

Hence there is **one** and **only one** particular integral of the inhomogeneous equation, called, say $x_P(t)$ which satisfies the full equation

$$\mathcal{L}[x_P(t)] = f(t).$$

We were expecting two arbitrary constants and now we have none! Where do we get them from?

The solution comes from considering the homogeneous equation

$$\mathcal{L}[x(t)] = 0,$$

i.e., the same left-hand side as the full inhomogeneous equation, but with $f(t)$ replace by 0 on the right-hand side. We know that this “reduced” equation will have a general solution involving two arbitrary constants A, B of the form

$$x_{CF}(t) = Ax_1(t) + Bx_2(t),$$

where the $x_j(t)$ each independent satisfy the linear homogeneous equation

$$\mathcal{L}[x_j(t)] = 0, \quad j = 1, 2.$$

We call the general solution to the homogeneous equation $x_{CF}(t)$ the **complementary function** (hence the subscript “CF” attached to it).

Now consider what happens when we add the particular integral $x_P(t)$ and the complementary function $x_{CF}(t)$ and substitute them into the left-hand side of the **full inhomogeneous** equation and use the linearity property of the operator:

$$\mathcal{L}[x_P(t) + x_{CF}(t)] = \mathcal{L}[x_P(t)] + \mathcal{L}[x_{CF}(t)] = f(t) + 0 = f(t).$$

In other words the sum of the particular integral plus the complementary function **also** satisfies the full inhomogeneous equation. Since the complementary function includes two arbitrary constants, we have found the general solution of the full inhomogeneous equation:

$$x(t) = x_P(t) + x_{CF}(t) = x_P(t) + Ax_1(t) + Bx_2(t),$$

where

$$\mathcal{L}[x_P(t)] = f(t), \quad \mathcal{L}[x_j(t)] = 0, \quad j = 1, 2.$$

In general the complementary function is easy to find: you just use the same techniques as we did for solving the homogeneous equations. The difficulty arises in finding the particular integral, which must satisfy the full inhomogeneous equation.

There are several ways of doing this. It is natural to first consider the **method of undetermined coefficients** (aka “educated guesswork”!).

This method considers the form of $f(t)$ and then looks for a particular integral according to the following rule of thumb:

- If $f(t)$ is an n^{th} order polynomial in t , assume a particular integral of the form

$$x_P(t) = A_n t^n + A_{n-1} t^{n-1} + \cdots A_1 t + A_0,$$

i.e., an n^{th} order polynomial with coefficients A_i that must be found by direct substitution into the inhomogeneous equation and balancing powers of t .

- If $f(t) = k e^{\alpha t}$, with k and α given constants, assume a particular integral of the form

$$x_P(t) = A e^{\alpha t}$$

where the exponent α is the same as in $f(t)$ and A is a constant to be found by direct substitution into the inhomogeneous equation and balancing powers of $e^{\alpha t}$.

- If $f(t) = p \sin \beta t + q \cos \beta t$, with p, q and β given constants, assume a particular integral of the form

$$x_P(t) = A \sin \beta t + B \cos \beta t$$

where the frequency β is the same as in $f(t)$ and A, B are constants to be found by direct substitution into the inhomogeneous equation and balancing powers of $\sin \beta t$ and $\cos \beta t$.

- If $f(t)$ is a combination of the above three cases, e.g., an n^{th} order polynomial in t multiplied by an exponential $e^{\alpha t}$, with α a given constant, then assume a particular integral of the form

$$x_P(t) = e^{\alpha t} (A_n t^n + A_{n-1} t^{n-1} + \cdots A_1 t + A_0),$$

and determine the unknown A_i by direct substitution into the full inhomogeneous equation and balancing of terms as above.

- If any component of the proposed particular integrals also satisfies the homogeneous equation, then this will not be a suitable particular integral, as it will only duplicate the complementary function. Hence multiply the proposed particular integral by t , to obtain a new proposed particular integral, repeating this until the guess no longer contains any terms that satisfy the homogeneous equation.

3.4 Resonance

Many oscillatory systems exhibit a phenomenon called **resonance**. This is an large (sometimes infinite) increase in the amplitude of an oscillation when the system undergoes forcing at (or near) its natural frequency.

Suppose we have a system⁵:

$$\ddot{x} + \omega^2 x = \cos \omega_0 t, \quad x(0) = 0, \quad \dot{x}(0) = 0.$$

This represents an oscillatory system with a natural frequency ω being forced at a frequency ω_0 , chosen by us, where for the time being, we assume that $\omega \neq \omega_0$. The equation could model an undamped pendulum with periodic forcing.

We have:

⁵Note that the driving force is time dependent.

- a complementary function

$$x_{CF} = A \sin \omega t + B \cos \omega t$$

- a particular integral

$$x_P = C \sin \omega_0 t + D \cos \omega_0 t$$

so for $\omega \neq \omega_0$ there is no duplication between the complementary function or particular integral.

We have

$$\begin{aligned} x_P(t) &= C \sin \omega_0 t + D \cos \omega_0 t, \\ \dot{x}_P(t) &= C \omega_0 \cos \omega_0 t - D \omega_0 \sin \omega_0 t \\ \ddot{x}_P(t) &= -C \omega_0^2 \sin \omega_0 t - D \omega_0^2 \cos \omega_0 t \end{aligned}$$

Hence substitution into the full inhomogeneous equation gives

$$\begin{aligned} -C \omega_0^2 \sin \omega_0 t - D \omega_0^2 \cos \omega_0 t + \omega^2 (C \sin \omega_0 t + D \cos \omega_0 t) &= \cos \omega_0 t \\ C (\omega^2 - \omega_0^2) \sin \omega_0 t + D (\omega^2 - \omega_0^2) \cos \omega_0 t &= \cos \omega_0 t \end{aligned}$$

and we have

$$C = 0, \quad D = \frac{1}{\omega^2 - \omega_0^2}$$

Thus the particular integral is

$$x_P = \left(\frac{1}{\omega^2 - \omega_0^2} \right) \cos \omega_0 t.$$

It should be immediately obvious that if we tune the forcing frequency ω_0 to the natural frequency ω then the coefficient of the particular integral, and so the amplitude of the motion, diverges. Although catastrophic for the solution, this makes sense on both physical and mathematical grounds.

On **physical** grounds the system is receiving a forcing input at precisely the frequency it would like to oscillate. All the energy of the forcing can be transmitted into increasing the amplitude of the natural oscillation. This is very similar to pushing a child on a swing. Push the child at the wrong frequency and the amplitude will not increase (and you'll get hit in the face!). Push the child at the right frequency (just as the seat reaches its highest point in the swing) and the amplitude of the swing will increase.

On **mathematical** grounds when $\omega_0 = \omega$ the particular integral becomes $\cos \omega t$, which is identical to one of the components of the complementary function. Hence if we had started with $\omega_0 = \omega$, mathematically a particular integral of the form

$$x_P(t) = t (C \sin \omega t + D \cos \omega t),$$

should have been used. A short exercise shows that taking this approach we obtain

$$x_P(t) = \frac{t}{2\omega_0} \sin \omega t.$$

However, you can see that this particular integral will also grow in time due to the presence of the multiplicative factor of t . Hence, the amplitude of the oscillation will still grow. So regardless of whether we examine the limit of $\omega_0 \rightarrow \omega$, or first take $\omega_0 = \omega$ and then consider $t \rightarrow \infty$, the amplitude will diverge. The divergence of an amplitude is often indicative of a breakdown in the assumptions underlying the actual mathematical model.

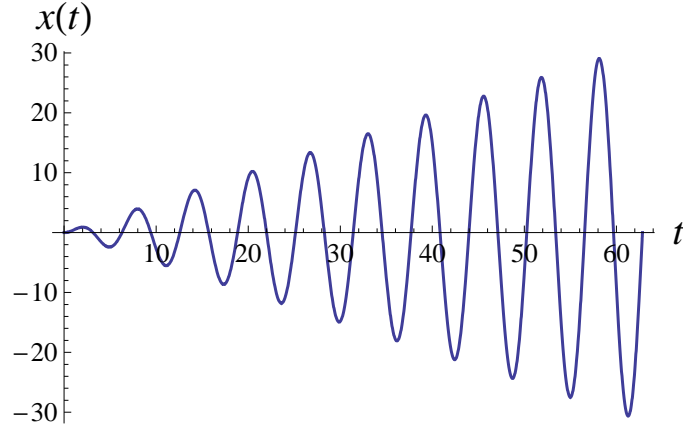


Figure 3.5: The growth of the amplitude of the oscillation when an undamped system is forced at its natural frequency ω_0 (here = 1), resulting in the behaviour $x(t) = t \sin(\omega t)/2\omega_0$.

For example, if x in this equation were modelling the angle of swing of a pendulum, the differential equation would only be valid for small values of $|x| \ll 1$. Hence if the amplitude of the particular integral is increasing unboundedly as $\omega_0 \rightarrow \omega$, then the assumptions that allowed us to linearise the equation from a $\sin x(t)$ gravitational component to a simple $x(t)$ term is false and hence at that frequency it would be more appropriate to consider the full nonlinear equation

$$\ddot{x} - \omega^2 \sin x(t) = \cos \omega t.$$

Another reason for a physical failure of the model might be because we have neglected the effect of damping (see below). Hence a good **mathematical** rule of thumb is:

“Divergence in a linear model for a set of parameters means you should consider a more complete nonlinear model near those parameter values”.

3.4.1 Beating

We continue to analyse the same situation as above, but we use it to discuss the phenomenon of **frequency beating**.

If we now keep $\omega \neq \omega_0$, the full solution is

$$x = A \sin \omega t + B \cos \omega t + \frac{1}{\omega^2 - \omega_0^2} \cos \omega_0 t ,$$

which leads to

$$x(0) = B + \frac{1}{\omega^2 - \omega_0^2} = 0 ,$$

and

$$\dot{x}(0) = A\omega = 0 \Rightarrow A = 0.$$

That is, we have

$$x(t) = x_{CF}(t) + x_P(t) = \left(\frac{1}{\omega^2 - \omega_0^2} \right) (\cos \omega_0 t - \cos \omega t).$$

Subtracting the angle addition formulae

$$\cos(A \pm B) = \cos A \cos B \mp \sin A \sin B,$$

from one another, we have

$$2 \sin A \sin B = \cos(A - B) - \cos(A + B).$$

Hence, setting $A - B = \omega_0$ and $A + B = \omega$, the full solution can be written as

$$x(t) = \left(\frac{2}{\omega^2 - \omega_0^2} \right) \sin \left[\left(\frac{\omega + \omega_0}{2} \right) t \right] \sin \left[\left(\frac{\omega - \omega_0}{2} \right) t \right].$$

The structure of this is extremely informative. It is an amplitude, multiplied by two sines. If ω is approximately (but not exactly) equal to ω_0 , the first sine oscillates with a high frequency, being the mean of ω and ω_0 . The second sine oscillates at a lower frequency being the half the difference of ω and ω_0 . The effect can be seen graphically:

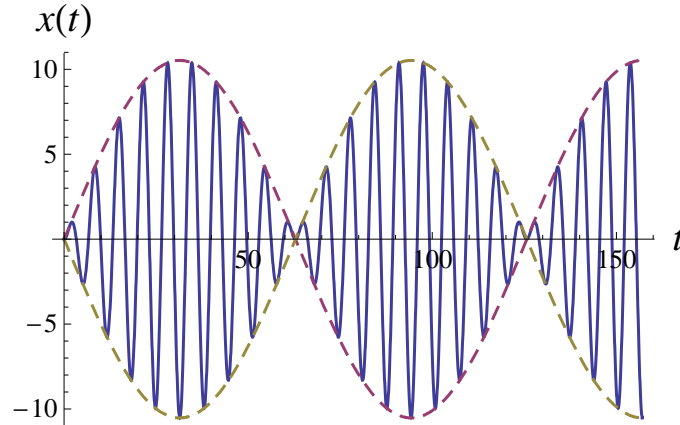


Figure 3.6: The effect of beating. Here the natural frequency is $\omega = 1$ and we have the forcing frequency set at $\omega_0 = 9/10$. The full line is the exact solution in the text. It is clear that high frequency waves, $(\omega + \omega_0)/2 = 19/20$, are modulated in amplitude by low frequency ones, $(\omega - \omega_0)/2 = 1/20$. The dashed lines denote the modulation of the amplitude given from the formula above by $\pm \sin \{(\omega - \omega_0)t/2\}$.

The amplitude higher frequency component is modulated by the lower frequency component. This is a phenomenon known as **beating**. It occurs, for example, when two musical instruments, violins say, that are almost, but not quite in tune. The audience hears a note that is at the mean of the two violins, but the sound varies in loudness at half the difference of the notes (a wowing sound).

3.4.2 Near resonance and damping.

The system we studied above had no damping $\dot{x}(t)$ term. Damping terms have the effect of “smearing out” the resonance. To see this we study the equation

$$\ddot{x} + \lambda \dot{x} + \omega^2 x = \cos \omega_0 t$$

where λ is a positive damping constant such that $\lambda^2 < 4\omega^2$ and the forcing frequency ω_0 is a parameter that can be changes to cycle through the value of the natural frequency of the undamped system ω .

We now have: a complementary function

$$x_{CF} = e^{-\lambda t/2} (A \cos \sigma t + B \sin \sigma t)$$

where

$$\sigma = \sqrt{4\omega^2 - \lambda^2}/2 > 0.$$

Clearly, this is a damped (actually underdamped) oscillatory motion, and so decreases in magnitude with time, whatever the forcing frequency ω_0 or (finite) initial conditions (i.e., the (finite) values of A and B). This type of motion is thus said to be **transient**: eventually the system will increasingly be increasingly less dominated by any component that corresponds to its **natural** frequency and will become increasingly dominated by motion at the **forced** frequency. Thus we won't bother finding A and B and pass on to the particular integral.

Due to the presence of $e^{-\lambda t/2}$ prefactor in the complementary function, the particular integral which takes the form

$$x_P = C \sin \omega_0 t + D \cos \omega_0 t$$

now has no duplication with the complementary function whether $\omega_0 = \omega$ or not. The linear growth is no longer apparent.

We have (yet again)

$$\begin{aligned} x_P(t) &= C \sin \omega_0 t + D \cos \omega_0 t, \\ \dot{x}_P(t) &= C \omega_0 \cos \omega_0 t - D \omega_0 \sin \omega_0 t \\ \ddot{x}_P(t) &= -C \omega_0^2 \sin \omega_0 t - D \omega_0^2 \cos \omega_0 t \end{aligned}$$

Hence substitution into the full inhomogeneous equation gives

$$\begin{aligned} -C \omega_0^2 \sin \omega_0 t - D \omega_0^2 \cos \omega_0 t + \lambda (C \omega_0 \cos \omega_0 t - D \omega_0 \sin \omega_0 t) + \\ + \omega^2 (C \sin \omega_0 t + D \cos \omega_0 t) &= \cos \omega_0 t \\ \{C (\omega^2 - \omega_0^2) - \lambda D \omega_0\} \sin \omega_0 t + \{D (\omega^2 - \omega_0^2) + \lambda C \omega_0\} \cos \omega_0 t &= \cos \omega_0 t \end{aligned}$$

and we have

$$C (\omega^2 - \omega_0^2) - \lambda D \omega_0 = 0, \quad D (\omega^2 - \omega_0^2) + \lambda C \omega_0 = 1$$

which can be solved simultaneously to give

$$C = \frac{\lambda \omega_0}{(\omega^2 - \omega_0^2)^2 + (\lambda \omega_0)^2} \quad D = \frac{(\omega^2 - \omega_0^2)}{(\omega^2 - \omega_0^2)^2 + (\lambda \omega_0)^2}$$

Thus the particular integral is

$$x_P = \frac{\lambda\omega_0}{(\omega^2 - \omega_0^2)^2 + (\lambda\omega_0)^2} \sin \omega_0 t + \frac{(\omega^2 - \omega_0^2)}{(\omega^2 - \omega_0^2)^2 + (\lambda\omega_0)^2} \cos \omega_0 t$$

This can be put into the form of just a single, phase lagged, oscillatory term using the standard addition formulae for sine and cosine:

$$\begin{aligned} C \sin \omega_0 t + D \cos \omega_0 t &= \sqrt{C^2 + D^2} \left(\underbrace{\frac{C}{\sqrt{C^2 + D^2}}}_{=\cos \phi} \sin \omega_0 t + \underbrace{\frac{D}{\sqrt{C^2 + D^2}}}_{=\sin \phi} \cos \omega_0 t \right) \\ \Rightarrow x_P(t) &= \sqrt{C^2 + D^2} \sin(\omega_0 t + \phi), \\ \tan \phi &= \frac{D}{C}. \end{aligned}$$

Using this approach we see that the amplitude of the oscillation is $R(\omega_0) = \sqrt{C^2 + D^2}$. For our C and D above this is

$$R(\omega_0) = \frac{1}{\sqrt{(\omega^2 - \omega_0^2)^2 + (\lambda\omega_0)^2}}, \quad \tan \phi = \frac{(\omega^2 - \omega_0^2)}{\lambda\omega}.$$

If we take the driving amplitude to be 1, the ratio of the response to the forcing is $R(\omega_0)$. A graph of this as a function of ω_0 is plotted below for different values of λ . Note that at resonance $\omega_0 = \omega$, the response climbs rapidly in amplitude, but is not infinite. The size of the response at resonance increases as the damping λ decreases.

Clearly the damping in the system means avoids a divergence in the response. This is a more physically realistic model of forced oscillatory systems.

This is a simple model of the mechanism that is responsible for the serious problems that arose with the Millennium footbridge in London. Soon after it was opened to foot passengers in June 2000, it became apparent that the bridge was undergoing a resonance swaying due to the frequency of footfalls. Extra masses and dampers had to be added to change the position and amplitude of the resonance. A more catastrophic example of resonance is the Tacoma bridge collapse of 1940. There a resonance of the bridge was excited by abnormal wind conditions. This resonance built up and ultimately led to the collapse of the bridge. Resonance is still an extremely important consideration in engineering construction (be it civil or electrical).

3.5 Going further: The Wronskian

So far we have focussed on second order constant equations. These are nice—because we can solve them—but they do not (necessarily) represent reality. In order to go further, we need to develop a bit more computational technology. As this machinery can be applied to more general problems, we take as our starting point an equation of form

$$\frac{d^2 x}{dt^2} + p(t) \frac{dx}{dt} + q(t)x(t) = r(t),$$

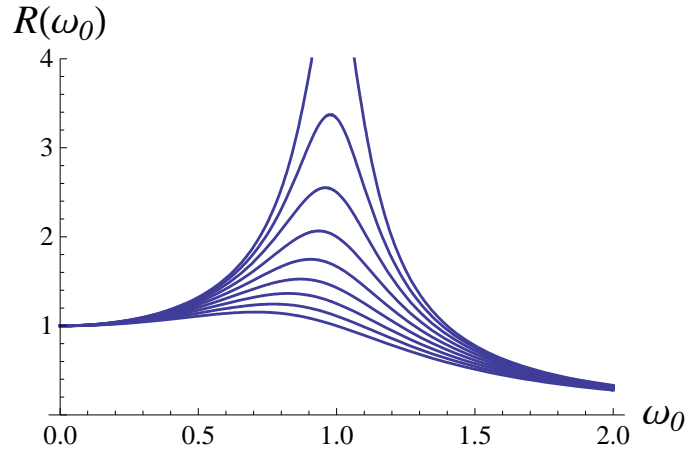


Figure 3.7: The growth of the ratio of the amplitude of response to the amplitude of forcing, $R(\omega_0)$ as the damping λ decreases, plotted as a function of the forcing frequency ω_0 near to the resonance frequency $\omega = 1$. As $\lambda \rightarrow 0$, the width of the resonance peak narrows and the amplitude at resonance $R(\omega_0 = \omega = 1) \rightarrow \infty$

where p, q and r are known functions of t . We also need initial data:

$$x(0) = x_0 \quad \dot{x}(0) = u_0 .$$

We first consider the **homogeneous** case where $r(t) = 0$:

$$\frac{d^2 x}{dt^2} + p(t) \frac{dx}{dt} + q(t)x(t) = 0.$$

and pose two questions we have so far (conveniently) avoided. How many solutions does this equation have? How do we know that these solutions are “independent”?

The first step towards answering these questions involves noting that, since the equation is linear, we can add solutions to get solutions. In essence, a linear combination of solutions is also a solution. This is easy to prove, but it does not take us very far. We may also note that, since we have a second order equation, we need to “integrate” twice which means that we would expect to have two integration constants floating around. This suggests that (perhaps) we should expect that there are two solutions. This turns out to be correct, but how do we prove this?

Suppose we start by assuming that there is a single solution, $x_1(t)$, corresponding to given initial data at (say) $t = t_0$. Setting $x(t) = \alpha x_1(t)$ with α we then require

$$\begin{aligned} \alpha x_1(t_0) &= x_0 &\Rightarrow &\alpha = x_0/x_1(t_0) \\ \alpha \dot{x}_1(t_0) &= u_0 &\Rightarrow &\alpha = u_0/\dot{x}_1(t_0) \end{aligned}$$

and it is easy to see that we cannot (in general) satisfy both equations with a single value of α . Hence, a single solution is not sufficient to provide a unique solution to the second order system of the equation plus two initial conditions.

What happens if we consider two solutions, $x_1(t)$ and $x_2(t)$? Now set $x(t) = \alpha x_1(t) + \beta x_2(t)$ with α and β constants. Then the initial conditions require

$$\begin{aligned}\alpha x_1(t_0) + \beta x_2(t_0) &= x_0 \\ \alpha \dot{x}_1(t_0) + \beta \dot{x}_2(t_0) &= u_0\end{aligned}$$

This can be written as a matrix equation

$$\begin{pmatrix} x_1(t_0) & x_2(t_0) \\ \dot{x}_1(t_0) & \dot{x}_2(t_0) \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \begin{pmatrix} x_0 \\ u_0 \end{pmatrix}$$

We see that, in order that solutions α and β exist, we require the existence of the inverse of the 2×2 matrix containing the initial data, $x_i(t_0)$.

From standard matrix theory (=linear algebra!), we know that the inverse exists if the determinant of the square matrix is non-zero, i.e.,

$$\begin{vmatrix} x_1(t_0) & x_2(t_0) \\ \dot{x}_1(t_0) & \dot{x}_2(t_0) \end{vmatrix} \neq 0.$$

We learn that two solutions will be sufficient, as long as this condition is satisfied.

Suppose that this determinant were actually zero. What would that mean? If we evaluate it explicitly we obtain⁶

$$x_1(t_0)\dot{x}_2(t_0) - \dot{x}_1(t_0)x_2(t_0) = 0 \quad \Rightarrow \quad \frac{x_1(t_0)}{x_2(t_0)} = \frac{\dot{x}_1(t_0)}{\dot{x}_2(t_0)} = c, \text{ say.}$$

Hence if the determinant were to vanish, we would have

$$x_1(t_0) = cx_2(t_0), \quad \dot{x}_1(t_0) = c\dot{x}_2(t_0)$$

which implies that the two solutions $x_1(t)$ and $x_2(t)$ are proportional:

$$x_1(t) = cx_2(t).$$

Suppose instead that there are three fundamental solutions. Set $x(t) = \alpha x_1(t) + \beta x_2(t) + \gamma x_3(t)$ where $x_3(t) \not\equiv 0$ and α, β, γ constants to be found by satisfaction of the initial conditions. Then the initial conditions require the following equations to be satisfied:

$$\begin{aligned}\alpha x_1(t_0) + \beta x_2(t_0) + \gamma x_3(t_0) &= x_0 \\ \alpha \dot{x}_1(t_0) + \beta \dot{x}_2(t_0) + \gamma \dot{x}_3(t_0) &= u_0\end{aligned}$$

This is an **underdetermined** system with 3 unknowns, but only 2 equations. Hence it is impossible to determine unique values of α, β and γ —we cannot have a unique solution. A similar argument applies when we have more than three candidate fundamental solutions.

We have learned something very important. A necessary and sufficient condition to find the complete solution of a second order homogenous linear ordinary differential equation with initial conditions is that we find two solutions $x_1(t), x_2(t)$ such that

$$\begin{vmatrix} x_1(t_0) & x_2(t_0) \\ \dot{x}_1(t_0) & \dot{x}_2(t_0) \end{vmatrix} \neq 0.$$

⁶Assuming here that $x_2(t_0) \neq 0, \dot{x}_2(t_0) \neq 0$, but an analogous satisfactory argument can be found when they are.

Then the general solution of the differential equation is

$$x(t) = \alpha x_1(t) + \beta x_2(t)$$

with α and β determined by the initial conditions.

The arguments leads us to define the **Wronskian**

$$W[x_1(t), x_2(t)] = \begin{vmatrix} x_1(t) & x_2(t) \\ \dot{x}_1(t) & \dot{x}_2(t) \end{vmatrix}$$

Its non-vanishing is intimately related to the concept of **linear independence**.

Example 4: Linearly independents solutions

Consider the equation for a simple harmonic oscillator

$$x'' + x = 0$$

for which we know the solutions

$$x_1(t) = \cos t, \quad x_2(t) = \sin t.$$

Then we have

$$x'_1(t) = -\sin t, \quad x'_2(t) = \cos t,$$

so that

$$W[x_1(t), x_2(t)] = \begin{vmatrix} x_1(t) & x_2(t) \\ x'_1(t) & x'_2(t) \end{vmatrix} = \begin{vmatrix} \cos t & \sin t \\ -\sin t & \cos t \end{vmatrix} = \cos^2 t + \sin^2 t = 1, \forall t.$$

Hence we deduce that $\cos t$ and $\sin t$ are linearly independent, i.e., it is not possible to find non-zero constants c_1 and c_2 such that $c_1 \cos t + c_2 \sin t = 0$ on an interval of t . (Of course we can find these constants so that $c_1 \cos t + c_2 \sin t = 0$ at a point, but linear dependence would require this expression to vanish over a range of values of t .)

3.6 General force terms: Variation of parameters

Finally, let us go a step further in the direction of more realistic situations, e.g. where the involved forces do not lend themselves to “guessing” suitable particular integrals. This illustrates an important method, which “happens” to include our recent friend, the Wronskian. The key point is that it provides a systematic way to obtain a particular integral without any form of guesswork. The method of **variation of parameters** provides just such a method. It can be applied whether the coefficients in the linear equation are constant or functions of the independent variable.

Let us start from a linear second order equation of form:

$$x''(t) + p(t)x'(t) + q(t)x(t) = r(t),$$

and assume that we have found a complementary function

$$x_{CF}(t) = \alpha x_1(t) + \beta x_2(t)$$

which satisfies the homogeneous equation

$$x''(t) + p(t)x'(t) + q(t)x(t) = 0.$$

Suppose we now look for a particular integral of the form

$$x_P(t) = v_1(t)x_1(t) + v_2(t)x_2(t),$$

where the $x_i(t)$ are the complementary functions above and $v_i(t)$ are functions to be determined.

Note that, if $v_i(t)$ were constants then this would have an identical form to the complementary function, which would not be allowed. We therefore assume that $v_i(t)$ are **not** constants and that they will turn out to be such that the complementary function is linearly independent of the particular integral.

Let us further suppose that the following two conditions hold (we will soon see why these are useful):

$$\begin{aligned} v_1'(t)x_1(t) + v_2'(t)x_2(t) &= 0 \\ v_1'(t)x_1'(t) + v_2'(t)x_2'(t) &= r(t) \end{aligned}$$

We can make these two assumptions, since the two $v_i(t)$ are, as yet, undetermined.

To find the $v_i(t)$ we must differentiate $x_P(t)$ and substitute it into the inhomogeneous equation. We then have

$$\begin{aligned} x_P'(t) &= \underbrace{v_1'(t)x_1(t) + v_2'(t)x_2(t)}_{=0 \text{ by assumption}} + v_1(t)x_1'(t) + v_2(t)x_2'(t) = v_1(t)x_1'(t) + v_2(t)x_2'(t) \\ x_P''(t) &= v_1(t)x_1''(t) + v_2(t)x_2''(t) + \underbrace{v_1'(t)x_1'(t) + v_2'(t)x_2'(t)}_{=r(t) \text{ by assumption}} = v_1(t)x_1''(t) + v_2(t)x_2''(t) + r(t) \end{aligned}$$

Substitution thus gives

$$\begin{aligned} \text{LHS} &= x_P'' + p(t)x_P' + q(t)x_P(t) \\ &= \{v_1(t)x_1''(t) + v_2(t)x_2''(t) + r(t)\} + p(t)\{v_1(t)x_1'(t) + v_2(t)x_2'(t)\} + q(t)x_P(t) \\ &= v_1(t)\underbrace{\{x_1''(t) + p(t)x_1'(t) + q(t)x_1(t)\}}_{=0} + v_2(t)\underbrace{\{x_2''(t) + p(t)x_2'(t) + q(t)x_2(t)\}}_{=0} + r(t) \\ &= r(t) = \text{RHS} \end{aligned}$$

So what? Well... we have just shown that

$$x_P(t) = v_1(t)x_1(t) + v_2(t)x_2(t)$$

is the particular integral of

$$x''(t) + p(t)x'(t) + q(t)x(t) = r(t)$$

provided that

$$x_i''(t) + p(t)x_i'(t) + q(t)x_i(t) = 0, \quad i = 1, 2$$

and

$$\begin{aligned}v_1'(t)x_1(t) + v_2'(t)x_2(t) &= 0 \\v_1'(t)x_1'(t) + v_2'(t)x_2'(t) &= r(t).\end{aligned}$$

How does this help to find the $v_i(t)$? We have two simultaneous equations for $v_i'(t)$. These can be rewritten as a matrix equation:

$$\begin{pmatrix} x_1(t) & x_2(t) \\ x_1'(t) & x_2'(t) \end{pmatrix} \begin{pmatrix} v_1'(t) \\ v_2'(t) \end{pmatrix} = \begin{pmatrix} 0 \\ r(t) \end{pmatrix}$$

which can be solved by inverting the coefficient matrix to give

$$\begin{pmatrix} v_1'(t) \\ v_2'(t) \end{pmatrix} = \frac{\begin{pmatrix} x_2'(t) & -x_2(t) \\ -x_1'(t) & x_1(t) \end{pmatrix}}{\begin{vmatrix} x_1(t) & x_2(t) \\ x_1'(t) & x_2'(t) \end{vmatrix}} \begin{pmatrix} 0 \\ r(t) \end{pmatrix} = \frac{r(t)}{W[x_1(t), x_2(t)]} \begin{pmatrix} -x_2(t) \\ x_1(t) \end{pmatrix}$$

where we have recognised $W[x_1(t), x_2(t)]$ as the Wronskian of $x_1(t)$ and $x_2(t)$. Thus we have two equations to solve

$$v_1'(t) = -\frac{x_2(t)r(t)}{W[x_1(t), x_2(t)]}, \quad v_2'(t) = \frac{x_1(t)r(t)}{W[x_1(t), x_2(t)]}$$

These can be integrated immediately with respect to t to give the solutions

$$v_1(t) = -\int^t \frac{x_2(\zeta)r(\zeta)}{W[x_1(\zeta), x_2(\zeta)]} d\zeta \quad v_2(t) = \int^t \frac{x_1(\zeta)r(\zeta)}{W[x_1(\zeta), x_2(\zeta)]} d\zeta$$

At the end of the day, we find that the particular integral is given by

$$\begin{aligned}x_P(t) &= v_1(t)x_1(t) + v_2(t)x_2(t) \\&= -x_1(t) \int^t \frac{x_2(\zeta)r(\zeta)}{W[x_1(\zeta), x_2(\zeta)]} d\zeta + x_2(t) \int^t \frac{x_1(\zeta)r(\zeta)}{W[x_1(\zeta), x_2(\zeta)]} d\zeta \\ \Rightarrow x_P(t) &= \int^t \frac{\{x_2(t)x_1(\zeta) - x_1(t)x_2(\zeta)\} r(\zeta)}{W[x_1(\zeta), x_2(\zeta)]} d\zeta\end{aligned}$$

The final formula can be conveniently remembered as

$$x_P(t) = \int^t \frac{\begin{vmatrix} x_1(\zeta) & x_2(\zeta) \\ x_1(t) & x_2(t) \end{vmatrix}}{W[x_1(\zeta), x_2(\zeta)]} r(\zeta) d\zeta$$

- Note that the lower limit is not specified. This is because any arbitrary change in the lower limit just generates the complementary function, which adds nothing to the particular integral. Hence it is possible to ignore the arbitrary constant of integration.
- Note that the coefficients in the inhomogeneous equation $p(t)$ and $q(t)$ were not assumed to be constants. Hence this formula is valid for particular integrals of more than just inhomogeneous equations of differential equations with constant coefficients.

- An analogous formula can be derived for n^{th} order equations, but we will not consider this here.

Example 5: A simple case

Consider the following linear, inhomogeneous, second order, constant-coefficient equation:

$$\ddot{x} - \dot{x} - 2x = e^{3t}$$

for which it is easy to see that the particular integral has the form Ae^{3t} and we can substitute in to find A . Now let us instead use variation of parameters to systematically find the particular integral **without guessing**.

The complementary function is

$$x(t) = Ae^{-t} + Be^{2t} \quad \Rightarrow \quad x_1(t) = e^{-t}, x_2(t) = e^{2t}$$

The inhomogeneous term on the right-hand side is

$$r(t) = e^{3t}.$$

The Wronskian is

$$W[x_1(\zeta), x_2(\zeta)] = \begin{vmatrix} x_1(\zeta) & x_2(\zeta) \\ \dot{x}_1(\zeta) & \dot{x}_2(\zeta) \end{vmatrix} = \begin{vmatrix} e^{-\zeta} & e^{2\zeta} \\ -e^{-\zeta} & 2e^{2\zeta} \end{vmatrix} = 3e^{\zeta}$$

Thus the particular integral is given by:

$$\begin{aligned} x_P(t) &= \int^t \frac{\begin{vmatrix} e^{-\zeta} & e^{2\zeta} \\ e^{-t} & e^{2t} \end{vmatrix}}{3e^{\zeta}} e^{3\zeta} d\zeta \\ &= \frac{1}{3} \int^t (e^{2t-\zeta} - e^{2\zeta-t}) e^{2\zeta} d\zeta \\ &= \frac{1}{3} \int^t (e^{2t+\zeta} - e^{4\zeta-t}) d\zeta \\ &= \frac{1}{3} e^{2t} \int^t e^{\zeta} d\zeta - \frac{1}{3} e^{-t} \int^t e^{4\zeta} d\zeta \\ &= \frac{1}{3} e^{2t+t} - \frac{1}{12} e^{-t+4t} \\ &= \frac{1}{4} e^{3t}, \end{aligned}$$

where it is worth noting that we ignored the constants of integration in the integral.

Example 6: A more complicated situation

The variation of parameters is a powerful tool for finding the particular integral of inhomogeneous equations where it is not easy to guess its form from the inhomogeneity on the right-hand side. As an example of this, we seek the particular integral of the inhomogeneous equation:

$$\ddot{x} - 2\dot{x} + x = \frac{e^t}{t^2 + 1}.$$

The complementary function is given in straightforward manner by

$$x(t) = (A + Bt)e^t \quad \Rightarrow \quad x_1(t) = e^t, x_2(t) = te^t$$

The inhomogeneous term on the RHS is

$$r(t) = \frac{e^t}{t^2 + 1},$$

which is clearly like nothing we have studied before. Don't panic! Just use the formula from variation of parameters.

The Wronskian is

$$W[x_1(\zeta), x_2(\zeta)] = \begin{vmatrix} x_1(\zeta) & x_2(\zeta) \\ \dot{x}_1(\zeta) & \dot{x}_2(\zeta) \end{vmatrix} = \begin{vmatrix} e^\zeta & \zeta e^\zeta \\ e^\zeta & (1 + \zeta)e^\zeta \end{vmatrix} = e^{2\zeta}$$

Thus the particular integral is given by:

$$\begin{aligned} x_P(t) &= \int^t \frac{\begin{vmatrix} e^\zeta & \zeta e^\zeta \\ e^t & te^t \end{vmatrix}}{e^{2\zeta}} \left(\frac{e^\zeta}{\zeta^2 + 1} \right) d\zeta \\ &= \int^t (te^{t+\zeta} - \zeta e^{\zeta+t}) \left(\frac{e^{-\zeta}}{\zeta^2 + 1} \right) d\zeta \\ &= \int^t \frac{(te^t - \zeta e^t)}{\zeta^2 + 1} d\zeta \\ &= te^t \int^t \frac{d\zeta}{\zeta^2 + 1} d\zeta - e^t \int^t \frac{\zeta}{\zeta^2 + 1} d\zeta \\ &= te^t \arctan t - \frac{1}{2} e^t \log |t^2 + 1| \end{aligned}$$

Which clearly looks nothing like the original right-hand side: without the formula from the variation of parameters we would not have realistically been able to guess the form of the PI!

Finally, the general solution of the above equation is

$$x(t) = x_{CF} + x_P = (A + Bt)e^t + te^t \arctan t - \frac{1}{2} e^t \log |t^2 + 1|.$$

3.7 Eigenvalues and boundary value problems (vibrating strings)

So far we have assumed that the problem we consider involves only initial conditions, which serve to fix the integration constants in the problem and/or the relation between linearly independent solutions at the initial time. There is, however, an important alternative to this, where the solution is instead constrained by set boundary conditions. This changes the nature of the problem—instead of considering the evolution of the system we would typically end up discussing eigenvalues and characteristic solutions which help us understand the behaviour. We will not go very far in this direction, but it is very important that we understand the ideas involved as they lay the foundation for the standard method of solving linear **partial differential equations**.

As a step in this direction—aimed at illustrating the principle and the fact that the specification of boundary data may lead to either no solutions or an infinite number of them—let us consider

$$x''(t) + p(t)x'(t) + \lambda q(t)x(t) = 0, \quad x(a) = 0, \quad x(b) = 0.$$

The key point is that we will treat λ as a parameter, which turns out to determine the existence and type of solutions. Intuitively, thinking of this as an initially unspecified parameter, we can “tune” λ (exactly as in tuning the frequency of a violin string, say) to select the type of solution we want.

Typically a solution satisfying the boundary conditions will exist for only certain values of λ . Such problems are called **eigenvalue problems** and λ is the **eigenvalue**⁷. The solution of the equation+boundary value problem corresponding to a particular value of λ is called the **eigenfunction**.

The relationship between eigenvalues of differential equations and matrices through linear algebra is intimate, but beyond the scope of the present discussion.

3.7.1 Determination of Eigenvalues.

Starting from the equation⁸

$$x''(t) + \lambda x(t) = 0, \quad x(0) = 0, \quad x(L) = 0, \quad L > 0.$$

let us figure out which values of λ lead to non-trivial solutions? We have to consider all (real) possibilities.

- Suppose we have $\lambda = 0$:

⁷*Eigen* is a German word meaning “own” in the sense of “special” or “characteristic” value.

⁸This problem is motivated by the one-dimensional wave equation

$$\frac{\partial^2 Y}{\partial t^2} - \frac{\partial^2 Y}{\partial x^2} = 0$$

which, if we assume that the time dependence is “harmonic” (read: $Y \propto e^{i\omega t}$ for some frequency ω), reduces to

$$\frac{\partial^2 Y}{\partial x^2} + \omega^2 Y = 0$$

If $\lambda = 0$, the equation becomes $x''(t) = 0$, with a general solution

$$x = At + B,$$

Substitution into the endpoint conditions gives $A = B = 0$. Hence the only solution when $\lambda = 0$ is the trivial one $x(t) = 0$. Such static solutions are often rejected in modelling situations as being too boring, or irrelevant. Hence $\lambda = 0$ is rejected as an eigenvalue.

- If, instead, the parameter is negative, $\lambda < 0 = -\alpha^2$:

Now the equation

$$x''(t) - \alpha^2 x(t) = 0$$

has a general solution

$$x(t) = A \cosh \alpha t + B \sinh \alpha t.$$

Substitution into the left endpoint conditions gives:

$$x(0) = A \cosh 0 + B \sinh 0 = A = 0.$$

Substitution into the right endpoint conditions gives:

$$x(L) = A \cosh \alpha L + B \sinh \alpha L = B \sinh \alpha L = 0.$$

Now $\sinh \alpha L$ can only vanish when $\alpha L = 0$. Hence we deduce that $B = 0$. Again the only solution when $\lambda < 0$ is the trivial one $x(t) = 0$. Hence there are no eigenvalues when $\lambda < 0$.

- Finally, the parameter may be positive, $\lambda > 0 = +\alpha^2$:

The equation is now

$$x''(t) + \alpha^2 x(t) = 0$$

with a general solution

$$x(t) = A \cos \alpha t + B \sin \alpha t.$$

Substitution into the left endpoint conditions gives:

$$x(0) = A \cos 0 + B \sin 0 = A = 0.$$

Substitution into the right endpoint conditions gives:

$$x(L) = A \cos \alpha L + B \sin \alpha L = B \sin \alpha L = 0.$$

Clearly $B \sin \alpha L$ can vanish when $B = 0$. This would again lead to a trivial solution. However $B \sin \alpha L$ can also vanish if we tune λ so that

$$\alpha L = n\pi, \quad n \in \mathbb{N} \quad \Rightarrow \quad \alpha = \frac{n\pi}{L} \quad \Rightarrow \quad \lambda \equiv \lambda_n = \left(\frac{n\pi}{L}\right)^2.$$

Hence for a set of discrete eigenvalues λ_n , we can find eigenfunctions

$$x(t) = B \sin \lambda_n t = B_n \sin \frac{n\pi t}{L}$$

that satisfy the equation and the boundary conditions. Note that B is still undetermined. Any possible finite value of B will do, and it could vary between the λ_n . As we shall see in the example below, additional information is often used to determine which values of B_n are appropriate.

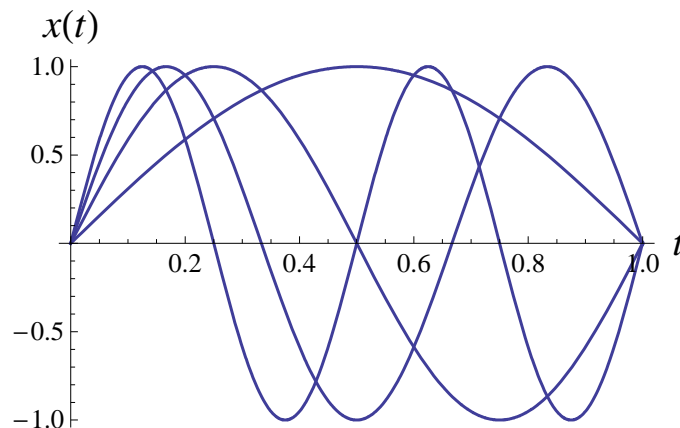


Figure 3.8: The eigenfunctions $\sin n\pi t/L$ for $L = 1$, $n = 1, 2, 3, 4$.

We have effectively just solved the problem of an oscillating violin/guitar string. The eigenvalues we have found are the frequencies you hear when the string instrument is played. The eigenfunctions describe ways in which the strings can vibrate.

3.8 Euler-type equations

An Euler-type equation⁹ is the simplest type of second order differential equations that does **not** have constant coefficients. The class of equations is easy extended to higher orders, but we will only consider the second-order example:

$$a_2 t^2 \frac{d^2 x}{dt^2} + a_1 t \frac{dx}{dt} + a_0 x = 0.$$

(one factor of t per derivative). The problem is easily solved because it can be turned into a second-order linear constant-coefficient equation by the change of variables:

$$z = \ln t \quad \Rightarrow \quad t = e^z.$$

Using the chain rule, we have

$$\frac{dx}{dt} = \frac{dx}{dz} \frac{dz}{dt} = \frac{1}{t} \frac{dx}{dz} = e^{-z} \frac{dx}{dz}$$

$$\frac{d^2 x}{dt^2} = \frac{d}{dt} \left(\frac{dx}{dt} \right) = \frac{d}{dt} \left(e^{-z} \frac{dx}{dz} \right) = e^{-z} \frac{d}{dz} \left(e^{-z} \frac{dx}{dz} \right) = e^{-2z} \frac{d^2 x}{dz^2} - e^{-2z} \frac{dx}{dz}.$$

Substitution of these results into the Euler equation gives

$$\begin{aligned} a_2 e^{2z} \left(e^{-2z} \frac{d^2 x}{dz^2} - e^{-2z} \frac{dx}{dz} \right) + a_1 e^z \left(e^{-z} \frac{dx}{dz} \right) + a_0 x &= 0, \\ a_2 \frac{d^2 x}{dz^2} + (a_1 - a_2) \frac{dx}{dz} + a_0 x &= 0 \end{aligned}$$

⁹Note that Euler was so prolific that there are many different equations that bear the name “Euler equation”, e.g., the main equation in fluid dynamics.

This is (obviously) a homogeneous second-order constant-coefficient equation and so can be solved by the usual method, for x as a function of z . Then x can be written in terms of t by using the substitution $z = \ln t$.

In practice, note that (at least for distinct roots) the solutions of the transformed equation will take the form $e^{k_i z}$. Hence after transforming back to the original variables x , then the solutions take the form t^{k_i} . Hence it is possible to attack Euler equations by a direct substitution of solutions of the form t^k and solving the indicial equation (=auxiliary equations) for k .

Example 7: Stellar oscillations

A typical situation where Euler equations arise is in *astero-seismology* (the oscillations of stars)—perhaps ironically, involving the Euler equations from fluid dynamics. The relevant equation relates to the dependence on the radial coordinate, which we will call t in order to stay close to the previous description. One arrives at an ordinary differential equation by the method of separation of variables (which you will find out a lot more about next year). This equation can be written

$$t^2 x'' + 2tx' - l(l+1)x = 0$$

with $l \geq 0$ and integer. We want the solution to be i) regular at the centre, $t = 0$, and ii) match smoothly to a solution in the star's exterior. We will ignore the second of these conditions for now.

Trying a power-law solution, $x = t^\alpha$, we get

$$\alpha(\alpha-1)t^2 \times t^{\alpha-2} + 2\alpha t \times t^{\alpha-1} - l(l+1)t^\alpha = 0$$

or, dividing through by the common factor t^α ;

$$\alpha(\alpha-1) + 2\alpha - l(l+1) = 0$$

or

$$\alpha^2 + \alpha - l(l+1) = 0 \implies \alpha = -\frac{1}{2} \pm \sqrt{\frac{1}{4} + l(l+1)} = -\frac{1}{2} \pm \left(l + \frac{1}{2}\right)$$

We have two options. Either $\alpha = l$, which means that the solution behaves like t^l . That is, it satisfies the boundary condition at $t = 0$. The other option is $\alpha = -l - 1$ and a solution that behaves as t^{-l-1} , which diverges as $t \rightarrow 0$ and therefore must be discarded.

3.8.1 Repeated roots of the indicial equation

Suppose we want to solve

$$t^2 \frac{d^2 x}{dt^2} + 3t \frac{dx}{dt} + x = 0.$$

Try a solution of the form $x(t) = t^k$, as before. Then

$$x'(t) = kt^{k-1}, \quad x''(t) = k(k-1)t^{k-2}.$$

Substitution into the equation gives

$$t^2 \times \underbrace{k(k-1)t^{k-2}}_{x''(t)} + 3t \times \underbrace{kt^{k-1}}_{x'(t)} + \underbrace{t^k}_{x(t)} = 0 \quad \Rightarrow \quad \{k(k-1) + 3k + 1\}t^k = 0$$

This equation must be valid for all t , hence we obtain the **indicial** equation as

$$k(k-1) + 3k + 1 = 0 \quad \Rightarrow \quad k^2 + 2k + 1 = 0 \quad \Rightarrow \quad k = -1, \text{ twice.}$$

We obtain but one solution as $x = 1/t$, but what about the other one? The hint comes from what happens when you have equal roots in a second order linear constant coefficient equations. In that case the solution takes the form $(A + Bz)e^{k_1 z}$. We could have solved the Euler equation by a transformation $z = \ln t$ as a second order linear constant coefficient equation, in which case $k_1 = -1$. Hence the general solution takes the form:

$$x(t) = \frac{A}{t} + \frac{B}{t} \ln t.$$

In general when there are equal roots k_1 for the indicial equation of an Euler equation, the general solution is then

$$x(t) = (A + B \ln t)t^{k_1}$$

3.8.2 Complex roots of the indicial equation

It is also possible to obtain complex values for the roots of the indicial equation, say $k = \alpha \pm \beta i$. In that case, using the rules of logarithms, the general solution can be written as

$$x(t) = At^{\alpha+\beta i} + Bt^{\alpha-\beta i} = t^\alpha (Ae^{\beta i \ln t} + Be^{-\beta i \ln t}) = t^\alpha (C \cos(\beta \ln t) + D \sin(\beta \ln t)),$$

where A, B , or alternatively, C, D are arbitrary constants determined by the two initial/boundary conditions.

Chapter 4

Kinematics and relative motion

In general, kinematics is concerned with the description of moving bodies, without considering how the motion is produced. This typically involves the motion in terms of

- the position

$$\vec{r}(t),$$

- the velocity

$$\dot{\vec{r}}(t) \equiv \frac{d\vec{r}(t)}{dt} = \vec{v}(t),$$

- the acceleration

$$\ddot{\vec{r}}(t) \equiv \frac{d^2\vec{r}(t)}{dt^2} = \frac{d\vec{v}(t)}{dt} = \vec{a}(t).$$

For any given force, the motion is then determined from Newton's second law (N2). We have seen aspects of this already. Now we are going to take a closer look at the role of the observer. In particular, we are going to touch upon aspects relating to Newton's first law, which brings in the concept of an *inertial frame*. We know that, in a coordinate system in which there is no force acting on a body, it does not accelerate (relative to these coordinates). Let us now add the desire that different observers should infer the same physics. In order for that to be the case, the velocity of the observer (or the experimental device) must not impact on the dynamics. This gives us a hint of what we are after. In order to make the idea precise, it helps if we understand how measurements are affected by relative motion.

4.1 Frames of Reference

An event takes place at a point in space at a given time. For example, Man landed on the Moon on July 16th 1969 at the Sea of Tranquility; JFK was assassinated on 26th November 1963 in Dallas, Texas.

We can all agree that these specific events happened¹, but we may disagree on the exact coordinate location or time. We may not have synchronised our clocks and we may be using different rulers.

Events can be characterised both in time and space and for this we need a **reference frame**, which is a set of spatial coordinates plus a time coordinate. In general the motion of a body is also described by reference to a set of axes, for example the Cartesian coordinate system with a specific origin. Sometimes a set of axes, called a **frame** may itself be moving² and different observers may choose different points for their origin.

For example, someone sitting in a train may naturally choose to centre his (Cartesian) coordinate description of the world on his seat. Someone waiting on the platform as the train passes through the station would more likely centre her (Cartesian) coordinate description of the world on her position on the platform. If the passenger on the train throws an apple into the air as the train passes through the station, he will describe its motion according to his set of coordinates which travel with the train. The trajectory of the apple will look quite different to the person waiting on the platform as will her description of the trajectory in terms of the platform coordinates. How can we reconcile the two descriptions?

To begin to do this we need to define the concept of an inertial frame of reference.

An **inertial frame of reference** is a set of coordinate axes in which an isolated body under no forces (i.e., external influence) travels with constant velocity.

Experiment demonstrates that a frame fixed relative to the sun is inertial for all macroscopic practical purposes. However, a frame which rotates relative to another frame is **not** inertial because any motion in a curve must produce an acceleration and hence involves a non-constant velocity. Thus technically speaking a frame fixed in a laboratory on the surface of the rotating earth is **not** inertial, since the laboratory itself is rotating. However for most practical purposes the earth is spinning so slowly that the rotational complications are unimportant (the anticlockwise spinning of the water down the plug-hole in the northern hemisphere is a myth!). However for situations involving large distances or timescales, such as meteorology or intercontinental ballistic missiles, the effects of rotation do introduce additional external influences that must be taken into account.

Basically, in what follows we are always working with approximations.

4.2 Galilean Transformations

Let \mathcal{O} be the origin of a coordinate frame \mathcal{S} and \mathcal{O}' be the origin of a coordinate frame \mathcal{S}' . Suppose \mathcal{S} and \mathcal{S}' have the same orientation³, but that the position vector of \mathcal{O}' relative to \mathcal{O} be $\vec{R}(t)$.

We shall also assume that there is a universal time. In other words, in the example above say, both the commuter \mathcal{O} and the passenger on the train \mathcal{O}'

¹Assuming we are not conspiracy theorists.

²This will be important later!

³As we are dealing with non-accelerated motion, we can always do this.

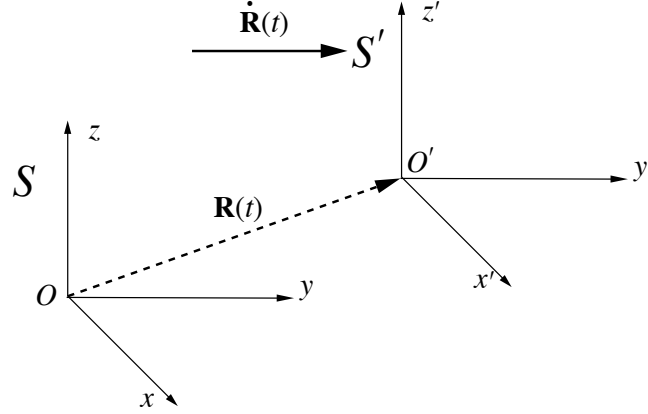


Figure 4.1: Frame \mathcal{S}' moves with velocity $\dot{\vec{R}}(t)$ relative to \mathcal{S} . Note that, since \mathcal{S}' moves with constant velocity we can always align the x -axis with that of the fixed coordinate system.

agree on the precise time at which the apple was thrown into the air and its time of flight.⁴

- Define $\vec{r}(t)$ to be the position vector of a point P in \mathcal{S} relative to \mathcal{O} .
- Define $\vec{r}'(t)$ to be the position vector of a point P in \mathcal{S}' relative to \mathcal{O}' .

Then it follows at time t from the law of the addition of vectors the relative positions of the point as seen by the differing observers at \mathcal{O} and \mathcal{O}' is instantaneously

$$\vec{r}(t) = \vec{R}(t) + \vec{r}'(t).$$

If we differentiate this with respect to time we have

$$\frac{d\vec{r}(t)}{dt} = \frac{d\vec{R}(t)}{dt} + \frac{d\vec{r}'(t)}{dt},$$

or replacing derivatives by \vec{v} for velocities

$$\vec{v}_{P \text{ relative to } \mathcal{S}} = \vec{v}_{P \text{ relative to } \mathcal{S}'} + \vec{v}_{\mathcal{S}' \text{ relative to } \mathcal{S}}.$$

The last equation can be abbreviated to

$$\vec{v}_{PS} = \vec{v}_{PS'} + \vec{v}_{S'S}.$$

or (in slightly more natural notation)

$$\vec{v} = \vec{V} + \vec{v}'$$

⁴That a universal time exists irrespective of motion seems perfectly natural to us. However, it is also wrong. In the theory of relativity there is no universal measurement of time that is the same between inertial frames. If the speed of the train were near to the speed of light, the commuter and passenger would disagree as to the exact time the apple was thrown.

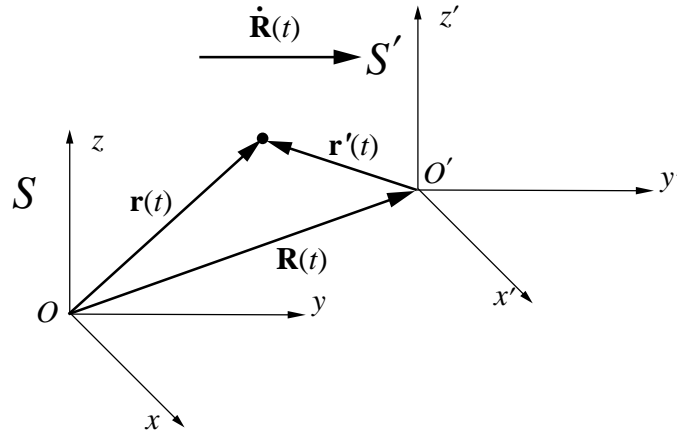


Figure 4.2: Frame S' moves with velocity $\dot{\vec{R}}(t)$ relative to S .

This is the equation that links the relative velocities of the point as observed in each frame, when the two frames are moving with respect to each other. We see that velocities add, just as we expect from our everyday intuition.

Relative velocities are confusing. However the confusion can often be minimised by drawing a diagram⁵. Also, the following alternative notation might make things clearer. The crucial equation we use is of the simple form

$$\vec{v}_{AC} = \vec{v}_{AB} + \vec{v}_{BC}$$

In words, the velocity of A relative to C is equal to the sum of the velocity of A relative to B , plus the velocity of B relative to C . This follows from the equation relating displacement vectors:

$$\vec{AC} = \vec{AB} + \vec{BC}$$

or, in other notation,

$$\vec{r}_{AC} = \vec{r}_{AB} + \vec{r}_{BC}$$

The equation for relative velocities is simply the time derivative of these equations.

If $\vec{v}_{S'S}$ is a constant $= \vec{V}$, say then S' is also an inertial frame. The coordinates in each frame are then related by the following set of transformations, known as **Galilean transformations**:

$$\vec{R}(t) = \vec{V}t \Rightarrow \vec{r}(t) = \vec{V}t + \vec{r}'(t).$$

where we have assumed $\vec{r}(t) = \vec{r}'(t)$ at $t = 0$ (the two coordinate systems have a common origin at the initial time).

⁵In fact, you should always do this!

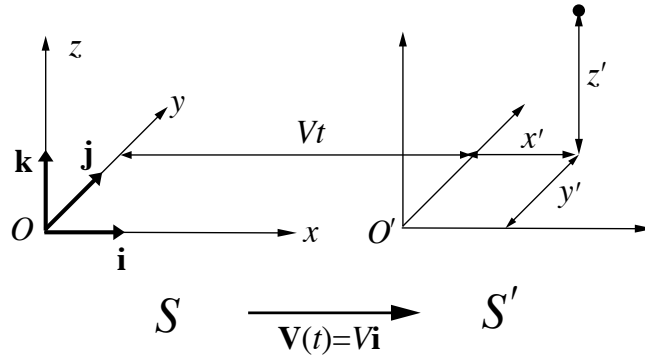


Figure 4.3: Frame \mathcal{S}' moves with velocity $V\hat{i}$ relative to \mathcal{S} .

Example 8: Galilean transformation

Suppose \mathcal{S} and \mathcal{S}' have a uniform speed $\vec{V} = V\hat{i}$ relative to one another, with the respective axes aligned (see figure 4.3). The Galilean transformation between coordinates (x, y, z) relative to \mathcal{O} in \mathcal{S} and the coordinates (x', y', z') relative to \mathcal{O}' in \mathcal{S}' is then given by the set of vector equation

$$x\hat{i} + y\hat{j} + z\hat{k} = Vt\hat{i} + x'\hat{i} + y'\hat{j} + z'\hat{k}$$

or in component form

$$\begin{aligned} x &= Vt + x' \\ y &= y' \\ z &= z' \end{aligned}$$

Note that because the axes were aligned an identical set of basis vectors \hat{i} , \hat{j} and \hat{k} could be used in both frames.

Example 9: Relative motion in a straight line

A train moves at 30 ms^{-1} (about 70 mph). A boy runs to the front of the train at 2.5 ms^{-1} (about 5 mph). What is the velocity of the boy relative to the track? Let

- \mathcal{S} be a coordinate frame that is stationary with respect to the track,
- \mathcal{S}' be a coordinate frame that is stationary with respect to the train.

with the corresponding x and x' axes aligned in the same direction along the track/train and pointing towards the front of the train. Then let

- \vec{v}_{PS} = velocity of the boy P relative to a stationary observer on the track.

- $\vec{v}_{PS'}$ = velocity of the boy P relative to a stationary observer on the train = $2.5 \hat{i} \text{ ms}^{-1}$.
- $\vec{v}_{S'S}$ = velocity of the stationary observer on the train relative to the stationary observer on the track = $30 \hat{i} \text{ ms}^{-1}$.

From the relative velocity formula

$$\vec{v}_{PS} = \vec{v}_{PS'} + \vec{v}_{S'S},$$

we deduce that

$$\vec{v}_{PS} = 30 \hat{i} + 2.5 \hat{i} = 32.5 \hat{i} \text{ ms}^{-1}.$$

If the boy turns round and runs at the same speed towards the back of the train, his new velocity relative to the train is $\vec{v}_{PS'} = -2.5 \hat{i} \text{ ms}^{-1}$. Consequently his velocity relative to the stationary track observer is then

$$\vec{v}_{PS} = 30 \hat{i} - 2.5 \hat{i} = 27.5 \hat{i} \text{ ms}^{-1}.$$

Example 10: Relative motion of particles

Two particles have position vectors given by

$$\vec{r}_1(t) = 4t\hat{i} + 5t^2\hat{j} - 2\sin\pi t\hat{k}$$

and

$$\vec{r}_2(t) = 2t\hat{i} - 5t\hat{j} + \cos\pi t\hat{k}.$$

Find the relative velocity and the relative acceleration of the second particle with respect to the first at $t = 1$.

First define the relevant reference frames:

- Frame S : the stationary observer.
- Frame P_1 : the frame of particle 1.
- Frame P_2 : the frame of particle 2.

We require $\vec{v}_{P_2P_1}$, which is given by

$$\vec{v}_{P_2P_1} = \vec{v}_{P_2S} + \vec{v}_{SP_1} = \vec{v}_{P_2S} - \vec{v}_{P_1S}$$

The velocities of the particles are:

$$\vec{v}_{P_1S} = \frac{d}{dt}(\vec{r}_{P_1S}) = \frac{d}{dt}(\vec{r}_1) = \dot{\vec{r}}_1(t) = 4\hat{i} + 10t\hat{j} - 2\pi\cos\pi t\hat{k}$$

$$\vec{v}_{P_2S} = \frac{d}{dt}(\vec{r}_{P_2S}) = \frac{d}{dt}(\vec{r}_2) = \dot{\vec{r}}_2(t) = 2\hat{i} - 5\hat{j} - \pi\sin\pi t\hat{k}$$

The velocity of 2 relative to 1 is:

$$\vec{v}_{P_2P_1} = \dot{\vec{r}}_2(t) - \dot{\vec{r}}_1(t) = -2\hat{i} - (5 + 10t)\vec{j} + (2\pi \cos \pi t - \pi \sin \pi t) \hat{k}.$$

So at $t = 1$ we have

$$\vec{v}_{P_2P_1}(t = 1) = \dot{\vec{r}}_2(1) - \dot{\vec{r}}_1(1) = -2\hat{i} - 15\vec{j} - 2\pi\hat{k}.$$

By a similar logic the acceleration of 2 relative to 1 is

$$\vec{a}_{P_2P_1} = \ddot{\vec{r}}_2(t) - \ddot{\vec{r}}_1(t) = 0\hat{i} - 10\vec{j} + (-2\pi^2 \sin \pi t - \pi^2 \cos \pi t) \hat{k}.$$

so at $t = 1$ we have

$$\vec{a}_{P_2P_1}(t = 1) = \ddot{\vec{r}}_2(1) - \ddot{\vec{r}}_1(1) = -10\hat{j} + \pi^2\hat{k}.$$

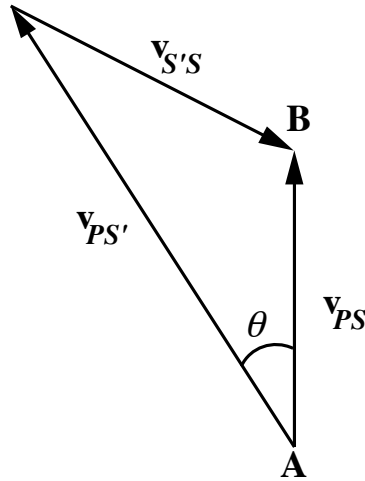


Figure 4.4: Wind velocities

Example 11: Airspeed versus ground speed

An aeroplane P is to fly from a point A to an airfield B which is 300 km due north of A . The plane can fly at a speed of 150 km h^{-1} relative to the air. A wind of speed 30 km h^{-1} blows from the north-west. Find the direction the plane should be pointing in if it is to reach B and its flight time.

Let

- \mathcal{S} be a coordinate frame that is stationary with respect to the ground (A and B),
- \mathcal{S}' be a coordinate frame that is stationary with respect to the wind.
- us take basis vectors \hat{i} pointing east and \hat{j} pointing north.

Then we have

- \vec{v}_{PS} = velocity of the plane P relative to a stationary observer on the ground.
- $\vec{v}_{PS'}$ = velocity of the plane P relative to the wind.
- $\vec{v}_{S'S}$ = velocity of the wind relative to the stationary observer on the ground.

Hence we know that:

- $\vec{v}_{S'S} = 30 \times \frac{1}{\sqrt{2}}(\hat{i} - \hat{j}) \text{ km h}^{-1}$;
- $|\vec{v}_{PS'}| = 150 \text{ km h}^{-1}$;
- if the plane is to fly from A to B , $\vec{v}_{PS} = V\hat{j}$ for some V we must find.

The best way to solve this is to draw a vector diagram corresponding to the relative velocity equation (see figure 4.4)

$$\vec{v}_{PS} = \vec{v}_{PS'} + \vec{v}_{S'S}.$$

There are two unknowns in the problem, the bearing the plane must take and the speed of the plane relative to the ground. A short study of the vector triangle in the diagram shows the following.

- The unknown angle θ is the bearing (west of north) that the plane must fly, so that its velocity relative to the ground is due north. It is given by the sine rule:

$$\frac{150}{\sin 135} = \frac{30}{\sin \theta} \quad \Rightarrow \quad \sin \theta = \frac{1}{5\sqrt{2}} \quad \Rightarrow \quad \theta = \arcsin\left(\frac{1}{5\sqrt{2}}\right)$$

or about $8^\circ 8'$ west of north. ($1' = 1 \text{ minute} = 1/60$ -th of a degree).

- The speed of the aircraft relative to the ground is also given by the sine rule

$$\frac{|\vec{v}_{PS}|}{\sin(180 - 135 - \theta)} = \frac{150}{\sin 135}$$

Hence

$$|\vec{v}_{PS}| = 127.3 \text{ km h}^{-1}$$

and the flight time from A to B is $300/127.28 = 2 \text{ hrs } 21 \text{ min.}$

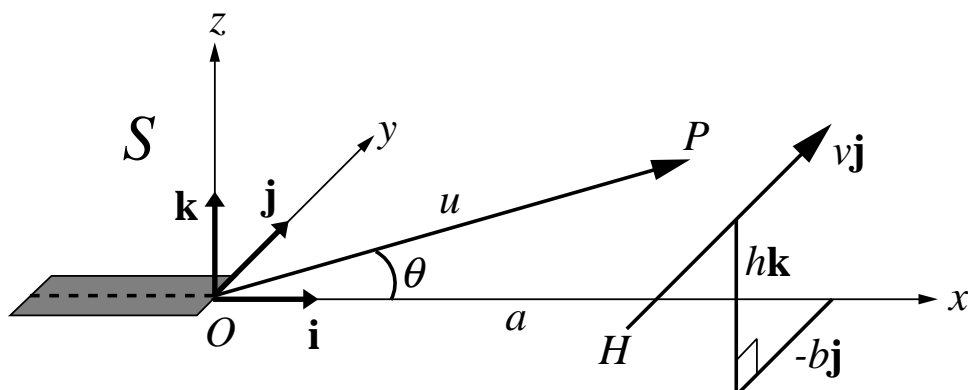


Figure 4.5: Geometry of flight paths of plane P and helicopter H .

Example 12: Closest approach

An aircraft P takes off from a runway that is aligned from west to east parallel to \hat{i} . The speed of the aircraft (relative to the stationary air) is u and after takeoff it flies due east, climbing at angle θ to the horizontal. Meanwhile a helicopter H is flying due north at speed $v\hat{j}$ at a constant height $h\hat{k}$ (with $h > 0$ and \hat{k} pointing upwards!).

At the moment the aircraft takes off at $t = 0$, the helicopter is at the point $a\hat{i} - b\hat{j} + h\hat{k}$ ($a, b > 0$). Find the time at which the two machines are closest and distance between them at that point.

Clearly it makes sense to draw a diagram (figure 4.5).

Start by defining frames. Let:

- \mathcal{S} be a coordinate frame that is stationary with respect to the end of the runway,
- \mathcal{S}' be a coordinate frame that is stationary with respect to the helicopter.

Recall the relationship between position vectors

$$\vec{r}(t) = \vec{R}(t) + \vec{r}'(t).$$

It would be useful to calculate $\vec{r}'(t)$. Note that

- the position of the plane relative to the end of the runway at $t = 0$ is $\vec{r}(0) = \vec{0}$.
- the position of the helicopter relative to the end of the runway at $t = 0$ is

$$\vec{R}(0) = a\hat{i} - b\hat{j} + h\hat{k}.$$

- the position of the plane at $t = 0$ relative to the helicopter at $t = 0$ is

$$\vec{r}'(0) = \vec{r}(0) - \vec{R}(0) = -a\hat{i} + b\hat{j} - h\hat{k}.$$

The corresponding relative velocities are then given by

$$\vec{v}_{PS} = \vec{v}_{PS'} + \vec{v}_{S'S}.$$

and

- the velocity of the plane relative to the end of the runway is $\vec{v}_{PS} = u \cos \theta \hat{i} + u \sin \theta \hat{k}$.
- the velocity of the helicopter relative to the end of the runway is $\vec{v}_{S'S} = v \hat{j}$.
- the velocity of the plane relative to the helicopter is

$$\vec{v}_{PS'} = \vec{v}_{PS} - \vec{v}_{S'S} = u \cos \theta \hat{i} - v \hat{j} + u \sin \theta \hat{k}.$$

Hence we can write down a vector equation for the position of the plane relative to the helicopter. It is

$$\vec{r}'(t) = \vec{r}'(0) + \vec{v}_{PS'} t = -a \hat{i} + b \hat{j} - h \hat{k} + t (u \cos \theta \hat{i} - v \hat{j} + u \sin \theta \hat{k})$$

which is nothing more than the equation of a line in 3D.

Now the distance of closest approach will occur when $|\vec{r}'(t)|$ or analogously $|\vec{r}'(t)|^2$ is a minimum.

This occurs when

$$\frac{d}{dt} |\vec{r}'(t)|^2 = \frac{d}{dt} (\vec{r}'(t) \cdot \vec{r}'(t)) = \frac{d\vec{r}'(t)}{dt} \cdot \vec{r}'(t) + \vec{r}'(t) \cdot \frac{d\vec{r}'(t)}{dt} = 2\vec{r}'(t) \cdot \vec{v}_{PS'} = 0$$

Geometrically this is when $\vec{v}_{PS'}$ and $\vec{r}'(t)$ are perpendicular.

Hence we have

$$\begin{aligned} 0 &= (u \cos \theta \hat{i} - v \hat{j} + u \sin \theta \hat{k}) \cdot (-a \hat{i} + b \hat{j} - h \hat{k} + t (u \cos \theta \hat{i} - v \hat{j} + u \sin \theta \hat{k})) \\ \Rightarrow t_{\min} &= -\frac{(u \cos \theta \hat{i} - v \hat{j} + u \sin \theta \hat{k}) \cdot (-a \hat{i} + b \hat{j} - h \hat{k})}{|u \cos \theta \hat{i} - v \hat{j} + u \sin \theta \hat{k}|^2} \\ &= \frac{au \cos \theta + bv + hu \sin \theta}{u^2 + v^2}. \end{aligned}$$

The closest distance between the two machines is then

$$|\vec{r}'(t_{\min})| = \sqrt{\vec{r}'(t_{\min}) \cdot \vec{r}'(t_{\min})} = \text{exercise for student!}$$

Can this ever be zero? In practice the relative speeds of aircraft are so fast (up to over 1000 mph $\approx 450 \text{ ms}^{-1}$ for civil aircraft) that air traffic controllers try to allow aircraft never to come within a certain non-zero distance of each other. Usually over the Atlantic routes into the UK this is 1000 ft for vertical separation and a few miles for horizontal separation—yes, the Imperial-measures dominate air traffic control!

Chapter 5

Special relativity

Newtonian mechanics ruled supreme for centuries, describing motion from every day events to the planets orbiting the Sun. However, as experimenters started probing higher velocities (or, equivalently, energies) it became clear that the theory had problems. In particular, at the end of the 19th century there was an apparent conflict between Newtonian physics and the theory of electromagnetism developed by James Clerk Maxwell and Michael Faraday. This new theory suggested that the speed of light (in vacuum) ought to be constant, regardless of the motion of the observer. This (obviously) does not agree with the Galilean transformations we have considered.

The theory of electromagnetism demonstrates that light is a wave. In order to make sense of this, physicists introduced a medium for these waves to propagate through¹: the luminiferous ether. However, experiments aimed at demonstrating the absolute motion of the Earth through the ether failed (most famously, the effort by Michelson and Morley in 1881). As a resolution, physicists introduced two, seemingly ad hoc, hypotheses:

- rigid bodies are contracted in the direction of travel, and
- moving clocks appear to run slow.

Mathematically, the results were contained in (fairly) simple relations called the *Lorentz transformation*. The effects would impact on any effort to detect the relative motion with respect to the ether, basically cancelling the effects one might intuitively expect.

This description was logically consistent, but there was no way of verifying the assumptions. At least not until 1905, when Albert Einstein provided an elegant derivation based on two postulates. The first is known as *the principle of relativity*. The second concerns the speed of light, which is taken to be a universal constant. This assumption has a range of curious implications – it forces us to abandon the concept of universal time (which is key to Newtonian mechanics). We also lose the notion of simultaneity. Events may appear to happen at the same time according to one observer, but not according to another. This may seem confusing, but the results are quite easy to demonstrate. We

¹The logic is simple: In order for there to be a wave there must be something bobbing up and down (or sideways, for that matter).

will do this using the *k-calculus* that was developed by Hermann Bondi in the 1950s.

5.1 k-calculus

Let us start with some familiar. An **event** in space and time simply refers to something happening. Everyone can agree on what the event is (what happened), but since different observers may use their own clocks (which may not be synchronised) and measuring devices (perhaps rulers with different scales or simply a different assumed origin), they will not (typically) agree on when and where the event took place. There is nothing mysterious about this.

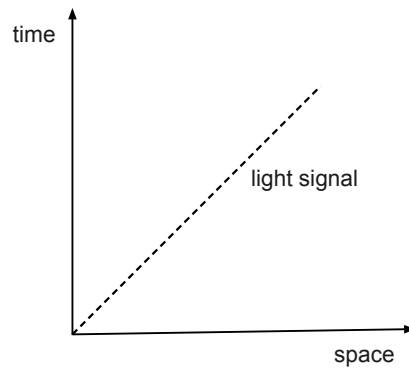


Figure 5.1: Spacetime diagram in units where the speed of light $c = 1$.

Mathematically, we can illustrate the collection of events throughout history in a spacetime diagram, like that in figure 5.1. This would (obviously) involve recording how objects move – how they trace out their individual **world lines** through space and time. By following each world line we track the object’s history. Now, let us introduce the notion of an **observer**. This can be any object that can be equipped with a clock and a ruler, which he/she carries along on the journey through spacetime. Each individual combination of clock and ruler makes up the observer’s **frame of reference**. In Newtonian mechanics, we assume that if two observers synchronise their clock then they will continue to agree on the time of all events. Time is absolute.

In order to develop Einstein’s picture, let us start with the principle of relativity:

- All inertial observers are equivalent.

What does this mean? Well, it says that, if one observer carries out an experiment and discovers a physical law, then other observers carrying out the same experiment should agree on this law. At least as long as the observers have uniform relative velocity – they must not accelerate because this brings additional forces into play. However, there is more to it than that. In reality we can not determine an absolute position in space or time, we can only ever measure positions relative to some origin. Similarly, a velocity of a body is always relative

to some other. Position and velocity are both **relative**. This is important. Any experiment our observer may carry out must involve some kind of observation. Therefore the statement of the principle of relative does not refer to the results of particular experiments being equal. This follows by logic.

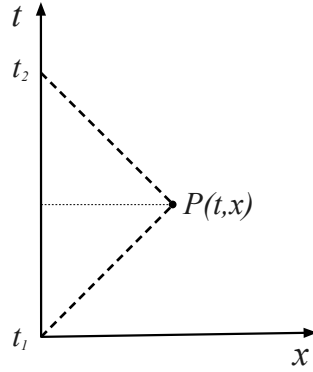


Figure 5.2: The “radar” method for measuring points in spacetime.

We have suggested that we can use a set of clock and rulers to measure events. Let us now dispense of the rulers and make use of the second postulate, the constant speed of light². This leads us to (using terminology that may have seemed modern in the 1950s, but which seems unfashionable today) the so-called radar method. If we know the speed of light, then we can easily figure out the position in space of any event (let’s call it P). Simply assuming that the event is associated with a “mirror” we can bounce a light signal off it. In order to avoid confusion, we will work in two dimensions, time and space. This means that we can think of the observers as sitting along a straight line, each with a clock and a mirror (or a flashlight). A given observer (let’s call him/her A , which seems quite imaginative) then carries out measurements like in figure 5.2, where the event $P(t, x)$ could be on another observer’s (B) world line. As we are focussing on inertial observers, this second world line will be a straight line at some angle in the coordinate system associated with A . The angle simply depends on the relative velocity, see figures 5.3 and 5.4. As the relative speed approaches that of light, the angle approaches 45° .

In this picture, the spatial distance to the event P in figure 5.2 is inferred as half the difference between the times of emission and reception. In order to make the analysis simple, it is common to use units such that the speed of light $c = 1$. That is, we measure distance as time intervals (e.g. in light seconds) and we have

$$x = \frac{1}{2}(t_2 - t_1)$$

Similarly, it is easy to see that the time associated with the event will be

$$t = \frac{1}{2}(t_2 + t_1)$$

²Note that we assume there is not matter between events and observers in this discussion.

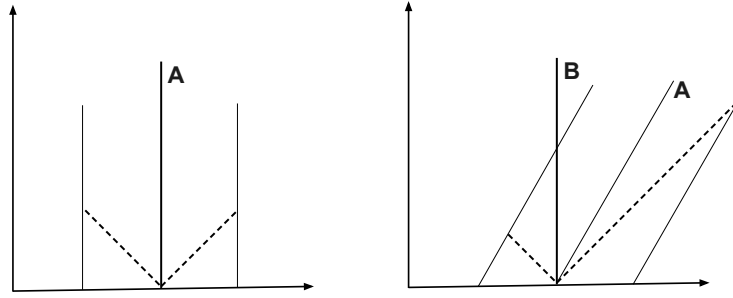


Figure 5.3: Illustrating the lack of simultaneity in relativity.

An immediate implication of this is that we have to abandon the notion of simultaneity. This is easily illustrated with a thought experiment (the kind of which Einstein is famous for). Let us consider a passenger on a train, equipped with two flashlights (and no hesitation using them). Let the person stand in the middle of the carriage and shine the two flashlights towards the front and back of the carriage. The ask, when do the light rays hit the two walls. Our observer would answer that this happens at the same time – the light rays have to travel the same distance at the same (fixed) speed. The events are *simultaneous*. This is the situation illustrated in the left panel of figure 5.3. Let us now consider the experiment from the point of view of another observer, B standing on a platform as the train whizzes by. Because the train is moving, the two walls will move as the light flashes travel from emission to impact. As a result the distance to the back wall is shorter than that to the front wall. This means that observer B will not agree. The light hits the back wall first. Basically, the order of events is relative depending on who carries out the observation. This is an important lesson.

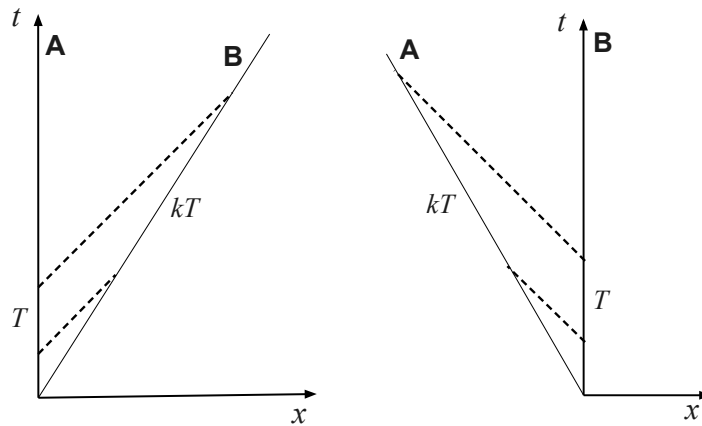


Figure 5.4: Relating measured time intervals in k -calculus. Left: from the point of view of A . Right: From the point of view of B .

Let us now relate measurements by two observers moving relative to one

another, as in figure 5.4. In order to do this, we assume that the time intervals involved are proportional to one another. This makes sense since the relative velocity is fixed. In effect, if observer A measures an interval T (say between two flashes of a flash light or two ticks on a clock), the B measures kT , where k is constant. This is the fundamental assumption of k -calculus. Note that, from the point of view of the other observer (B) measured time intervals must be related by the same constant of proportionality, as in the right panel of figure 5.4. Of course, if B moves in the positive x -direction relative to A , then A moves in the negative direction relative to B . It is important to keep this in mind.

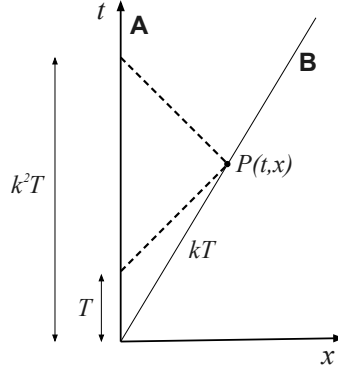


Figure 5.5: Working out points in spacetime using k -calculus.

To make progress, consider the set up in figure 5.5. Assume that the two observers synchronise their clocks at the origin (when they pass through the same point in space). Comparing this figure to figure 5.2, and making use of the inferred relations between time intervals it is easy to see that we have $t_1 = T$ and $t_2 = k^2T$. As T varies, this allows us to work out the coordinates associated with events along the world line of B . We have

$$t = \frac{1}{2}(k^2 + 1)T$$

and

$$x = \frac{1}{2}(k^2 - 1)T$$

However, we can relate the result to the relative velocity, since

$$v = \frac{x}{t} = \frac{k^2 - 1}{k^2 + 1}$$

Solving for k in terms of v (noting that k must be greater than 1 if A and B move apart), we have

$$k = \left(\frac{1+v}{1-v} \right)^{1/2}$$

As we will see later, this is the relativistic formula for the Doppler shift. If B moves away from A then $k > 1$ which leads to frequencies being shifted towards the “red”, while they shift towards the “blue” when B moves towards A .

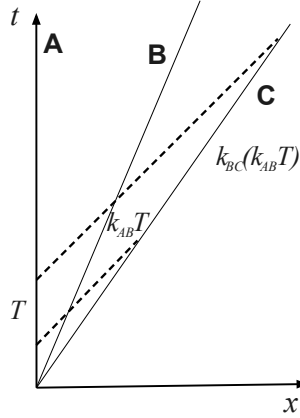


Figure 5.6: The idea behind relativistic velocity addition.

As a slight extension of the example, let us ask how velocities add in this picture. In this case we need to introduce a third observer C . This also involves extra k factors. We denote these as k_{AB} , k_{BC} and k_{AC} as in figure 5.6. However, these factors can not be independent. We must have

$$k_{AC} = k_{AB}k_{BC}$$

Re-writing this relation in terms of the three relative velocities (which involves a bit of algebra) we arrive at the composition law for velocities:

$$v_{AC} = \frac{v_{AB} + v_{BC}}{1 + v_{AB}v_{BC}}$$

This is very important result. In the limit of low velocities ($v_{AB} \ll 1$ etcetera as we assume units where $c = 1$), we have

$$v_{AC} \approx v_{AB} + v_{BC}$$

This accords with our usual (Newtonian) experience. Velocities add. However, now replace one of the observers with a light signal. Taking $v_{BC} = 1$ we see that the resulting speed relative to A is

$$v_{AC} = \frac{v_{AB} + 1}{1 + v_{AB}} = 1$$

This is new. We learn that we cannot exceed the speed of light by adding velocities – there is an upper speed limit. If we add two velocities, which are both below the speed of light, we arrive at a composite velocity which is also below the speed of light.

5.2 The Lorentz transformation

We are now ready to get to the core argument – the derivation of the Lorentz transformation. In order to discuss this, we consider the setting represented in

figure 5.7. We let the event P have coordinates (t, x) with respect to observer A and (t', x') with respect to observer B . We know from before that observer A has to send out a light signal at $t_1 = t - x$ in order to observe the event at $t_2 = t + x$ (see the symmetry in figure 5.2). Similarly, observer B has to send out a light signal at $t'_1 = t' - x'$ in order to observe the event at $t'_2 = t' + x'$ (it helps to ask what figure 5.2 would look like according to B).

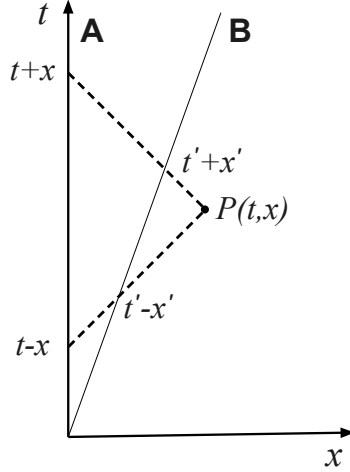


Figure 5.7: Deriving the Lorentz transformation.

Assuming that the observers have synchronised their clocks at the origin (as before), we know from the k -calculus argument that

$$t' - x' = k(t - x)$$

and

$$t + x = k(t' + x')$$

Re-arranging these relations (and making use of the result for v in terms of k) we arrive at

$$t' = \frac{1}{2} \left[\frac{1}{k}(t + x) + k(t - x) \right] = \gamma(t - vx)$$

$$x' = \frac{1}{2} \left[\frac{1}{k}(t + x) - k(t - x) \right] = \gamma(x - vt)$$

where

$$\gamma = (1 - v^2)^{-1/2}$$

This is the Lorentz transformation. The derivation is surprisingly simple – once we have laid the foundations by discussing measurements carried out by observers in relative motion.

It is worth noting (and quite easy to show) that the quantity $t^2 - x^2$ is invariant under the Lorentz transformation. That is, we have

$$(t')^2 - (x')^2 = t^2 - x^2$$

This will be important later.

5.3 Implications: length contraction and time dilation

With the Lorentz transformation in hand, let us consider a couple of implications. The basic story is simple – we have to accept that space and time are flexible, and that measurements depend on the motion of the observer.

First of all, consider a rod moving with speed v in the x -direction relative to a fixed observer, as in figure 5.8. If we measure the length of the rod at rest (in the moving frame), we have

$$l_0 = x'_B - x'_A$$

But this is not the length measured by the fixed non-moving observer. Instead, this observer measures

$$l = x_B - x_A$$

However, if this measurement is done at time $t = t_0$ (on the non-moving clock), then we know that

$$x'_A = \gamma(x_A - vt_0)$$

and

$$x'_B = \gamma(x_B - vt_0)$$

so we find

$$l_0 = x'_B - x'_A = \gamma(x_B - x_A) = \gamma l \longrightarrow l < l_0$$

The moving rod appears to be contracted in the direction of travel.

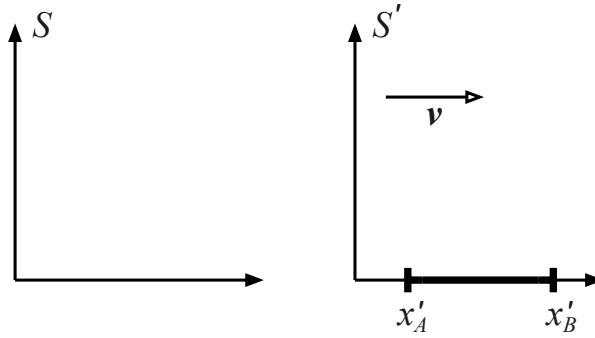


Figure 5.8: Standard configuration for systems moving relative to one another. This is the set-up for the derivation of length contraction, with a moving rod measured by a fixed observer.

Now consider a moving clock. With the clock at rest in the primed frame (the same as in figure 5.8), two ticks of the clock are given by coordinates (x'_1, t'_1) and $(x'_1, t'_2) = (x'_1, t'_1 + T_0)$ (say). Transforming this into the non-moving frame³ we have

$$t_1 = \gamma(t'_1 + vx'_1)$$

³This is easiest done by assuming that the observer moves away from the clock in the $-x$ direction.

and

$$t_2 = \gamma(t'_1 + T_0 + vx'_1)$$

This means that

$$T = t_2 - t_1 = \gamma T_0 \longrightarrow T > T_0$$

Time intervals on the moving clock appear longer, so the clock seems to run slow. This is called *time dilation*.

It is also instructive to consider the Doppler shift. We know from everyday experience that the frequency of a moving object (say the siren of an ambulance) is affected by relative motion. To see how this comes about, consider a source emitting waves with wavelength λ_0 in the frame of the emitter. Suppose the source is a flashing light. If two flashes are emitted a time $\delta t'$ apart, then the second flash has to travel a distance $v\delta t'$ further in order to reach a non-moving observer. This means that it will arrive a time $v\delta t'/c$ later. That is, the flashes arrive with a time difference

$$\delta t = \left(1 + \frac{v}{c}\right) \delta t'$$

Translating this into wavelength (inverse of frequency=proportional to time interval) we see that

$$\frac{\lambda}{\lambda_0} = 1 + \frac{v}{c}$$

This is an entirely classical effect. Of course, it will be affected by relativity. Adding in the time dilation the time difference will be measured by

$$\delta t \rightarrow \gamma \delta t = \gamma \left(1 + \frac{v}{c}\right) \delta t'$$

so now we have

$$\frac{\lambda}{\lambda_0} = \gamma \left(1 + \frac{v}{c}\right) = \left(\frac{1 + v/c}{1 - v^2/c^2}\right)^{1/2} = \left(\frac{1 + v/c}{1 - v/c}\right)^{1/2}$$

But we have seen this result before. The relativistic Doppler shift is simply the k -factor.

5.4 Paradoxes

Special relativity is (famously) associated with a number of (apparent) paradoxes. This is not surprising – we are messing with fundamental concepts like space and time, and the implications may be uncomfortable. These paradoxes are typically designed in such way that the theory appears to be inconsistent – but at a closer inspection one always finds that everything is fine.

As an example, let us consider the so-called “barn paradox”. This story involves a pole (perhaps carried by an olympic pole vaulter), parallel to the ground, moving at relativistic speed towards a barn. The pole is imagined passing through the open front and rear doors of barn which is shorter than its rest length, so if the pole was not moving it would not fit inside. To a stationary observer (perhaps a cheeky farmer), due to length contraction, the moving pole should fit inside the barn as it passes through. Of course, from the point of view of the person carrying the pole it is the barn which will be Lorentz

contracted. Therefore the pole can not possibly fit inside the barn. This poses an apparent contradiction.

This paradox results from the mistaken assumption of absolute simultaneity. The pole is said to fit into the barn if both of its ends can be made to be simultaneously inside. The paradox is resolved when we consider that in relativity, simultaneity is relative to each observer, making the answer to whether the pole fits inside the barn relative to each of them. We can illustrate this by drawing the different coordinate systems arising from the Lorentz contraction on top of one another. This leads to figure 5.9.

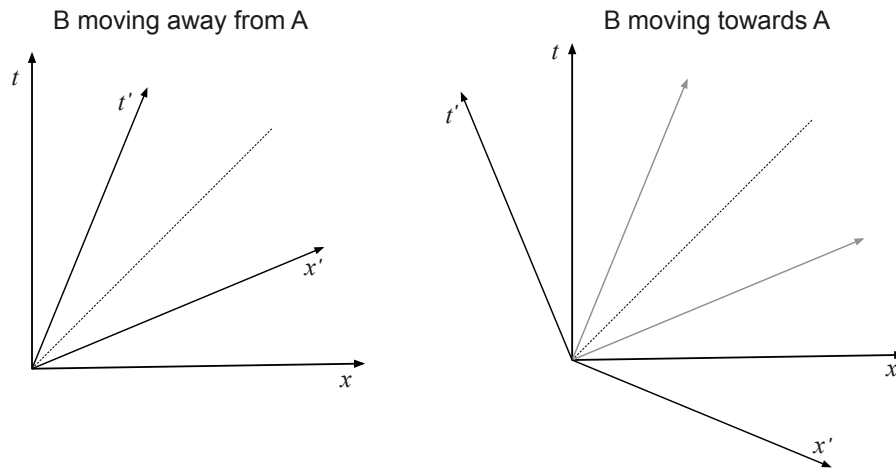


Figure 5.9: The relation between coordinate axes according to the Lorentz transformation.

With this picture in hand it is easy to see how the issue of simultaneity enters the length contraction argument (and ultimately resolves the barn paradox). An illustration of this – for the case of a rod moving away from the observer – is provided in figure 5.10.

Another famous paradox involves a pair of twins, one of whom makes a journey into space in a high-speed rocket and returns home to find that the twin who remained on Earth has aged more. At first, this result is puzzling because each twin should see the other twin as moving, and so each should (paradoxically) find the other to have aged less. Can you figure out how to use a spacetime diagram to resolve the issue?

5.5 Going further: Relativistic dynamics

Having explored the kinematics of special relativity, let us move on to dynamics. The first (important) step connects with the Lorentz transformation. If we consider small space and time intervals then we have seen that the *spacetime interval*⁴

$$ds^2 = -dt^2 + dx^2$$

⁴For simplicity, we will restrict our discussion to a single space dimension.

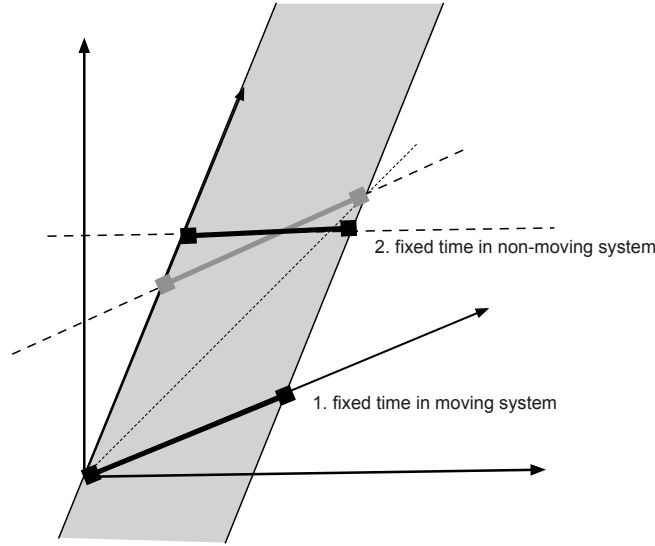


Figure 5.10: Illustrating length contraction.

is invariant – we get the same value for ds^2 in a fixed and a moving frame. Let us now define the proper time (=the time measured on a co-moving clock) through

$$d\tau^2 = -ds^2 \quad (5.1)$$

(recalling that we are using units where the speed of light, $c = 1$).

In order to define the squared distance in space-time, we need to introduce a new dot-product which reflects the combination used in the invariant spacetime interval. First we need the concept of a four-vector⁵;

$$d\vec{X} = \begin{pmatrix} dt \\ dx \end{pmatrix}$$

Recalling the usual dot product;

$$v^2 = \mathbf{v} \cdot \mathbf{v} = (v_1 \ v_2) \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = (v_1 \ v_2) \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = v_1^2 + v_2^2$$

and comparing to the spacetime interval, it is easy to see that we now need

$$ds^2 = d\vec{X} \cdot d\vec{X} = (dt \ dx) \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} dt \\ dx \end{pmatrix} = -dt^2 + dx^2$$

For future reference, the matrix

$$\eta = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}$$

⁵We use capital letters for four-vectors, and denote them with an arrow, to contrast with the usual spatial vectors, which will be lower case and indicated as bold. We are also suppressing two of the space dimensions here...

is called the metric – so named because it allows us to measure distance in space-time. However, for our present purposes we only need to know the rule for the new dot-product; basically that the product of the time-components come with a minus sign, while the squares of spatial components are added, as usual.

As we are interested in dynamics, we need a velocity for objects moving in space-time. Thus we define the four-velocity

$$\vec{U} = \frac{d\vec{X}}{d\tau}$$

and it is easy to see (using (5.1)) that

$$\vec{U} \cdot \vec{U} = \frac{d\vec{X}}{d\tau} \cdot \frac{d\vec{X}}{d\tau} = \frac{ds^2}{d\tau^2} = -1$$

(if we put the the correct units in, then the right-hand side should be $-c^2$). In effect, the normalisation of the four-velocity means that it has only three undetermined components – just like the velocity in Newtonian physics.

As a slight aside (at least at this point) it is worth considering the four-velocity as measured by an observer not riding along with the object in question. This observer would (still in 2D) measure a (four-) vector

$$\vec{U} = (U^0 \quad U^1)$$

such that

$$\vec{U} \cdot \vec{U} = -(U^0)^2 + (U^1)^2 = -1$$

or

$$(U^0)^2 \left[1 - \left(\frac{U^1}{U^0} \right) \right] = -1$$

Now let $U^0 = A$ and define the relative velocity of the observer and the measured object as

$$v = \frac{U^1}{U^0}$$

Then it follows that

$$A = \gamma = (1 - v^2)^{-1/2}$$

We recognise this as the Lorentz factor from before. This should, of course, have been expected. In general, it means that the four-velocity takes the form

$$\vec{U} = \gamma (1 \quad \mathbf{v}) \tag{5.2}$$

It is also worth noting that the Lorentz transformation applies to general four-vectors. Staying in 2D, we have⁶

$$(U^0)' = \gamma(V) (U^0 - VU^1)$$

$$(U^1)' = \gamma(V) (U^1 - VU^0)$$

⁶Recall that we have

$$t' = \gamma(t - vx) .$$

where V is the relative velocity between the two frames and

$$\gamma(V) = (1 - V^2)^{-1/2}$$

Applying this transformation to (5.2), we have

$$(U^1)' = \gamma(V)\gamma(v)(v - V)$$

which vanishes if we make the observer move along with the object ($V \rightarrow v$), which is kind of obvious. In this co-moving frame, we also find that

$$(U^0)' = \gamma^2(v)(1 - v^2) = 1$$

as expected.

Moving towards dynamics, we define the four-momentum as

$$\vec{P} = m\vec{U}$$

where m is the mass of the object under consideration. It is easy to see that this means that we have

$$\vec{P} = m\gamma(1 - \mathbf{v})$$

and it also follows (immediately) that

$$\vec{P} \cdot \vec{P} = -m^2$$

Consider the expression for the four momentum in the low-velocity limit where $\mathbf{v} \ll 1$ ($= c$). First of all, the spatial components clearly lead back to the Newtonian momentum $\mathbf{p} = m\mathbf{v}$. Meanwhile, the time component leads to

$$P^0 = m\gamma \approx m + \frac{1}{2}mv^2$$

or, if we put the speed of light back in

$$P^0 = m\gamma \approx mc^2 + \frac{1}{2}mv^2$$

We recognise the second term as the kinetic energy. This means that the first term has to be an energy, as well. There is an energy associated with mass! Letting $E = P^0$ we see that we have

$$E = mc^2$$

for a body at rest. This is perhaps the most famous equation of all, and we have just derived it...

In general, the energy of an object follows from (again with $c = 1$)

$$E = -\vec{U} \cdot \vec{P} = -m \frac{ds^2}{d\tau^2} = m$$

These quantities are important because, it is the four-momentum that is conserved in relativistic collisions—as explored by the Large Hadron Collider.

Chapter 6

Conservation Laws

Let us now introduce the (hugely important) idea of conservation laws. As the name suggests, a conserved quantity is one that does not change throughout the motion of a system. Conservation laws can often be used as an alternative formulation of dynamical problems—they allow us to introduce suitable variables that remain fixed as the system evolves¹. We will deal with two major conservation properties, relating to **energy** and **momentum**. These conservation laws extend the range of mechanical problems we can attack. We shall also use the concept of energy to discuss the *stability* of a static situation: if a static system is displaced slightly, will it return to the same static state, or will there be a dramatic (or even catastrophic) change in its structure. We have already alluded to these properties in the context of potential motion.

6.1 Energy

The history of the concept of energy is long and varied. We might easily be able to mathematically define an quantity that we could call the energy, and have an intuitive idea of what it means. However historically there were great issues with the concept of physical mechanisms that could convert, say, thermal energy into energy of motion. Were they the same “thing”? How was this achieved? Did they have the same origin? What was the mechanical equivalent of heat? This was not some arcane academic debate: it underlay the whole of the industrial revolution in the UK, which was largely powered by the conversion of the thermal energy of steam into the kinetic energy of pumps/wheels/looms etc. and so the industrial revolution of the whole world.

Originally heat energy was assumed to be a fluid (the caloric theory) that was transferred between bodies and could neither be created nor destroyed. This theory was demolished by the major experimental advances of James Joule who showed that heat can be produced by motion. For this reason the unit of energy is named the “joule”. A joule can be defined in many equivalent ways, for example:

- The energy expended by a force of 1N moving an object a distance of 1 m in the direction of the force.

¹It always helps to be able to with constant quantities.

- The energy expended when a current of 1 amp to pass through an (ohmic) electrical resistance of 1 ohm.

We will not be concerned with the philosophical issues of energy. Rather, we will use it as a mathematical tool.

6.2 Work and Energy

It is natural to start by introducing the idea of energy from Newton's laws of motion. We know that Newton's second law of motion states that the rate of change of momentum \vec{p} with time is equal to the force applied \vec{F} :

$$\vec{F} = \frac{d\vec{p}}{dt}.$$

Here we shall consider this body to be of *constant* mass m and at time t at position $\vec{r}(t)$ relative to a certain inertial frame. Using the definition $\vec{p} = m\dot{\vec{r}}(t)$, we can then write **N2** as:

$$\vec{F} = m\ddot{\vec{r}}(t).$$

We can dot-product both sides with the velocity $\dot{\vec{r}}(t)$.

$$\begin{aligned}\vec{F} \cdot \dot{\vec{r}}(t) &= m \left(\ddot{\vec{r}}(t) \cdot \dot{\vec{r}}(t) \right) \\ &= \frac{1}{2}m \frac{d}{dt} \left(\dot{\vec{r}}(t) \cdot \dot{\vec{r}}(t) \right) \\ &= \frac{d}{dt} \left(\frac{1}{2}m |\dot{\vec{r}}(t)|^2 \right). \quad (6.1)\end{aligned}$$

The kinetic energy T of a particle of constant mass m and velocity $\dot{\vec{r}}(t)$ relative to an inertial frame is then defined to be:

$$T \equiv \frac{1}{2}m |\dot{\vec{r}}(t)|^2.$$

Obviously the kinetic energy of a stationary particle is zero relative to its own rest frame. Thus the quantity on the left-hand side of (6.1) is the rate of change of kinetic energy with time

$$\vec{F} \cdot \dot{\vec{r}}(t) = \frac{dT}{dt}.$$

Suppose that between $t = a$ and $t = b$, the particle moves from $r(a)$ to $r(b)$. We can then integrate both sides with respect to t to give:

$$\int_a^b \vec{F} \cdot \frac{d\vec{r}}{dt} dt = \int_a^b \frac{dT}{dt} dt = T(b) - T(a).$$

The right-hand side is the difference in kinetic energy of the body at times a and b . The left-hand side is the amount of energy that must have been expended by the force to make this change. It is defined to be the work done by the force on the particle. Hence we have:

The work W done by a force F on a particle moving along a trajectory $r(t)$ in time $a \leq t \leq b$ is the amount of energy given by the integral

$$W(\vec{r}(a), \vec{r}(b)) \equiv \int_a^b \vec{F} \cdot \frac{d\vec{r}}{dt} dt = \int_a^b \vec{F} \cdot d\vec{r} \quad (6.2)$$

and we have derived the law of conservation of energy:

$$W(a, b) = T(b) - T(a)$$

In words the law of conservation of energy states that the work done equals the change in kinetic energy of a system. If $W > 0$, the force has increased the speed of the particle. If $W < 0$, the force has resisted the motion of the particle, it has reduced in speed and dissipated energy.

Overall energy is conserved. The work done decelerating (say) a body will be dissipated as frictional heat and/or acoustic energy. The work done accelerating a body to increase its kinetic energy must come from a source, such as chemical combustion of a fuel. However for now, we will not be concerned with the form of the energy that the work done takes.

Before we can use (6.2) we need to define how we evaluate the weird looking work integral. Such integrals are called **line integrals** because they are integrals that depend on the path over which the integration is carried out.

6.3 A slight aside: Vector Fields and Line Integrals

As a useful reminder, let us outline some jargon that you will come across in reading texts about how to evaluate line integrals.

First of all, a *scalar field* is just a fancy description for an ordinary function defined at each point in space. The input is a vector $\vec{r} = x\hat{i} + y\hat{j} + z\hat{k}$, (where x, y , and z may themselves depend on another parameter, e.g., time t) the output is a number $f(\vec{r}) = f(x, y, z)$. Example of scalar fields are:

- the altitude of a landscape above (or below) sea level;
- temperature of the atmosphere;
- the pressure of the atmosphere;
- the wind *speed* at a point.

Scalar fields are often represented by drawing contour plots of curves/surfaces where $f(\vec{r}) = \text{constant}$.

Similarly, a vector field is a function whose input is a vector $\vec{r} = x\hat{i} + y\hat{j} + z\hat{k}$ while the output is a vector quantity. In a Cartesian frame it takes the form

$$\vec{F}(\vec{r}) = F_1(\vec{r})\hat{i} + F_2(\vec{r})\hat{j} + F_3(\vec{r})\hat{k}.$$

Examples of vector fields are:

- the wind *velocity* on a weather map;
- the Earth's magnetic field at any point on its surface;
- the electric field near a pylon;
- the vorticity of a fluid flow.

Vector fields are usually represented with arrow plots, where the length of the arrow is proportional to the strength of the field. The forces in **N2** can be regarded as vector fields.

If we want to quantify the work done by a certain process, we need to evaluate a line integral. An ordinary 1D integral is the integration of a scalar function $f(x)$ over a straight line segment $a \leq x \leq b$

$$\int_a^b f(x) dx$$

It represents the area between $f(x)$ and the x -axis between $x = a$ and $x = b$. If $f(x)$ represents the height of a straight fence, then the integral is the area of the fence panels.

The work integral we need to evaluate

$$\int_a^b \vec{F} \cdot \frac{d\vec{r}}{dt} dt$$

is a generalisation of this definition.²

- First the integration is now over a given curve C , or $\vec{r}(t)$ with $a \leq t \leq b$. This is equivalent to working out the area of a fence that is placed along a curved base.
- Second, the integrand involves vector functions \vec{F} and $\dot{\vec{r}}$. However these combine as a dot product to form a scalar function $\vec{F} \cdot \dot{\vec{r}}$.

We represent the trajectory by the curve C :

$$\vec{r}(t) = x(t)\hat{i} + y(t)\hat{j} + z(t)\hat{k}, \quad a \leq t \leq b$$

which has a tangent vector

$$\frac{d\vec{r}}{dt} = \frac{dx}{dt}\hat{i} + \frac{dy}{dt}\hat{j} + \frac{dz}{dt}\hat{k}, \quad a \leq t \leq b.$$

²We have introduced this type of integral using the concept of work, but it is generally defined for any vector field, independently of its physical meaning. In fact t need not be time: it is just a parameter describing positions on C . Furthermore, $\vec{F}(\vec{r})$ really only depends on the position \vec{r} on the curve C , not on the parameter describing it. How you represent C or \vec{r} in terms of parameters should not affect the answer.

With these definitions, the force experienced by the particle will (in general) differ at each point of the trajectory.³

To evaluate the integral, we evaluate \vec{F} at each point on the trajectory as a function of the curve parameter t , take the dot product

$$\vec{F} \cdot \frac{d\vec{r}}{dt},$$

and then integrate the resulting scalar function of t between a and b . Sometimes the work done depends on the path taken by the body. Sometimes it does not.

6.4 Conservation of Linear Momentum

Energy is not the only quantity that is conserved in Newtonian mechanics. We may also consider the linear momentum.

N2 tells us that

$$\vec{F} = \frac{d\vec{p}}{dt},$$

where $\vec{p} = m\vec{v}$ is the linear momentum of the body. If there are no external forces acting, then

$$\vec{F} = 0 \Rightarrow \frac{d\vec{p}}{dt} = 0 \Rightarrow \vec{p} = \text{constant}.$$

Hence we see that the linear momentum of a body is conserved if the total force on it is zero.

Components of momentum can also be conserved. Suppose \vec{s} is a constant vector such that $\vec{F} \cdot \vec{s} = 0$. Thus the force \vec{F} has no component in the \vec{s} direction. If we take the scalar product of both sides of **N2** with \vec{s} we obtain:

$$\vec{F} = \frac{d\vec{p}}{dt} \Rightarrow \vec{F} \cdot \vec{s} = \frac{d\vec{p}}{dt} \cdot \vec{s} \Rightarrow 0 = \frac{d(\vec{p} \cdot \vec{s})}{dt}$$

(since \vec{s} is constant). Hence we deduce that

$$\vec{p} \cdot \vec{s} = \text{constant}.$$

In words, the *component* of linear momentum in a direction in which the component of force vanishes is constant in time.

Example 13: Golf on the moon

An astronaut plays golf on the surface of the moon, where there is little atmosphere (i.e., negligible air resistance). The golf ball of mass m is launched and moves with velocity $\vec{v}(t)$, subject only to the moon's gravity $-g\hat{k}$.

³Thus the force is not constant and so we can't apply the usual formula of "Work done = Component of force in direction of motion \times distance moved". However we can split the trajectory into very small sections of length $d\vec{r}$ over which the force is approximately constant, apply the old formula and then sum (= integrate) over them.

N2 gives:

$$\begin{aligned}\vec{F} = \frac{d\vec{p}}{dt} \Rightarrow -mg\hat{k} &= m \frac{d\vec{v}}{dt} \Rightarrow -mg\hat{k} \cdot \hat{i} = m \frac{d\vec{v}}{dt} \cdot \hat{i} \\ \Rightarrow 0 &= m \frac{d}{dt} (\vec{v} \cdot \hat{i}).\end{aligned}$$

In other words, the \hat{i} component of $\vec{v}(t)$ is constant throughout the motion. This is simply a reflection of the fact that the force has no component in this direction.

6.5 Conservative Forces and potentials

In the context of mechanics, conservative forces represent a special class of force fields for which the work done on a body by the field is independent of the motion⁴.

The idea is simple. If a force field $\vec{F}(\vec{r})$ at any point $\vec{r}(t) = x(t)\vec{i} + y(t)\vec{j} + z(t)\vec{k}$ can be represented by the gradient of a scalar function $V(\vec{r})$ of the form⁵

$$\vec{F}(\vec{r}) \equiv -\nabla V(\vec{r})$$

then it is a **conservative force field**.

The work done by such a field is independent of the path taken, it depends only on the initial and final states.

If a force is conservative such that $\vec{F} = -\nabla V$, then the function V is called the **potential energy** of the field (or **potential** for short).

Let us consider a simple example.

Example 14: Gravity

Gravity is a conservative force field. If we take the z -axis to point upwards, then

$$\begin{aligned}\vec{F} = -mg\hat{k} \Rightarrow \vec{F} &= -\frac{d}{dz}(mgz)\hat{k} \\ \Rightarrow \vec{F} &= -\frac{d}{dz}(mgz)\end{aligned}$$

The gravitational potential is thus

$$V(z) = +mgz.$$

If the z -axis had instead pointed downwards, we would have obtained $V = -mgz$. In both cases, an arbitrary constant can be added.

⁴This is useful as it significantly simplifies the evaluation of the work integral.

⁵Note that this requires us to use partial derivatives, which we are (for the moment) trying to avoid.

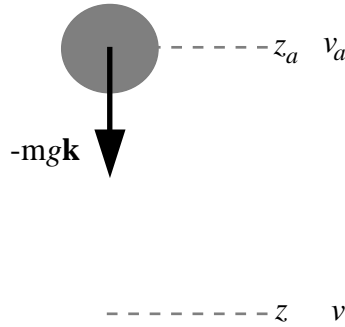


Figure 6.1: A body falling in a gravitational potential

Example 15: Motion Under Gravity

Conservation of energy can often be used as an alternative to the solution of a dynamical system using **N2**. This example provides a demonstration.

A body of mass m is thrown vertically upward under the action of a gravitational field $-mg\hat{k}$. Suppose that the particle has initial speed v_a at height $z_a\hat{k}$. Find its speed $v = |\dot{\vec{r}}|$ as a function of position

Gravity is a conservative force with potential (see above) $V(z) = mgz$. In the absence of non-conservative fields we have

$$T(z) + V(z) = \text{constant} = E, \text{ say}$$

$$\frac{1}{2}m|\dot{\vec{r}}|^2 + mgz = \text{constant} = \frac{1}{2}mv_a^2 + mgz_a$$

$$\frac{1}{2}mv^2 + mgz = \frac{1}{2}mv_a^2 + mgz_a$$

$$\Rightarrow v(z) = \pm\sqrt{v_a^2 + 2g(z_a - z)} \text{ or}$$

$$v(z) = \pm\sqrt{\frac{2}{m}\sqrt{E - mgz}}$$

The maximum height of the body is reached when $v(z) = 0$, i.e., $z - z_a = v_a^2/(2g)$.

6.6 One-dimensional motion

Let us now focus on a body moving along a line, so that its position is determined by a single function $x(t)$ (say). For now, suppose that the force on the particle depends only on the position, not the velocity. That is, we have $F = F(x)$. We

then define the potential $V(x)$ in such a way that

$$F(x) = -\frac{dV}{dx}$$

(note the sign!). Obviously, the potential is only defined up to a constant. We can always invert the equation by integrating both sides. The integration constant is now determined by the choice of lower limit of the integral,

$$V(x) = -\int_{x_0}^x F(x')dx'$$

where x' is a dummy variable. Given the definition of the potential, we can write the equation of motion as

$$m\ddot{x} = -\frac{dV}{dx} \quad (6.3)$$

For any force in one dimension which depends only on the position, there exists a conserved **energy**,

$$E = \frac{1}{2}m\dot{x}^2 + V(x)$$

As before, the fact that this is conserved means that $\dot{E} = 0$ for any trajectory that obeys the equation of motion. Introducing the kinetic energy,

$$T = \frac{1}{2}m\dot{x}^2$$

it is easy to show that the energy is conserved. We only need to differentiate to get

$$\dot{E} = m\dot{x}\ddot{x} + \frac{dV}{dx}\dot{x} = \dot{x}\left(m\ddot{x} + \frac{dV}{dx}\right) = 0$$

where the last equality holds courtesy of the equation of motion (6.3). We also see why the sign of the potential in Newton's second law makes sense.

In any physical system, conserved quantities of this kind are precious—because they help us simplifying the solution. Hence, we will spend some time finding them and showing how they help us simplify various problems.

6.7 Stability of Equilibrium

If a body is stationary it is said to be in equilibrium. From **N2** we have seen that a necessary condition for this is that the sum of the forces acting on the body \vec{F} must vanish. Now if all the forces acting on the system are conservative so that $\vec{F} = -V'$ we then have a necessary condition for equilibrium as

$$\vec{F} = \vec{0} \longrightarrow V' = 0.$$

Thus a system will be in equilibrium at a stationary point (or critical point) of the potential. This definition is valid in all dimensions.

To explore whether a position of equilibrium is stable, unstable (or otherwise), we will consider a 1-D potential $V(x)$.

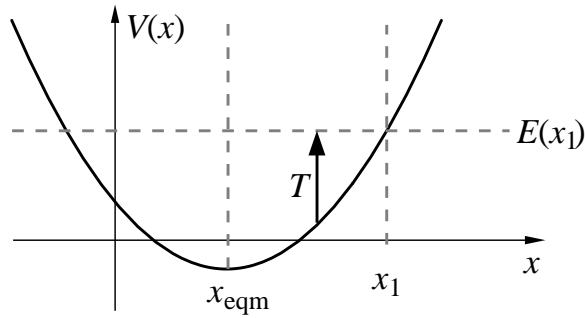


Figure 6.2: Stable equilibrium.

Example 16: Stable equilibrium

Consider a potential function $V(x)$ that behaves qualitatively like that in the diagram with a minimum of potential energy at a point x_0 . Suppose a particle is released from rest at a point x_1 relative to the minimum. By conservation of total energy, $E = E(x_1) = \text{constant}$, throughout the motion the kinetic, T , and potential, V , energies are related by

$$E = T + V \Rightarrow T = E - V.$$

Now since $T = m|\dot{r}|^2/2$ we must have $T \geq 0$. Hence we can write

$$T \geq 0 \Rightarrow E - V \geq 0 \Rightarrow V \leq E$$

i.e. the motion is confined to regions where $V \leq E$. Thus the particle will oscillate in a finite region around the minimum of V and so the motion is bounded. Hence the equilibrium point is said to be stable.

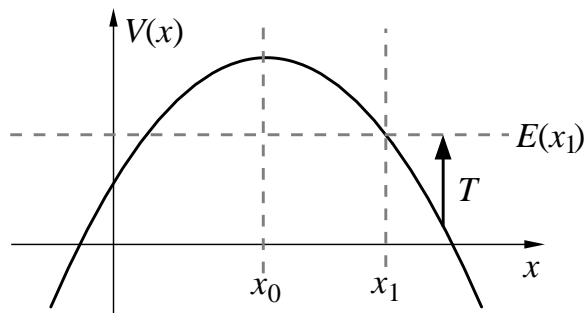


Figure 6.3: Unstable equilibrium.

Example 17: Unstable equilibrium

Consider now a potential that has a maximum at x_0 . If a particle released from a small distance away, x_1 , again energy is conserved $E = E(x_1) = \text{constant}$ throughout the motion. Now since the kinetic energy must be non-negative, we can again write

$$T \geq 0 \Rightarrow E - V \geq 0 \Rightarrow V \leq E$$

so again, the only classically allowed motion must have $V \leq E$. Hence the particle cannot approach the maximum and must move away from it. As it does so the kinetic energy T and hence speed will increase. Hence the equilibrium point is said to be unstable.

In general, an equilibrium point in a conservative system where $\vec{F} = -\nabla V$ occurs when

$$\nabla V = 0.$$

- It is **stable** if it is a **minimum** (in 1-D $\Rightarrow V''(x) > 0$).
- It is **unstable** if it is a **maximum** (in 1-D $\Rightarrow V''(x) < 0$).
- Otherwise it may be stable or unstable and needs further investigation (in 1-D $\Rightarrow V''(x) = 0$).

In 1-D, max/min points where $T = 0$, i.e., $E = V$ are called **turning points**. On one side of such points the classical motion is allowed, and on the other it is not. Hence a moving particle is said to “turn around” there⁶.

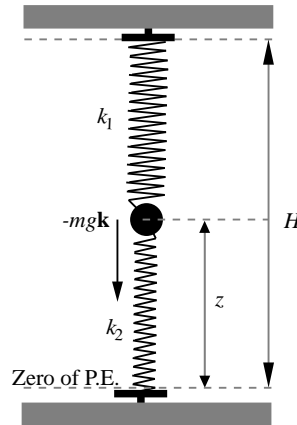


Figure 6.4: Elastic energy is stored in two springs

⁶An analogous argument can be applied in more than one dimension. In higher dimensions isolated maxima of the potential are unstable equilibrium points, isolated minima are stable points. (Just think of rolling a ball down a hill.) However, saddle points also satisfy the equilibrium condition $\nabla V = 0$ and so are also equilibrium points. Their stability or otherwise, depends on the motion perturbing the equilibrium.

Example 18: Coupled springs

A mass m is attached to two Hookean springs with spring constants k_1, k_2 and natural lengths l_1, l_2 , as in figure 6.4. Determine the equilibrium point and whether it is stable or unstable.

Potential energy:

- Gravitational: mgz
- Spring 2: $\frac{1}{2}k_2(z - l_2)^2$
- Spring 1: $\frac{1}{2}k_1(H - z - l_1)^2$

$$V(z) = mgz + \frac{1}{2}k_2(z - l_2)^2 + \frac{1}{2}k_1(H - z - l_1)^2$$

Equilibrium position:

$$\begin{aligned} V'(z) &= mg + k_2(z - l_2) - k_1(H - z - l_1) = 0 \\ z_{\text{eqm}} &= \frac{k_1(H - l_1) + k_2l_2 - mg}{k_1 + k_2} \end{aligned}$$

Stability:

$$V''(z) = k_1 + k_2 > 0$$

The equilibrium point is a minimum and hence it is stable.

6.8 Moving in a potential

Let us consider the general case of a potential $V(x)$. It turns out that we can learn quite a lot about the motion without actually solving the problem. For example, the existence of a conserved energy magically allows us to turn the problem into a first order differential equation

$$E = \frac{1}{2}m\dot{x}^2 + V(x) \quad \longrightarrow \quad \frac{dx}{dt} = \pm \sqrt{\frac{2}{m} [E - V(x)]}$$

Noting that the right-hand side does not depend on t , we have our first hint of the importance of conserved quantities in solving a problem. Of course, to go from a second order equation (like **N2**) to a first order equation, we must have chosen an integration constant. In this case, this is the energy E itself. Given a first order equation, we can always write down a formal solution for the dynamics by integrating. We have

$$t - t_0 = \pm \int_{x_0}^x \frac{dx'}{\sqrt{\frac{2}{m} [E - V(x')]} \quad (6.4)$$

If we can work out the integral, we are done. If we can't do the integral, you sometimes hear that the problem has been “reduced to quadrature”. This really means “I can't do the integral”, but it is often the case that having a solution in this form still sheds light on some of the properties. And, if nothing else, one can always evaluate the integral numerically – computers are really good at integration.

6.9 Getting a feel for the solution

Given the potential $V(x)$, it is often easy to figure out the qualitative nature of any trajectory simply by looking at the form of $V(x)$. This allows us to answer some questions with very little work⁷. For example, we may want to know whether the particle is trapped within some region of space or can escape to infinity.

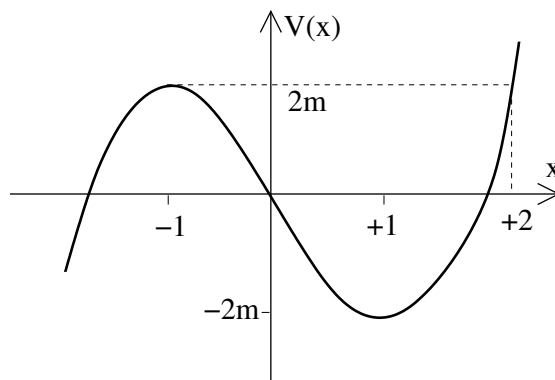


Figure 6.5: A cubic potential.

Let us illustrate this with an example. Consider the cubic potential

$$V(x) = m(x^3 - 3x)$$

If we were to substitute this into the general form (6.4), we have a fearsome looking integral which we struggle to solve⁸. Luckily, we can make progress even without solving the integral. The potential is plotted in figure 6.5. Let us start with the particle sitting stationary at some position x_0 . This means that the energy is

$$E = V(x_0)$$

and we know that this must remain constant during the subsequent motion. What happens next depends only on x_0 . We can identify the following possibilities:

⁷... and we like this.

⁸It can be solved – it is called an elliptic integral – but we do not yet have the tools to work this out.

- $x_0 = \pm 1$. These are the local maximum and minimum. If we drop the particle at these points, it stays there for all time.
- $x_0 \in (-1, +2)$. Here the particle is “trapped” in the dip. It oscillates backwards and forwards between the two points with potential energy $V(x_0)$. The particle cannot climb to the right because it does not have the energy. In principle, it could live off to the left where the potential energy is negative, but to get there it would have to first climb the small bump at $x = 1$ and it does not have the energy to this either⁹.
- $x_0 > 2$. When released, the particle falls into the dip, climbs out the other side, before falling into the void $x \rightarrow -\infty$.
- $x_0 < -1$. The particle just falls off to the left.
- $x_0 = +2$. This is a special value, since $E = 2m$ which is the same as the potential energy at the local maximum $x = 1$. The particle falls into the dip and starts to climb up towards $x = 1$. It can never stop before it reaches $x = 1$ for at its stopping point it would have only potential energy $V < 2m$. But, similarly, it cannot arrive at $x = 1$ with any excess kinetic energy. The only option is that the particle moves towards $x = 1$ at an ever decreasing speed, only reaching the maximum at time $t \rightarrow \infty$.

One can easily play a similar game to that above if the starting speed is not zero. In general, one finds that the particle is trapped in the dip $x \in [-1, +1]$ if its energy lies in the interval $E \in [-2m, 2m]$.

Example 19: The harmonic oscillator

The harmonic oscillator is, by far, the most important dynamical system in all of mathematical physics. The good news is that it is very easy to deal with. (In fact, the reason that it is so important is precisely because it is easy!). The potential energy of the harmonic oscillator is defined to be

$$V(x) = \frac{1}{2}kx^2$$

The harmonic oscillator is a good model for, among other things, a particle attached to the end of a spring. The force resulting from the energy V is given by $F = -kx$ which, in the context of the spring, is Hooke’s law. The equation of motion is (again, as before)

$$m\ddot{x} = -kx$$

with general solution

$$x(t) = A \cos \omega t + B \sin \omega t \quad \text{with} \quad \omega = \sqrt{\frac{k}{m}}$$

A and B are integration constants and ω is called the angular frequency (also... as before).

⁹There is an assumption here which is implicit throughout all of classical mechanics: the trajectory of the particle $x(t)$ is a continuous function.

6.10 Why (almost) everything is a harmonic oscillator

A particle placed at an **equilibrium** point will stay there for all time. In the example with a cubic potential, we had two such equilibrium points: $x = \pm 1$. In general, if we want $\dot{x} = 0$ for all time, then clearly we must have $\ddot{x} = 0$, which, from the form of the equation of motion, tells us that we can identify the equilibrium points with the critical points of the potential,

$$\frac{dV}{dx} = 0$$

What happens to a particle that is close to an equilibrium point, x_0 (say)? In this case, we can Taylor expand the potential energy about $x = x_0$. Because, by definition, the first derivative vanishes, we have

$$V(x) \approx V(x_0) + \frac{1}{2}V''(x_0)(x - x_0)^2 + \dots \quad (6.5)$$

In order to proceed, we need to know the sign of $V''(x_0)$.

- $V''(x_0) > 0$. In this case, the equilibrium is a minimum of the potential and the potential energy is that of a harmonic oscillator. The particle oscillates backwards and forwards around x_0 with frequency

$$\omega = \sqrt{\frac{V''(x_0)}{m}}$$

Such equilibria are stable. As long as the amplitude of the oscillations is small enough (so that we can ignore the $(x - x_0)^3$ and higher order terms in the Taylor expansion), all systems oscillating around a stable fixed point behave like a harmonic oscillator.

- $V''(x_0) < 0$. In this case, the equilibrium point is a maximum of the potential. The equation of motion again reads

$$m\ddot{x} = -V''(x_0)(x - x_0)$$

but if the second derivative of the potential is negative, then we have $\ddot{x} > 0$ when $x - x_0 > 0$. This means that if we displace the system a little bit away from the equilibrium point, the acceleration pushes it further away. The equilibrium is unstable.

- Finally, we could have $V''(x_0) = 0$. In this case, there is nothing we can say about the dynamics of the system without Taylor expanding the potential further.

6.11 Addendum: Taylor expansions

The Taylor series is one of the most useful mathematical tools applied to real life problems. As in the example of potential motion near equilibrium, we can often use a Taylor series to approximate a complicated function by a (simpler) polynomial function.

The starting point is Taylor's theorem (which we state without proof, as we are mainly interested in the applications):

Let $f(x)$ be continuous and differentiable an infinite number of times at point $x = a$ (i.e., f and all of its derivatives f', f'', \dots exist at $x = a$). Then $f(x)$ can be expanded about the point $x = a$ in terms of powers of $(x - a)$, and this expansion is given by the formula

$$f(x) = f(a) + (x - a)f'(a) + \frac{(x - a)^2}{2!}f''(a) + \dots + \frac{(x - a)^n}{n!}f^{(n)}(a) + O(x - a)^{n+1}$$

This is a **Taylor expansion**. It has the form

$$f(x) = P_n(x) + O(x - a)^{n+1}$$

where $P_n(x)$ is a polynomial of degree n and $O(x - a)^{n+1}$ is a remainder term. The remainder term reminds us that in writing the Taylor expansion explicitly up to the term $\propto (x - a)^n$ ("up to order n ") we are omitting terms "of order $(x - a)^{n+1}$ and higher".

For x close to a , we can expect the remainder term to be small (since $x - a$ is small), so that $f(x) \simeq P_n(x)$. Thus Taylor's theorem enables us to approximate functions by polynomials.

In many cases the remainder term goes to zero as $n \rightarrow \infty$ and the Taylor series expansion of $f(x)$ takes the form of a possibly infinite, but convergent, series.

Example 20: The exponential

As an example, let us determine the Taylor series expansion of the function $f(x) = e^x$ about $x = 0$.

We first establish the necessary data. We know that we need the derivatives:

$$\begin{aligned} f(x) &= e^x & f'(x) &= e^x & f''(x) &= e^x & \dots \\ f(0) &= e^0 = 1 & f'(0) &= e^0 = 1 & f''(0) &= e^0 = 1 & \dots \end{aligned}$$

Then we use this in the Taylor series expansion with $a = 0$

$$f(x) = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!} + \dots = \sum_{n=0}^{\infty} \frac{1}{n!} x^n$$

Similarly, we can work out the Taylor series expansion of the function $f(x) = \cos x$ about $x = 0$.

In this case, we need:

$$\begin{aligned}f(x) &= \cos x, & f(0) &= 1 \\f'(x) &= -\sin x, & f'(0) &= 0 \\f''(x) &= -\cos x, & f''(0) &= -1 \\f'''(x) &= \sin x, & f'''(0) &= 0 \\f^{(4)}(x) &= \cos x, & f^{(4)}(0) &= 1\end{aligned}$$

and we arrive at

$$f(x) = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \cdots = \sum_{k=0}^{\infty} \frac{(-1)^k}{(2k)!} x^{2k}$$

Finally, let us make contact with our discussion of potentials and equilibrium points. We see that the statement (6.5) follows immediately from Taylor's theorem if we take $f = V$, $a = x_0$ and note that $V'(x_0) = 0$ (and we neglect terms beyond quadratic order).

Chapter 7

Collisions

As an application of the conservation of momentum, we will consider a set of problems involving the collision of two bodies. This is a common set-up in Nature, ranging from bar-room billiards to particle collisions in the Large Hadron Collider at CERN. We will discuss the classic Newtonian problem, but also take a look at how the results change if we are dealing with velocities near the speed of light.

In essence, we will make use of the concept of conservation of momentum, but before we do, we need to discuss what happens when momentum is **not** conserved. For this we need to introduce the idea of an **impulse**.

7.1 Impulses

In many situations a body is subjected to a rapidly applied force, e.g., the collision of two cars, a ball hitting the wall or even a punch in the face. In such situations, the velocity of the particle changes rapidly in a short time.



Figure 7.1: Short timescale, large amplitude force.

Consider an event where the force is applied for a brief period of time $t_1 < t < t_2$. According to **N2** the velocity change will involve a momentum change:

$$\vec{F}(t) = \frac{d\vec{p}}{dt}.$$

This momentum change is given by integrating both sides of **N2** to give

$$\vec{I} \equiv \int_{t_1}^{t_2} \vec{F}(t) dt = \vec{p}(t_2) - \vec{p}(t_1).$$

The time integral of \vec{F} is called the **impulse** of the force and is denoted by \vec{I} . The impulse is therefore the change in momentum of the body due to the rapidly applied force.

When two particles collide, in general an impulsive force acts on each one. If the collision is rapid, the effect of forces acting over significantly longer timescales may be insignificant. We can then use the concepts of conservation of momentum and impulse to study the effect of the collision on the particles.

7.2 Conservation of Total Momentum

Consider two bodies that collide:

- Body A with momentum $\vec{p}_A = m_A \vec{v}_A$.
- Body B with momentum $\vec{p}_B = m_B \vec{v}_B$.

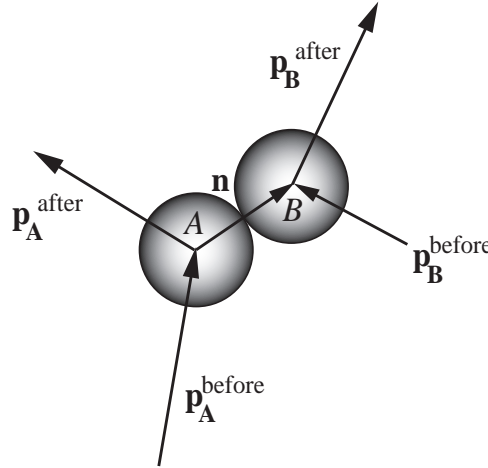


Figure 7.2: When worlds collide.

The collision is rapid and causes an impulse to each body. Assuming no other forces act during this rapid interaction, then by the time-integral of **N2** discussed above

$$\vec{p}_A^{\text{after}} - \vec{p}_A^{\text{before}} = \vec{I}_A, \quad \vec{p}_B^{\text{after}} - \vec{p}_B^{\text{before}} = \vec{I}_B$$

Now by **N3** at the moment of contact the bodies exert equal, but opposite forces on each other, so we must have

$$\vec{I}_A = -\vec{I}_B$$

and moreover these act in the direction of $\pm \vec{n}$, where \vec{n} is a unit vector along the line from the centre of one body to the centre of the other. Thus adding the two momentum equations we have

$$(\vec{p}_A^{\text{after}} - \vec{p}_A^{\text{before}}) + (\vec{p}_B^{\text{after}} - \vec{p}_B^{\text{before}}) = \vec{0}$$

$$\therefore \underbrace{(\vec{p}_A^{\text{after}} + \vec{p}_B^{\text{after}})}_{\text{total momentum after}} = \underbrace{(\vec{p}_A^{\text{before}} + \vec{p}_B^{\text{before}})}_{\text{total momentum before}}$$

The **total** momentum of the system is conserved in the impact.

7.3 Newton's Law of Restitution

Conservation of total momentum alone does not allow us to determine the momentum of each particle separately after the collision: it gives us but one vector equation for two vector unknowns \vec{p}_A^{after} and \vec{p}_B^{after} . We need more information.

The nature of the forces acting between the bodies depends on their material properties. A pair of steel ball bearings will rebound in a different way than say, two cars. To assess the effect of differing material properties on the collision would be (far too) complicated. Fortunately, experiment leads to a simple empirical law.

Newton's Law of Restitution: When two bodies collide, their relative parting velocity in the direction of their line of contact \vec{n} ($|\vec{n}| = 1$), is $-e$ times their relative approaching velocity in the same direction.

$$(\vec{v}_A^{\text{after}} - \vec{v}_B^{\text{after}}) \cdot \vec{n} = -e (\vec{v}_A^{\text{before}} - \vec{v}_B^{\text{before}}) \cdot \vec{n}$$

Note that this law contains the **velocities** and **not** the momenta of the bodies.

The number e is an empirically determined constant that depends on the nature of the particles and encompasses the effect of their material properties on the collision. It is called the **coefficient of restitution**.

- If $e = 1$ the collision is said to be perfectly elastic (a theoretical ideal).
- If $e = 0$, the collision is perfectly inelastic (like throwing mud at a wall).

In practice the coefficient of restitution takes a value in the range $0 < e < 1$. Only in the case of a perfectly elastic collision is the total kinetic energy of the system conserved. In an inelastic collision a proportion of the kinetic energy will be converted into another form of energy, such as heat, light or sound.

7.4 Constancy of velocities perpendicular to line of collision

We still don't have enough information to calculate the velocities of the bodies after the collision. The final ingredient follows from the definition of an impulse: *if there is no friction between the bodies*, no impulsive force will act perpendicular to their line of centres. Hence we deduce that, for frictionless collisions:

- the momentum component of each particle in a direction orthogonal to the line of centres \vec{n} , is conserved ;
- hence if the masses remain unchanged, the velocity of each particle in the direction orthogonal to \vec{n} is unchanged.

Mathematically, we have **conservation of perpendicular velocities**: When the two bodies collide, if the unit vector \vec{t} is perpendicular to the direction of their line of contact \vec{n} , then

$$\vec{v}_A^{\text{after}} \cdot \vec{t} = \vec{v}_A^{\text{before}} \cdot \vec{t}, \quad \vec{v}_B^{\text{after}} \cdot \vec{t} = \vec{v}_B^{\text{before}} \cdot \vec{t}.$$

This is important: It gives us enough information to close the system of equations and solve for the motion of the bodies after the collision.

7.5 Collisions and Non-conservation of Kinetic Energy

As the ideas we have outlined are best illustrated by actual examples, we will consider a series of (increasingly involved) problem settings. It is natural to start with a one-dimensional situation.

Two particles of mass m_1 and m_2 are moving in the same direction with speeds u_1 and u_2 , respectively and collide. Find their subsequent velocities u'_1 and u'_2 if their coefficient of restitution is e .

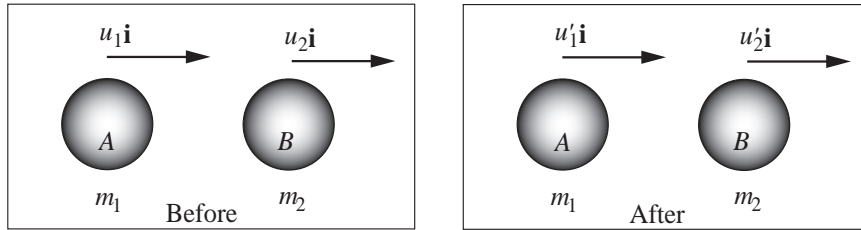


Figure 7.3: Collision parallel to the line of centres.

When the particles collide, their line of contact will be along (say) the $\vec{n} = \hat{i}$ direction and they each experience an equal impulse of magnitude I , but in opposite directions.

- Conservation of momentum for ball 1:

$$(A) \quad m_1 u'_1 \hat{i} = m_1 u_1 \hat{i} + I \hat{i},$$

where the sign of the impulse has been chosen arbitrarily.

- Conservation of momentum for ball 2:

$$(B) \quad m_2 u'_2 \hat{i} = m_2 u_2 \hat{i} - I \hat{i},$$

where the sign of the impulse has been determined to be opposite to the choice in (A).

- Newton's law of restitution acts in the \hat{i} direction is:

$$(C) \quad (u'_2 \hat{i} - u'_1 \hat{i}) \cdot \hat{i} = -e (u_2 \hat{i} - u_1 \hat{i}) \cdot \hat{i}$$

- Since there is no initial motion perpendicular to the line of contact of the bodies, there is no motion perpendicular afterwards. It is enough to consider the components u_1 , u_2 , u'_1 , u'_2 .
- Resolving into components we have three equations for the three unknowns,

$$\begin{aligned} \text{(A)} \quad m_1 u'_1 &= m_1 u_1 + I \\ \text{(B)} \quad m_2 u'_2 &= m_2 u_2 - I \\ \text{(C)} \quad u'_2 - u'_1 &= -e(u_2 - u_1) \end{aligned}$$

- Remove I by considering (A)+(B) \Rightarrow

$$m_1 u'_1 + m_2 u'_2 = m_1 u_1 + m_2 u_2$$

With (C), we now have two equations for the two unknowns u'_1 and u'_2 . These can be solved simultaneously to give:

$$\begin{aligned} u'_1 &= \frac{u_1(m_1 - em_2) + m_2 u_2(1 + e)}{(m_1 + m_2)} \\ u'_2 &= \frac{u_2(m_2 - em_1) + m_1 u_1(1 + e)}{(m_2 + m_1)} \end{aligned}$$

I may be calculated by back-substitution of u'_1 or u'_2 into (A) or (B). This gives

$$I = \frac{m_1 m_2}{m_1 + m_2} (1 + e)(u_2 - u_1).$$

Note that often we are not so much interested in I as the velocities after the collision. Hence we do not always bother to determine it.

7.5.1 Elastic and inelastic collisions

Note that if $e = 1$ and $m_1 = m_2$, the speeds of the particles interchange at the collision, i.e., $u'_1 = u_2$, $u'_2 = u_1$. Calculations of this type underlie the modelling of the executive toy known as “Newton’s Cradle”.

Note also that when the collision is perfectly elastic the kinetic energy is conserved.

The total kinetic energy before collision

$$T_{\text{before}} = \frac{1}{2} m_1 u_1^2 + \frac{1}{2} m_2 u_2^2$$

The total kinetic energy after the collision

$$\begin{aligned}
 T_{\text{after}} &= \frac{1}{2}m_1u_1'^2 + \frac{1}{2}m_2u_2'^2 \\
 &= \frac{1}{2}m_1 \left\{ \frac{u_1(m_1 - m_2) + 2m_2u_2}{(m_1 + m_2)} \right\}^2 + \frac{1}{2}m_2 \left\{ \frac{u_2(m_2 - m_1) + 2m_1u_1}{(m_2 + m_1)} \right\}^2 \\
 &= \dots \\
 &= \frac{1}{2}m_1u_1^2 + \frac{1}{2}m_2u_2^2
 \end{aligned}$$

If $e \neq 1$ then the total kinetic energy is **not** conserved. Where does this energy go? The **total** energy **has** been conserved, but some of the kinetic energy has been converted into other forms of energy. This could be (for example) heat, acoustic, or even light energy. The mechanical problem can predict the amount of the kinetic energy that is transformed, but not the form the energy takes.

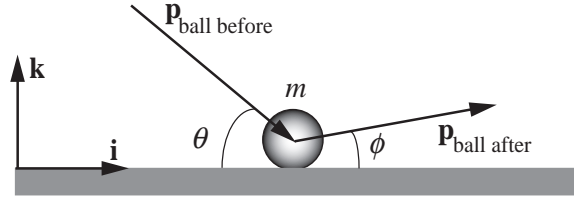


Figure 7.4: Ball bouncing off a wall.

Example 21: Bouncing ball

A ball of mass m hits the ground. The ball is travelling at speed u at an angle θ to the ground, as in figure 7.4. If the coefficient of restitution of impact is e , (and there is no spin on the ball) find the velocity after the impact.

Before impact:

$$\vec{p}_{\text{ball before}} = m(u \cos \theta \hat{i} - u \sin \theta \hat{k})$$

At impact: line of centres is along \hat{k} , hence the impulse takes the form

$$\vec{I} = I \hat{k}$$

After impact: assume that ground does not move,

$$\vec{p}_{\text{ball after}} = m(u' \cos \phi \hat{i} + u' \sin \phi \hat{k})$$

Conservation of Momentum for ball:

$$m(u' \cos \phi \hat{i} + u' \sin \phi \hat{k}) = m(u \cos \theta \hat{i} - u \sin \theta \hat{k}) + I \hat{k}$$

Now apply Newton's law of restitution:

$$\hat{k} \cdot [\vec{v}_{\text{ball, after}} - \vec{v}_{\text{ground, after}}] = -e \hat{k} \cdot [\vec{v}_{\text{ball, before}} - \vec{v}_{\text{ground, before}}]$$

Assume that ground does not move, so that $\Rightarrow \vec{v}_{\text{ground, before}} = \vec{v}_{\text{ground, after}} = \vec{0}$,

$$(u' \cos \phi \hat{i} + u' \sin \phi \hat{k} - \vec{0}) \cdot \hat{k} = -e(u \cos \theta \hat{i} - u \sin \theta \hat{k} - \vec{0}) \cdot \hat{k}$$

Resolving the conservation of momentum equations into \hat{i} and \hat{k} components these give 3 equations for the three unknowns, u' , ϕ and I ,

$$(A) \quad mu' \cos \phi = mu \cos \theta$$

$$(B) \quad mu' \sin \phi = -mu \sin \theta + I$$

$$(C) \quad u' \sin \phi = eu \sin \theta$$

$$(C)/(A) \Rightarrow$$

$$\tan \phi = e \tan \theta$$

$$(A)^2 + (C)^2 \Rightarrow$$

$$\begin{aligned} u'^2(\cos^2 \phi + \sin^2 \phi) &= u^2(\cos^2 \theta + e^2 \sin^2 \theta) \\ \Rightarrow u' &= u \sqrt{1 - (1 - e^2) \sin^2 \theta} \end{aligned}$$

If we require it, (B) can then be used to determine the impulse as

$$I = me u \sin \theta + mu \sin \theta = (1 + e)mu \sin \theta$$

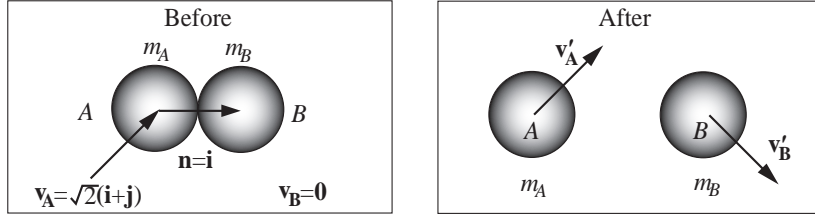


Figure 7.5: Oblique impact. (Two billiards are said to collide *obliquely*, if their initial velocities point in directions different to their line of centres at collision.)

Example 22: Oblique collision

Consider two colliding billiard balls A and B of equal mass m whose line of centres lie in the direction \hat{i} at the point of collision, see figure 7.5. Let the velocities before the collision be $\vec{v}_A = \sqrt{2}(\hat{i} + \hat{j})$ and $\vec{v}_B = \vec{0}$. The coefficient of restitution between the balls is e . Find the velocities of the balls after collision.

We shall use the following notation for the post-collision velocities for A and B , respectively:

$$\vec{v}'_A = a\hat{i} + b\hat{j}, \quad \vec{v}'_B = \alpha\hat{i} + \beta\hat{j}.$$

Conservation of momentum:

$$\begin{aligned} m_A \vec{v}_A + m_B \vec{v}_B &= m_A \vec{v}'_A + m_B \vec{v}'_B \\ \Rightarrow \vec{v}_A + \vec{0} &= \vec{v}'_A + \vec{v}'_B \\ \Rightarrow \sqrt{2}(\hat{i} + \hat{j}) &= a\hat{i} + b\hat{j} + \alpha\hat{i} + \beta\hat{j} \\ \Rightarrow \sqrt{2} &= a + \alpha \quad \text{and} \quad \sqrt{2} = b + \beta. \end{aligned}$$

Newton's law of restitution along the line of centres $\vec{n} = \hat{i}$:

$$\begin{aligned} (\vec{v}'_A - \vec{v}'_B) \cdot \hat{i} &= -e(\vec{v}_A - \vec{v}_B) \cdot \hat{i} \\ \left(\{a\hat{i} + b\hat{j}\} - \{\alpha\hat{i} + \beta\hat{j}\} \right) \cdot \hat{i} &= -e \left(\sqrt{2} \{ \hat{i} + \hat{j} \} - \vec{0} \right) \cdot \hat{i} \\ a - \alpha &= -\sqrt{2}e. \end{aligned}$$

Conservation of (momentum) perpendicular to line of motion $\vec{t} = \hat{j}$. The mass of either particle does not change (classically) as a result of the collision, so conservation of momentum perpendicular to the line of centres just becomes conservation of velocities:

$$\begin{aligned} \vec{v}'_A \cdot \hat{j} = \vec{v}_A \cdot \hat{j} &\Rightarrow b = \sqrt{2}, \\ \vec{v}'_B \cdot \hat{j} = \vec{v}_B \cdot \hat{j} &\Rightarrow \beta = 0. \end{aligned}$$

Combining the momentum and restitution equations and simultaneously solving for a and α , we have:

$$\begin{aligned}\sqrt{2} &= a + \alpha \\ -e\sqrt{2} &= a - \alpha \\ \Rightarrow a &= \frac{1}{\sqrt{2}}(1 - e) \quad \alpha = \frac{1}{\sqrt{2}}(1 + e).\end{aligned}$$

Hence drawing everything together, we have

$$\begin{aligned}\vec{v}'_A &= \frac{1}{\sqrt{2}}(1 - e)\hat{i} + \sqrt{2}\hat{j}, \\ \vec{v}'_B &= \frac{1}{\sqrt{2}}(1 + e)\hat{i}.\end{aligned}$$

Note that the originally stationary ball B moves off along the line of centres at impact. If you play snooker you may already know that one aims to hit the coloured ball with the cue ball so that the line of centres points towards the pocket. This is a mathematical explanation of this phenomenon.

This is, of course, an idealised model. In practice, friction between balls and table may affect the results.

7.6 Centre of mass

Let us now consider what happens if the impact takes place at high velocities. In order to do this it is useful (as it simplifies the mathematics) to focus on elastic collisions¹. In general all velocities are vectors, but in order to keep the discussion simple we will consider motion in one dimension, in which case the velocities can be treated as scalars.

Suppose we know the velocities of two particle before they collide, v_1 and v_2 . These are measured in the “lab frame”. If the two masses are m_1 and m_2 then the corresponding total (kinetic) energy is

$$T = T_1 + T_2 = \frac{1}{2}m_1v_1^2 + \frac{1}{2}m_2v_2^2$$

However, in many situations it is convenient to consider the problem in a different coordinate system, e.g. one that is moving along with the centre of mass. In this system, the two velocities are u_1 and u_2 (say) and, if the centre of mass moves with velocity V , we have the simple transformations

$$u_1 = v_1 - V$$

$$u_2 = v_2 - V$$

because (as we know from our everyday experience) velocities add (at least in Newtonian dynamics...).

¹We are being lazy. It would be straightforward to extend the discussion to the general case by accounting for inelastic change in the energy of the collision products, but we would not gain much additional understanding by doing this.

We determine V by demanding that the total momentum vanishes in the centre of mass frame. With the total mass $M = m_1 + m_2$ we then have

$$MV = m_1v_1 + m_2v_2 \longrightarrow V = \frac{m_1v_1 + m_2v_2}{M}$$

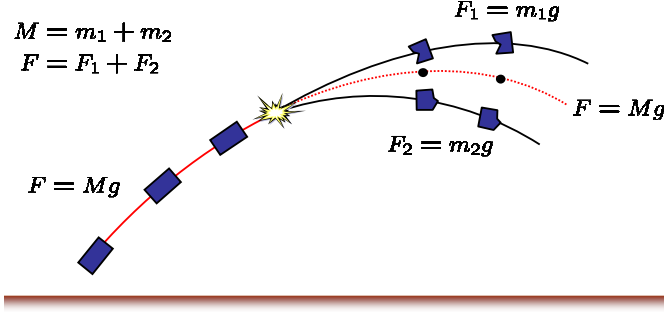


Figure 7.6: The motion of the centre-of-mass remains the same as if the object had stayed in one piece.

We can also assign a kinetic energy associated with the centre of mass:

$$T_{\text{cm}} = \frac{1}{2}MV^2$$

For an elastic collision the energy is conserved. In essence, this energy is not available for the collision, but... in general we have

$$T = T_{\text{cm}} + T_{\text{rel}}$$

where the remaining energy

$$T_{\text{rel}} = \frac{1}{2}\mu(v_2 - v_1)^2$$

is available for the collision. The quantity μ is called the reduced mass, and it is given by

$$\mu = \frac{m_1m_2}{M}$$

We now have all the quantities we need to consider a collision in the centre of mass frame. Letting primed velocities represent the velocity of the particles after the collision, we have momentum conservation

$$m_1u_1 + m_2u_2 = m_1u'_1 + m_2u'_2$$

and energy conservation

$$\frac{1}{2}m_1u_1^2 + \frac{1}{2}m_2u_2^2 = \frac{1}{2}m_1(u'_1)^2 + \frac{1}{2}m_2(u'_2)^2$$

We have two equations for two unknowns: u'_1 and u'_2 . We also know that in the centre of mass frame the total momentum is zero. So we actually have

$$m_1u_1 + m_2u_2 = 0 \longrightarrow u_1 = -\frac{m_2}{m_1}u_2$$

and as the same is true after the collision;

$$m_1 u'_1 + m_2 u'_2 = 0 \longrightarrow u'_1 = -\frac{m_2}{m_1} u'_2$$

Combining these results with the conservation of energy (7.6) we see that we must have

$$u'_2 = \pm u_2 \quad \text{and} \quad u_1 = \pm u_1$$

For an elastic collision (and in the centre of mass frame), there are only two options. Either the particles emerge with the same velocity as before (actually, a miss!) or the directions are reversed but the speeds stays the same.

Chapter 8

Rockets: Motion with variable mass

So far we have confined ourselves to studying the mechanics of bodies with constant mass. Let us now turn to an important variable mass problem, that of rocket motion (for which we no longer have $\vec{F} = m\vec{a}$). Although the mechanics is discussed in the context of a rocket, it could equally be applied to the acceleration of a jet fighter, the ascent of a balloon or even falling raindrops on a misty day.

We consider a rocket of mass $m(t)$ travelling at velocity $\vec{v}(t)$ in space (i.e., free from gravity and air resistance) by emitting a jet of gas at a speed $u > 0$ relative to the rocket. At an infinitesimally later time $t + dt$ suppose that the mass of the rocket is $m + dm$ due to the expulsion of a mass $-dm$ of gas through the back of the rocket.¹

We want to use conservation of momentum to study this situation. If we denote the fixed inertial frame by S , the rocket frame by R , and the exhaust frame by E , then

$$\vec{v}_{ES} = \vec{v}_{ER} + \vec{v}_{RS}$$

We therefore have

$$\vec{v}_{ES} = -u\vec{n} + \vec{v}$$

where \vec{n} is a unit vector pointing in whatever direction the rocket points.

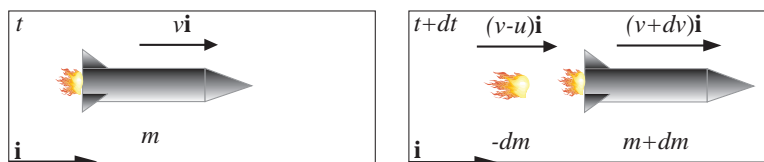


Figure 8.1: Rocket motion by ejection of gases.

¹How can you have a negative mass of gas you ask? Well, you don't really. Note that $-dm$ is a positive quantity if dm is assumed negative. This is the convention that the most people have adopted!

At time t :

$$\text{Momentum of gas : } \vec{0} \quad \text{Momentum of rocket : } m\vec{v}$$

At time $t + dt$:

$$\text{Momentum of gas : } -dm(-u\vec{n} + \vec{v}) \quad \text{Momentum of rocket : } (m+dm)(\vec{v} + d\vec{v})$$

Conservation of total momentum gives:

$$m\vec{v} = (m + dm)(\vec{v} + d\vec{v}) + (-dm)(-u\vec{n} + \vec{v})$$

Expanding:

$$m\vec{v} = m\vec{v} + \vec{v}dm + md\vec{v} + u\vec{n}dm - \vec{v}dm$$

where terms beyond linear have been neglected. This can be rearranged (and turned into a differential equation by recalling that the change was associated with a time interval dt)

$$\begin{aligned} m \frac{d\vec{v}}{dt} &= -u\vec{n} \frac{dm}{dt} \\ \Rightarrow m \frac{d\vec{v}}{dt} + \vec{v} \frac{dm}{dt} &= -u\vec{n} \frac{dm}{dt} + \vec{v} \frac{dm}{dt} \end{aligned}$$

which has the interpretation:

$$\underbrace{\frac{d}{dt}(m\vec{v})}_{\text{rate of change of rocket momentum}} = \underbrace{(\vec{v} - u\vec{n}) \frac{dm}{dt}}_{\text{force due to rocket motion}} = -\vec{v}_{ES} \frac{dm}{dt}$$

By **N2** we can generalise this to the full vector form of the equation:

$$\frac{d}{dt}(m\vec{v}) = (\vec{v} - u\vec{n}) \frac{dm}{dt} + \vec{F}$$

where \vec{F} is the sum of the other forces acting on the rocket, e.g., air resistance, gravity etc. Note that in the above derivation the magnitude u may also be a function of time.

Example 23: Rocket speed

Consider a rocket under gaseous propulsion in deep space, free from all external forces, travelling along the x -axis. If the initial speed of the rocket is $\vec{v}_0 = v_0\hat{i}$, its mass is m_0 and the propulsive gases are emitted at a constant speed relative to the rocket, calculate its speed as a function of mass.

Using the above calculation (with $\vec{n} = \hat{i}$) we obtain the differential equation

$$m \frac{dv}{dt} = -u \frac{dm}{dt}$$

This can be converted to a separable equation by dividing through by one of the derivatives

$$m \frac{dv}{dt} \bigg/ \frac{dm}{dt} = -u \longrightarrow m \frac{dv}{dm} = -u$$

$$\int dv = -u \int \frac{dm}{m} \longrightarrow v - v_0 = u \ln \left(\frac{m_0}{m} \right).$$

Thus the speed depends only on the ratio of initial to current masses (> 1). Hence to get the rocket to travel quickly we need $m_0/m \gg 1$, i.e., to reduce the mass of the rocket by as much as possible. That's (partly) why you have multistage rockets. Once a particular fuel tank is empty, it is jettisoned to reduce the mass as much as possible. If we want to know the speed at a particular time, we need to specify the rate of burn of rocket fuel, dm/dt , which is often a constant for most of the burn time. Suppose the rate of fuel burn for the above rocket is constant rate $-\alpha$. (Remember dm was negative!)

$$\frac{dm}{dt} = -\alpha \Rightarrow m(t) = m_0 - \alpha t$$

Hence we have

$$v = v_0 + u \ln \left(\frac{m_0}{m_0 - \alpha t} \right).$$

Note that we must have $m_0 - \alpha t > 0$, or else it will most certainly be out of fuel!

Example 24: Rocket burn under gravity

Suppose the above rocket is ejecting gas at a constant rate (α) and at a constant velocity relative to it but is moving vertically upward in a constant gravitational field $\vec{g} = -g\hat{k}$ (assuming that it won't go very high...). Calculate its speed as a function of time.

We use the full expression for **N2**

$$\frac{d}{dt}(m\vec{v}) = (\vec{v} - u\vec{n}) \frac{dm}{dt} + \vec{F}$$

where

$$\vec{F} = -m(t)g\hat{k}, \quad \vec{v} = v(t)\hat{k}, \quad \vec{n} = \hat{k}, \quad u = \text{constant}, \quad \frac{dm}{dt} = -\alpha.$$

The mass of the rocket is

$$\frac{dm}{dt} = -\alpha \Rightarrow m(t) = m_0 - \alpha t.$$

Inserting this into the full equation and noting that only the \hat{k} component is relevant we have

$$m \frac{d\vec{v}}{dt} = -u\vec{n} \frac{dm}{dt} + \vec{F}$$

If the rocket starts at $t = 0$ with speed $v(0) = 0$, then

$$0 = -u \ln(m_0) + \text{constant} \Rightarrow \text{constant} = u \ln(m_0)$$

and we then have

$$v = u \ln \left(\frac{m_0}{m_0 - \alpha t} \right) - gt.$$

The effect of gravity is to retard the speed of the rocket by an amount gt . Integrating once more with respect to time will give the position of the rocket.

Example 25: A rocket with linear drag

Here's a slightly more involved example. The initial mass of the rocket is m_0 and we still burn fuel at a constant rate, so $\dot{m} = -\alpha$. But now the rocket is subject to linear drag, $F = -\gamma v$, presumably because it has encountered some sticky intergalactic sludge or something. If the rocket starts from rest, how fast is it going after it has burned one half of its mass as fuel? With linear drag, the rocket equation becomes

$$m\dot{v} + u\dot{m} = -\gamma v$$

We can get a feel for what's going on by looking at this equation. Since $\dot{m} = -\alpha$, rearranging we get

$$m\dot{v} = \alpha u - \gamma v$$

This means that we will continue to accelerate through the sticky stuff if we are travelling slowly and burning fuel fast enough so that $\alpha u > \gamma v$, but as our speed approaches $v = \alpha u / \gamma$ the acceleration slows down and we expect this to be the limiting velocity. However, if we were travelling too fast to begin with, so $\gamma v > \alpha u$, then we will slow down until we again hit the (same) limiting speed. These expectations are brought out by the detailed solution (which follows from solving a separable equation).

Chapter 9

Systems, solid bodies and moments

So far, we have only considered the motion of a single body. If our goal is to understand everything in the Universe, this is a bit limiting. So let us take a small step forwards and consider the dynamics of N interacting particles.

9.1 Many particles

The first thing that we do is put a label $i = 1, \dots, N$ on everything¹. The i^{th} particle has mass m_i , position \vec{r}_i and momentum $\vec{p}_i = m_i \dot{\vec{r}}_i$.

Now **N2** states that, for each particle, $i = 1, 2, \dots, N$,

$$m_i \ddot{\vec{r}}_i(t) = \dot{\vec{p}}_i = \vec{F}_i.$$

where \vec{F}_i is the force acting on the i^{th} particle. The new thing is that, for a system of particles the force \vec{F}_i can be split into two parts: an external force \vec{F}^{ext} (for example, if the whole system sits in a gravitational field) and a force due to the presence of the other particles. We write

$$\vec{F}_i = \vec{F}_i^{\text{ext}} + \sum_{j \neq i}^N \vec{F}_{ij}$$

where \vec{F}_{ij} is the force on particle i due to the presence of particle j . This allows us to provide a more precise definition of Newton's third law. Recall that *every reaction has an equal and opposite reaction*. This means,

$$\vec{F}_{ij} = -\vec{F}_{ji}$$

We can (obviously) add all N equations to get

$$\sum_{i=1}^N m_i \ddot{\vec{r}}_i(t) = \sum_{i=1}^N \vec{F}_i.$$

¹In a sense, this is the first step towards the index notation you will later see for vectors and tensors (especially in relativity). However, then we will be labelling vector components, while we are now labelling particles.

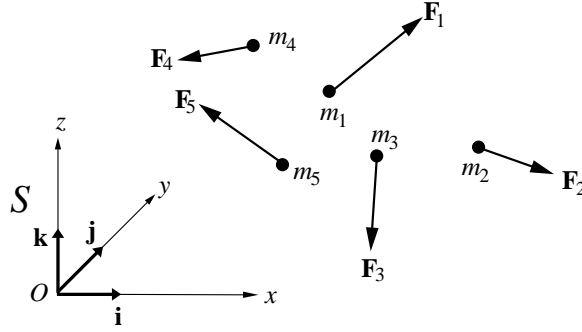


Figure 9.1: They're absolutely everywhere!

This is the equation of motion for the combined system of particles. However, it is more instructive to express the collective motion in terms of the **centre of mass**. Thus, we introduce the total mass:

$$M = \sum_{i=1}^N m_i$$

and the centre of mass (generalising the previous definition in the natural way)

$$\vec{R}(t) \equiv \frac{1}{M} \sum_{i=1}^N m_i \vec{r}_i(t)$$

The total momentum of the system, \vec{P} , can then be written entirely in terms of the centre of mass motion,

$$\vec{P} = \sum_{i=1}^N \vec{p}_i = M \dot{\vec{R}}$$

Let us now ask how the centre of mass moves. We have

$$\dot{\vec{P}} = \sum_{i=1}^N \dot{\vec{p}}_i = \sum_{i=1}^N \left(\vec{F}_i^{\text{ext}} + \sum_{j \neq i}^N \vec{F}_{ij} \right) = \sum_{i=1}^N \vec{F}_i^{\text{ext}} + \sum_{i < j} \left(\vec{F}_{ij} + \vec{F}_{ji} \right)$$

where the last term vanishes because of **N3**. That is, we are left with

$$\dot{\vec{P}} = \sum_{i=1}^N \vec{F}_i^{\text{ext}}$$

This is an important result. It tells us that if we just want to know the motion of the centre of mass of a system of particles, then **only the external forces count**. If you throw a wriggling, squealing cat then its internal forces \vec{F}_{ij} can change its orientation, but they can do nothing to change the path of its centre of mass. That is dictated by gravity alone.

It is hard to overstate the importance of this result. Without it, the whole Newtonian framework for mechanics would come crashing down. After all, nothing that we really describe is truly a point particle. Certainly not a planet or a cat, but even something as simple as an electron has an internal spin. Yet none of these details matter because everything, regardless of the details, acts as a point particle if we just focus on the position of its centre of mass.

The centre of mass can be regarded as the **average position** of the mass. Sometimes it is convenient to consider an extended body many up of a collection of particles as but one particle with the same total mass as the system, but located at the centre of mass². The idea is that the system of particles acts dynamically as if all the mass were located at the centre of mass. This is actually the idea underlying the abstraction of extended bodies into point particles that we have already been using. The centre of mass is sometimes called the *centre of gravity*³, and represents the point through which gravity acts in a uniform gravitational field.

For future reference, it is also worth noting that if the system of particles is static, then $\ddot{\vec{R}}(t) = \mathbf{0}$ and we must have

$$\sum_{i=1}^N \vec{F}_i = 0.$$

This is a necessary (but not sufficient) condition for the static equilibrium for a system of particles. The system may be moving with a constant velocity, but we can always make the situation stationary by viewing from an inertial frame that moves at the same velocity as the centre of mass. This is (again) the *centre of mass frame*.

9.2 Moments

If two forces of equal magnitude act in opposite directions, then the sum of the forces vanishes. However, if these forces do not act at the same point on a body, they exert a twist, or **torque** or **moment** about an axis.

This torque will lead to a rotational acceleration of the body. Experimentally it is found that if the force is applied at a greater distance from the axis of rotation, then the torque/moment will be greater. In fact the moment can be determined by experiment as follows.

Consider particle i of constant mass m_i and suppose we decide to take the moments about the origin of the coordinate system \mathcal{O} and a (resultant) force \vec{F}_i is acting on a particle located at $\vec{r}_i(t)$. The moment \vec{M}_i of the force about \mathcal{O} is the vector product

$$\vec{M}_i(t) \equiv \vec{r}_i(t) \times \vec{F}_i(t).$$

²If the distribution of masses is continuous, the individual masses of particles in the definition are replaced with a local mass density and the sums are replaced with integrals over the volume of the body.

³Note that the centre of mass is only equivalent to the centre of gravity in a uniform gravitational field.

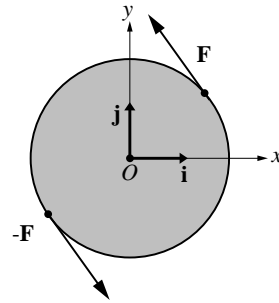


Figure 9.2: Zero-sum forces on a disc implies no translational acceleration, but there is rotational acceleration about the origin.

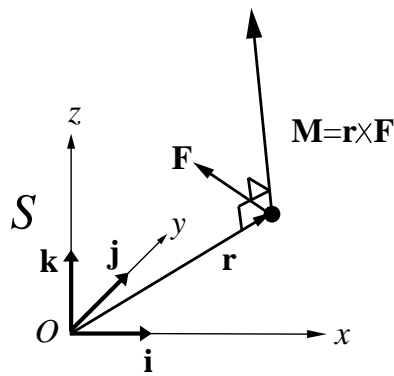


Figure 9.3: Definition of moment.

Combining the vector moment with **N2** for particle i , we obtain⁴

$$\vec{M}_i(t) \equiv \vec{r}_i(t) \times \vec{F}_i(t) = \vec{r}_i(t) \times m_i \ddot{\vec{r}}_i(t) = m_i \vec{r}_i(t) \times \ddot{\vec{r}}_i(t)$$

If you have studied mechanics before you can now see why the magnitude of a moment is said to be “force times *perpendicular* distance from the axis”. The vector product ensures that only the component of the force normal to $\vec{r}_i(t)$ appears in the magnitude of the component. The definition of the vector product combined with a simple sketch shows that in magnitude, these two statements are equivalent.

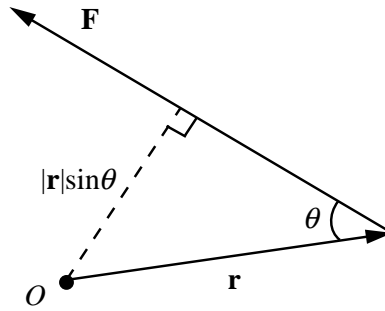


Figure 9.4: Magnitude of moment.

$$\vec{M} = \vec{r} \times \vec{F} \Rightarrow |\vec{M}| = |\vec{r}| \cdot |\vec{F}| \sin \theta \Rightarrow |\vec{M}| = \underbrace{|\vec{r}| \sin \theta}_{\perp \text{ distance}} \cdot |\vec{F}|$$

As with forces for the system of N particles, we may add the moments of individual particles about \mathcal{O} to give the total moment

$$\vec{M} = \sum_{i=1}^N \vec{M}_i = \sum_{i=1}^N \vec{r}_i(t) \times \vec{F}_i(t)$$

Note that if the body is in equilibrium, each individual particle has no acceleration $\ddot{\vec{r}}_i(t) = \mathbf{0}$ and we obtain

$$\vec{M} = \sum_{i=1}^N \vec{r}_i(t) \times \vec{F}_i(t) = \mathbf{0}$$

That is, a system in equilibrium does not experience a torque.

9.3 Angular Momentum and Torque

The angular momentum of a particle of mass m at position $\vec{r}(t)$ is defined to be

$$\vec{L} \equiv m(\vec{r} \times \dot{\vec{r}}) = \vec{r} \times \vec{p}$$

⁴Note that the moment is a vector quantity and that the \vec{r} comes before the \vec{F} (or the $\ddot{\vec{r}}_i$) in the vector product.

With this definition we see that

$$\frac{d\vec{L}}{dt} = \frac{d(\vec{r} \times m\dot{\vec{r}})}{dt} = (\dot{\vec{r}} \times m\dot{\vec{r}}) + \left(\vec{r} \times \frac{d(m\dot{\vec{r}})}{dt} \right) = \mathbf{0} + \left(\vec{r} \times \frac{d\vec{p}}{dt} \right)$$

From **N2**, the rate of change of momentum is nothing other than the force \vec{F} on the body. Hence we have

$$\frac{d\vec{L}}{dt} = (\vec{r} \times \vec{F}) = \vec{M}$$

which, in words, states that the moment of a force, or torque, about the origin O of a coordinate system is equal to the rate of change of the angular momentum. This is the rotational analogue of the translational form of Newton's second law:

$$\frac{d\vec{P}}{dt} = \vec{F}$$

We will return to the angular momentum later.

9.4 Energy

It is also useful to consider the energy of a system of particles. The total kinetic energy is

$$T = \frac{1}{2} \sum_i m_i \dot{\vec{r}}_i \cdot \dot{\vec{r}}_i$$

Suppose we consider the motion relative to the centre of mass, such that

$$\vec{r}_i = \vec{R} + \vec{y}_i$$

then

$$\sum_i m_i \vec{r}_i = M \vec{R} \Rightarrow \sum_i m_i \vec{y}_i = 0$$

This means that the kinetic energy can be written

$$\begin{aligned} T &= \frac{1}{2} \sum_i m_i (\dot{\vec{R}} + \dot{\vec{y}}_i)^2 \\ &= \frac{1}{2} \sum_i m_i \dot{\vec{R}}^2 + \dot{\vec{R}} \cdot \underbrace{\sum_i m_i \dot{\vec{y}}_i}_{=0} + \frac{1}{2} \sum_i m_i \dot{\vec{y}}_i^2 \\ &= \frac{1}{2} M \dot{\vec{R}}^2 + \frac{1}{2} \sum_i m_i \dot{\vec{y}}_i^2 \end{aligned}$$

Basically, the kinetic energy splits up into the kinetic energy of the centre of mass, together with the kinetic energy of the particles moving around the centre of mass.

9.5 Rigid bodies

So far, we have only discussed “particles”, objects with no extended size. But what happens to more complicated objects that can twist and turn as they move? The simplest example is a rigid body. This is a collection of N particles, constrained so that the relative distance between any two points, i and j , is fixed:

$$|\vec{r}_i - \vec{r}_j| = \text{fixed}$$

A rigid body can undergo only two types of motion: its centre of mass can move and it can rotate. We will focus on the rotation.

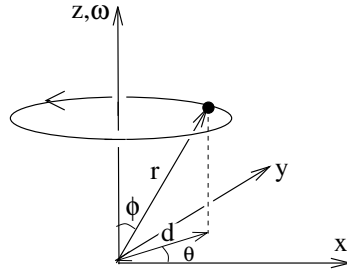


Figure 9.5: A rotating particle.

Suppose we fix some point in the rigid body and consider rotation about this point. To describe these rotations, we need the concept of angular velocity. We begin by considering a single particle which is rotating around the z -axis. The position and velocity of the particle are given by (in cylindrical coordinates)

$$\vec{r} = (d \cos \theta, d \sin \theta, z) \Rightarrow \dot{\vec{r}} = (-d\dot{\theta} \sin \theta, d\dot{\theta} \cos \theta, 0)$$

We can write this by introducing a new vector $\vec{\omega} = \dot{\theta} \hat{e}_z$:

$$\dot{\vec{r}} = \vec{\omega} \times \vec{r}$$

The vector $\vec{\omega}$ is called the angular velocity. In general we can write $\vec{\omega} = \omega \hat{n}$. Here the magnitude, $\omega = \dot{\theta}$, is the angular speed of rotation, while the unit vector \hat{n} points along the axis of rotation, defined in a right-handed sense⁵.

The speed of the particle is then given by

$$v = |\dot{\vec{r}}| = r\omega \sin \phi = d\omega$$

where

$$d = |\hat{n} \times \vec{r}| = r \sin \phi$$

is the perpendicular distance to the axis of rotation. Finally, we will also need an expression for the kinetic energy of this particle as it rotates about the \hat{n} axis through the origin; it is

$$T = \frac{1}{2} m \dot{\vec{r}} \cdot \dot{\vec{r}} = \frac{1}{2} m (\vec{\omega} \times \vec{r}) \cdot (\vec{\omega} \times \vec{r}) = \frac{1}{2} m d^2 \omega^2$$

⁵Curl the fingers of your right hand in the direction of rotation: your thumb points in the direction of $\vec{\omega}$.

9.6 The moment of inertia

Let us now consider a system with N particles which make up a rigid body. The fact that the body is rigid mean that all particles rotate with the same velocity:

$$\dot{\vec{r}}_i = \vec{\omega} \times \vec{r}_i$$

This ensure that the relative distance between the particles remains fixed

$$\begin{aligned} \frac{d}{dt} |\vec{r}_i - \vec{r}_j|^2 &= 2 (\dot{\vec{r}}_i - \dot{\vec{r}}_j) \cdot (\vec{r}_i - \vec{r}_j) \\ &= 2 [\vec{\omega} \times (\vec{r}_i - \vec{r}_j)] \cdot (\vec{r}_i - \vec{r}_j) = 0 \end{aligned}$$

We can now write the kinetic energy of a rigid body as

$$T = \frac{1}{2} \sum_i m_i \dot{\vec{r}}_i \cdot \dot{\vec{r}}_i = \frac{1}{2} I \omega^2$$

where we have introduced

$$I = \sum_{i=1}^N m_i d_i^2$$

where d_i is the distance of each particle away from the rotational axis. This is the *moment of inertia*. It is worth noting the similarity between the rotational kinetic energy and the translational kinetic energy ($Mv^2/2$) – the moment of inertia is to rotation what the mass is to translation.

The moment of inertia also plays a role in the angular momentum. We have

$$\vec{L} = \sum_i m_i \vec{r}_i \times \dot{\vec{r}}_i = \sum_i m_i \vec{r}_i \times (\vec{\omega} \times \vec{r}_i)$$

Introducing $\vec{\omega} = \omega \hat{n}$ (with \hat{n} a unit vector along the rotational axis), we have

$$\begin{aligned} \vec{L} \cdot \hat{n} &= \omega \sum_i m_i [\vec{r}_i \times (\hat{n} \times \vec{r}_i)] \cdot \hat{n} \\ &= \omega \sum_i m_i (\vec{r}_i \times \hat{n}) \cdot (\vec{r}_i \times \hat{n}) = I \omega \end{aligned}$$

We already know that acting with a torque $\vec{\tau}$ changes the angular momentum:

$$\dot{\vec{L}} = \vec{\tau}$$

For a rigid body, we see that if the torque is in the direction of the angular momentum (so that $\vec{\tau} = \tau \hat{n}$) we have

$$I \dot{\omega} = \tau$$

In practice, we calculate the moment of inertia by replacing the individual particles with a continuous density distribution $\rho(\vec{r})$. The total mass is given by the volume integral

$$M = \int \rho(\vec{r}) dV$$

and the moment of inertia follows as

$$I = \int \rho(\vec{r}) x_{\perp}^2 dV$$

where x_{\perp} is the perpendicular distance from the point \vec{r} to the rotation axis. For example, for a uniform density sphere (with radius a) we would have

$$I = \frac{2}{5} M a^2$$

9.7 Why the two-body problem is really a one-body problem

Solving the general problem of the dynamics of N mutually interacting particles is (very) hard. However, when there are no external forces present, the case of two particles (labelled 1 and 2) actually reduces to the kind of one particle problem we are familiar with.

We have already defined the centre of mass

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

Let us also introduce the relative separation

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

which allows us to write

$$\vec{r}_1 = \vec{R} + \frac{m_2}{M} \vec{r} \quad \text{and} \quad \vec{r}_2 = \vec{R} - \frac{m_1}{M} \vec{r}$$

As long as no external forces act on the system, we have $\ddot{\vec{R}} = 0$. Meanwhile, the relative motion is dictated by

$$\ddot{\vec{r}} = \ddot{\vec{r}}_1 - \ddot{\vec{r}}_2 = \frac{1}{m_1} \vec{F}_{12} - \frac{1}{m_2} \vec{F}_{21} = \frac{m_1 + m_2}{m_1 m_2} \vec{F}_{12}$$

after using **N3**. The equation of motion can now be written

$$\mu \ddot{\vec{r}} = \vec{F}_{12}$$

where μ is the *reduced mass*

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

This is nice. It means that we have already solved the problem of two mutually interacting particles because their centre of mass motion is trivial, while their relative separation reduces to the kind of problem we have already seen.

In the limit where one of the particles involved is very heavy, say $m_2 \gg m_1$, then $\mu \approx m_1$ and the heavy object remains essentially fixed, with the lighter object orbiting around it. For example, the centre of mass of the Earth and Sun is very close to the centre of the Sun. Even for the Earth and moon, the centre of mass is 1000 miles below the surface of the Earth.

Chapter 10

Central Forces

So far our main focus has been on motion that can be described in terms of a fixed Cartesian coordinate system. This takes us quite far, but it is by no means the only way we can describe a problem. Many problems have particular symmetries and dynamical problems simplify if we make use of this. To do this, we need to introduce other sets of coordinates – and these may not be fixed. For example, we can decide to use a coordinate system that is fixed in a rotating body. Then the coordinate basis vectors will vary in time. As an example of this – and an excuse to develop the required mathematics - we will discuss the motion of particles under central forces. These are forces that act towards a central point, for example the sun or an atom.

10.1 Cylindrical Polar Coordinates

When a 3D problem has a cylindrical symmetry (a rotational symmetry about a single axis), such as the magnetic field around a wire, or the flow of fluid through a circular pipe, it makes sense to use **cylindrical polar coordinates** (r, θ, z) , to specify the position of an object. In such situations the equations governing the system often simplify.

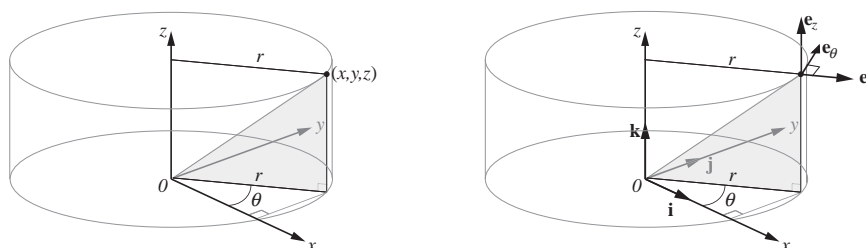


Figure 10.1: Cylindrical polar coordinate system and basis vectors

Effectively these are 2D polar coordinates (r, θ) in the (x, y) plane perpendicular to the symmetry axis, together with the usual Cartesian z -coordinate along the symmetry axis.

The transformation from cylindrical polar coordinates to Cartesian $(r, \theta, z) \rightarrow (x, y, z)$ is:

$$x = r \cos \theta, \quad y = r \sin \theta, \quad z = z.$$

while the inverse transformation follows from plane polar coordinates

$$r = \sqrt{x^2 + y^2},$$

with θ being the arctan of $\pm y/x$, depending on the quadrant of θ .

For motion confined to a 2D plane $z = \text{constant}$, it suffices to think of these coordinates as the plane 2D polar coordinates.

Instead of the unit Cartesian basis vectors $\hat{i}, \hat{j}, \hat{k}$ the unit basis vectors in cylindrical polars are:

- \hat{e}_r is in the outwards radial direction from the origin in the plane $z = 0$.
- \hat{e}_θ is in the azimuthal (increasing positive θ) direction $\hat{e}_r \times \hat{e}_z$, where
- \hat{e}_z is in the Cartesian direction \hat{k} (i.e., $\hat{e}_z \equiv \hat{k}$).

The relationship between cylindrical polar basis vectors at the point (r, θ, z) and Cartesian basis vectors is given by geometry as

$$\begin{aligned} \hat{e}_r &= \cos \theta \hat{i} + \sin \theta \hat{j} \\ \hat{e}_\theta &= -\sin \theta \hat{i} + \cos \theta \hat{j} \\ \hat{e}_z &= \hat{k} \end{aligned}$$

In general, the position vector of a point (r, θ, z) in cylindrical polar coordinates is given by

$$\vec{r} = r\hat{e}_r + z\hat{e}_z.$$

This looks crazy! We have point in 3D but only two apparent coordinates specifying it with **no** explicit behaviour on θ . This is because, as we have seen in the last point above, \hat{e}_r actually depends **itself** on θ . Thus if you use a radial basis vector \hat{e}_r , you are implicitly specifying a θ . Hence the third coordinate θ is actually there in the definition.

10.2 Velocity and Acceleration

The last point above shows that two of the unit vectors in cylindrical polars, \hat{e}_r and \hat{e}_θ , depend on their location θ . Hence as one moves around space in θ the orientation of these two unit vectors changes locally, depending on the point. This does not happen for the Cartesian basis vectors \hat{i}, \hat{j} and \hat{k} , which are independent of the position. (Note that since $\hat{e}_z = \hat{k}$ it does not change locally either.)

Suppose we wish to represent a derived quantity, such as velocity or acceleration in cylindrical polars. When we take the derivative(s) we must be sure to take into account the fact that the basis vectors are also changing from point to point, starting with the position representation.

Assume that the position of a particle at time t relative to an inertial frame can be represented by the cylindrical coordinates $(r(t), \theta(t), z(t))$. Then the vector position of the particle is

$$\vec{r} = r(t)\hat{e}_r + z(t)\hat{e}_z.$$

Since θ is a function of time, from the equation above relating polar and Cartesian bases, so will \hat{e}_r and \hat{e}_θ . The velocity of the particle is then given by

$$\begin{aligned}\frac{d\vec{r}}{dt} &= \frac{d}{dt}(r\hat{e}_r) + \frac{d}{dt}(z\hat{e}_z), \\ &= \dot{r}\hat{e}_r + r\dot{\hat{e}}_r + \dot{z}\hat{e}_z + z\dot{\hat{e}}_z.\end{aligned}$$

But $\dot{\hat{e}}_z = \dot{\hat{k}} = \vec{0}$ and

$$\begin{aligned}\dot{\hat{e}}_r &= \frac{d}{dt}(\cos\theta\hat{i} + \sin\theta\hat{j}) = \dot{\theta}(-\sin\theta\hat{i} + \cos\theta\hat{j}) = \dot{\theta}\hat{e}_\theta, \\ \dot{\hat{e}}_\theta &= \frac{d}{dt}(-\sin\theta\hat{i} + \cos\theta\hat{j}) = -\dot{\theta}(\cos\theta\hat{i} + \sin\theta\hat{j}) = -\dot{\theta}\hat{e}_r,\end{aligned}$$

where the last two derivatives also use the fact that $\dot{\hat{i}} = \dot{\hat{j}} = \vec{0}$. Hence substitution of these results into the expression for \vec{r} gives

$$\frac{d\vec{r}}{dt} = \dot{r}\hat{e}_r + r\dot{\theta}\hat{e}_\theta + \dot{z}\hat{e}_z.$$

Note that unlike the position vector in cylindrical polars, the velocity vector **does** involve a \hat{e}_θ component. That this is sensible follows from a qualitative description of each term in the velocity of the particle:

- $\dot{r}\hat{e}_r$ is the velocity in the outward radial direction;
- $r\dot{\theta}\hat{e}_\theta$ is the velocity in the azimuthal direction, i.e., around the z -axis and also perpendicular to \hat{e}_r ;
- $\dot{z}\hat{e}_z$ is the velocity in the z direction.

In order to work out the acceleration, we use the results from the velocity. Thus we have

$$\begin{aligned}\frac{d^2\vec{r}}{dt^2} &= \frac{d}{dt}(\dot{r}\hat{e}_r) + \frac{d}{dt}(r\dot{\theta}\hat{e}_\theta) + \frac{d}{dt}(\dot{z}\hat{e}_z) \\ &= \ddot{r}\hat{e}_r + \dot{r}\dot{\hat{e}}_r + \dot{r}\dot{\theta}\hat{e}_\theta + r\ddot{\theta}\hat{e}_\theta + r\dot{\theta}\dot{\hat{e}}_\theta + \ddot{z}\hat{e}_z + z\dot{\hat{e}}_z \\ &= \ddot{r}\hat{e}_r + \dot{r}\dot{\theta}\hat{e}_\theta + \dot{r}\dot{\theta}\hat{e}_\theta + r\ddot{\theta}\hat{e}_\theta - r\dot{\theta}^2\hat{e}_r + \ddot{z}\hat{e}_z \\ &= (\ddot{r} - r\dot{\theta}^2)\hat{e}_r + (2\dot{r}\dot{\theta} + r\ddot{\theta})\hat{e}_\theta + \ddot{z}\hat{e}_z\end{aligned}$$

That is,

$$\frac{d^2\vec{r}}{dt^2} = (\ddot{r} - r\dot{\theta}^2)\hat{e}_r + \frac{1}{r}\frac{d}{dt}(r^2\dot{\theta})\hat{e}_\theta + \ddot{z}\hat{e}_z.$$

so the three cylindrical components of acceleration are

- $(\ddot{r} - r\dot{\theta}^2)$ is the radial component of acceleration;
- $\frac{1}{r}\frac{d}{dt}(r^2\dot{\theta})$ is the azimuthal component of acceleration;
- \ddot{z} is the component of acceleration parallel to the z -axis.

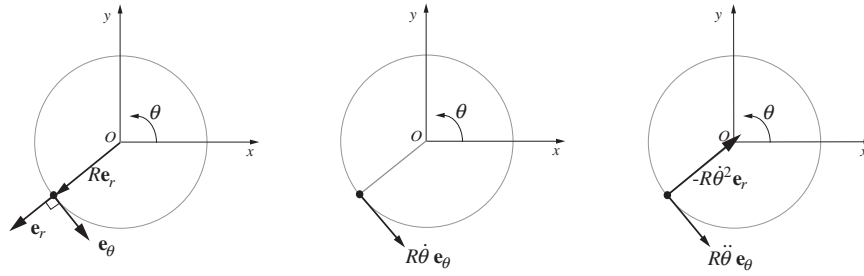


Figure 10.2: Position (left), velocity (centre) and acceleration (right) vectors for motion in a circle of radius R

Example 26: Motion in a circle

A particle moves in a circle of radius R in the $z = 0$ plane with position \vec{r} at time t . We study the problem using cylindrical coordinates with origin at the centre of the circle, with \hat{e}_z oriented perpendicular to the plane of the circle.

By definition we have the radial component of the position as

$$r(t) = R \Rightarrow \dot{r} = 0 \Rightarrow \ddot{r} = 0.$$

Likewise we have

$$z(t) = 0 \Rightarrow \dot{z} = 0 \Rightarrow \ddot{z} = 0.$$

Hence the velocity of the particle in radial coordinates is

$$\begin{aligned} \frac{d\vec{r}}{dt} &= \dot{r}\hat{e}_r + r\dot{\theta}\hat{e}_\theta + \dot{z}\hat{e}_z = 0\hat{e}_r + R\dot{\theta}\hat{e}_\theta + 0\hat{e}_z \\ \frac{d\vec{r}}{dt} &= R\dot{\theta}\hat{e}_\theta \end{aligned}$$

It is purely tangential to the circle (as it has to be if the circular motion is to be maintained!). The acceleration of the particle in radial components is

$$\begin{aligned} \frac{d^2\vec{r}}{dt^2} &= (\ddot{r} - r\dot{\theta}^2)\hat{e}_r + \frac{1}{r}\frac{d}{dt}(r^2\dot{\theta})\hat{e}_\theta + \ddot{z}\hat{e}_z = (0 - R\dot{\theta}^2)\hat{e}_r + R\ddot{\theta}\hat{e}_\theta + 0\hat{e}_z, \\ \frac{d^2\vec{r}}{dt^2} &= -R\dot{\theta}^2\hat{e}_r + R\ddot{\theta}\hat{e}_\theta. \end{aligned}$$

Note that the radial component of acceleration is directed towards the centre of the circle. This is called the **centripetal** (literally meaning “seeking the centre”) component of acceleration. If the particle has mass m and executes the motion with a constant angular speed, i.e.,

$$\dot{\theta} = \text{constant} = \omega, \Rightarrow \theta = \omega t + \text{constant},$$

then the velocity of the particle is

$$\frac{d\vec{r}}{dt} = R\omega\hat{e}_\theta.$$

The acceleration simplifies to only a radial component and **N2** gives the force as

$$\frac{d^2\vec{r}}{dt^2} = -R\dot{\theta}^2\hat{e}_r \Rightarrow \vec{F} = -mR\omega^2\hat{e}_r.$$

Note that the particle is only accelerating towards the centre of the circle. The interpretation of \vec{F} is that it is the force that is required to keep the particle in the circular orbit of the origin. Without the presence of \vec{F} , the particle would just continue in a straight line (**N1**). If the particle is attached to hoop or a string, by **N3** there is a balancing reactive force called the **centrifugal** force.

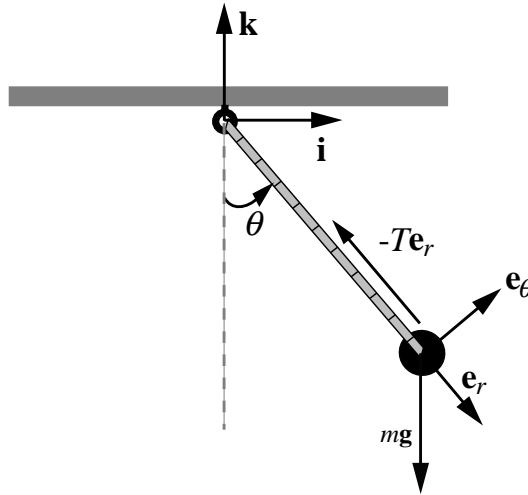


Figure 10.3: Forces acting on a pendulum represented in cylindrical polar coordinates.

Example 27: Motion of a plane pendulum revisited

Suppose a pendulum of mass m and fixed inextensible length l moves in a plane under the influence of gravity only, see figure 10.3. It is released from rest at $t = 0$ from an angle α radians, where $\alpha \ll 1$. Analyse the motion using cylindrical polar coordinates.

We fix coordinates with \hat{e}_r pointing along the length of the string, from the suspension point to the bob. \hat{e}_z passes horizontally through the suspension point, and points into the page. The motion of the bob will take place in the $z = 0$ plane. \hat{e}_θ is then determined by $\hat{e}_r \times \hat{e}_\theta$. The angle θ is fixed to be the angle of the string to the downwards vertical. The forces acting are the weight of the bob $m\vec{g}$ and the tension \vec{T} acting along the string. **N2** gives:

$$m\ddot{\vec{r}} = \vec{T} + m\vec{g}.$$

In terms of cylindrical polar coordinates, resolving the force components we have

$$\vec{T} = -T\hat{e}_r, \quad \vec{g} = g(\cos\theta\hat{e}_r - \sin\theta\hat{e}_\theta)$$

Thus **N2** becomes

$$m \left[\left(\ddot{r} - r\dot{\theta}^2 \right) \hat{e}_r + \frac{1}{r} \frac{d}{dt} \left(r^2 \dot{\theta} \right) \hat{e}_\theta \right] = -T \hat{e}_r + mg (\cos \theta \hat{e}_r - \sin \theta \hat{e}_\theta)$$

Now as the string is inextensible, $r(t) = l$, so in component form we have

$$\begin{aligned} -ml\dot{\theta}^2 &= -T + mg \cos \theta \\ ml\ddot{\theta} &= -mg \sin \theta \end{aligned}$$

For small oscillations $|\theta| \ll 1$, $\sin \theta \approx \theta$ and $\cos \theta \approx 1$ (remember θ will be measured in radians and 1 radian = $(180/\pi)^\circ$). Hence the last equation becomes

$$\ddot{\theta} + \frac{g}{l} \theta = 0.$$

This is simple harmonic motion, just as in the case of the vertical spring. It has solution

$$\theta(t) = C \cos \omega t + D \sin \omega t = A \cos(\omega t + B), \quad \omega = \frac{g}{l}.$$

The initial data is $\theta(0) = \alpha$, $\dot{\theta}(0) = 0$ which gives $A = \alpha$, $B = 0$. Thus the motion takes the form

$$\theta(t) = \alpha \cos \omega t.$$

Note that the frequency of such small oscillations is independent of m . The tension in the string is given by the first equation above as

$$T = mg \cos \theta + ml\dot{\theta}^2 \approx m (g + \alpha^2 \omega^2 l \sin^2 \omega t).$$

If the string is elastic, or the amplitude of the oscillations is large, the problem is much more difficult!

10.3 Spherical Polar Coordinates

When a problem has a spherical symmetry, such as the gravitational force felt by a satellite, or the precession of a spinning top, it makes sense to use **spherical polar coordinates**, (ρ, θ, ϕ) to simplify the governing equations.

Note that the θ angle does **not** have the same definition as in cylindrical coordinates.

The definitions of the coordinates are:

- ρ is the radial (shortest) distance from the origin of coordinates \mathcal{O} , $0 < \rho < \infty$
- θ is the angle of latitude as measured from the z axis or \hat{k} direction.
- ϕ is the longitude as measured from the Cartesian x -axis.

Note that the latitude θ is measured from the north pole and so the geographical latitude = $\theta - \pi/2$, $0 < \theta < \pi$. With these definitions we have the transformation

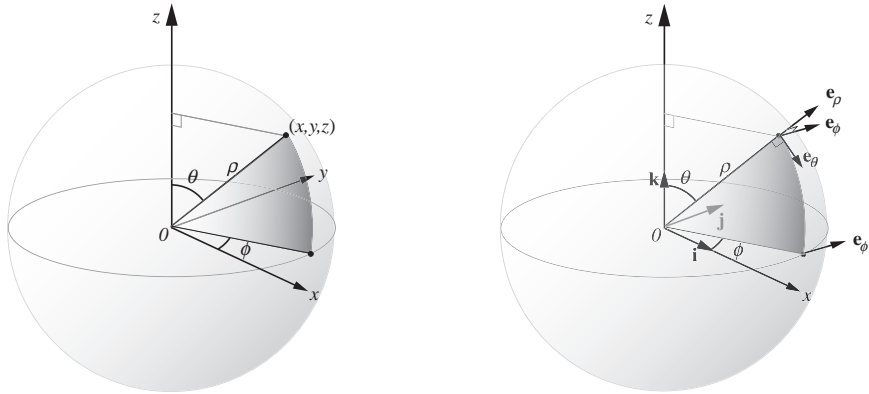


Figure 10.4: Spherical polar coordinate system and basis vectors. Note that the ϕ angle here is the same as the θ angle in the case of cylindrical coordinates.

$$(r, \theta, \phi) \rightarrow (x, y, z)$$

$$x = \rho \sin \theta \cos \phi \quad y = \rho \sin \theta \sin \phi \quad z = \rho \cos \theta.$$

$$\rho = \sqrt{x^2 + y^2 + z^2}$$

The unit basis vectors at any point P are given by:

→ \hat{e}_ρ which is the outwards radial unit vector in the direction \mathcal{OP} .

→ $\hat{e}_\phi = \hat{k} \times \hat{e}_\rho$.

→ $\hat{e}_\theta = \hat{e}_\phi \times \hat{e}_\rho$

In general, the position vector of a point (ρ, θ, ϕ) in spherical polar coordinates is given by

$$\vec{r} = \rho \hat{e}_\rho$$

Now the locally defined unit vector \hat{e}_r masks a dependence on both latitude θ and longitude ϕ .

Velocities and accelerations can be derived for the motion of a particle at $(\rho(t), \theta(t), \phi(t))$ in spherical polar coordinates at time t by techniques similar to those used for cylindrical polar coordinates. In summary, the results are: Velocity:

$$\dot{\vec{r}}(t) = \dot{\rho} \hat{e}_\rho + \rho \dot{\theta} \hat{e}_\theta + \rho \dot{\phi} \sin \theta \hat{e}_\phi$$

Acceleration:

$$\ddot{\vec{r}}(t) = \left(\ddot{\rho} - \rho \dot{\theta}^2 - \rho \dot{\phi}^2 \sin^2 \theta \right) \hat{e}_\rho + \left(2\dot{\rho} \dot{\theta} + \rho \ddot{\theta} - \rho \dot{\phi}^2 \sin \theta \cos \theta \right) \hat{e}_\theta + \left(2\rho \dot{\theta} \dot{\phi} \cos \theta + 2\dot{\rho} \dot{\phi} \sin \theta + \rho \ddot{\phi} \sin \theta \right) \hat{e}_\phi.$$

10.4 Non-inertial frames

We know that inertial frames provide the setting for Newtonian dynamics, but what happens if you find yourself in a frame that is not inertial? We have already seen examples of this, expressed in terms of the changing basis vectors

of cylindrical and spherical coordinates. Let us now consider what Newton's equations of motion look like in non-inertial frames. In order to be explicit, we will only consider rotating frames.

Let us start with an inertial frame S with coordinate axes x , y and z (as usual). Our goal is to understand the motion of particles as seen in a non-inertial frame S' , with axes x' , y' and z' , which is rotating with respect to S . We denote the angle between the x -axis of S and the x' -axis of S' as θ . Since S' is rotating, we clearly have $\theta = \theta(t)$ and $\dot{\theta} \neq 0$.

Our first task is to describe the rotation of the axes. For this, we can use the angular velocity vector $\vec{\omega}$ that we introduced earlier. Consider a particle that is sitting stationary in the S' frame. Then, from the perspective of frame S it will appear to be moving with velocity

$$\dot{\vec{r}} = \vec{\omega} \times \vec{r}$$

where, in the present case $\vec{\omega} = \dot{\theta} \hat{e}_z$.

We can extend the description of the rotation to the axes of S' themselves. Let \hat{e}'_i , $i = 1, 2, 3$ be the unit vectors that point along the x' , y' and z' directions. Then these also rotate with velocity

$$\dot{\hat{e}}'_i = \vec{\omega} \times \hat{e}'_i$$

This is the key result that will allow us to understand motion in rotating frames.

Consider now a particle which is not fixed in the S' frame, but moves along trajectory. We can measure the position of the particle in the inertial frame S , where

$$\vec{r} = \sum_i r_i \hat{e}_i$$

where \hat{e}_i are the basis vectors of S . Alternatively, we can measure the position of the particle in frame S' , where the position is

$$\vec{r} = \sum_i r'_i \hat{e}'_i$$

Of course, the actual position is the same in these two expressions. The coordinates differ because they are measured with respect to different axes. Let us now consider the velocity of the particle. In the fixed frame S we simply have

$$\dot{\vec{r}} = \sum_i \dot{r}_i \hat{e}_i$$

but in the rotating S' frame we have

$$\dot{\vec{r}} = \sum_i \left(\dot{r}'_i \hat{e}_i + r'_i \dot{\hat{e}}'_i \right) = \sum_i \left(\dot{r}'_i \hat{e}'_i + r'_i \vec{\omega} \times \hat{e}'_i \right) = \sum_i \dot{r}'_i \hat{e}'_i + \vec{\omega} \times \vec{r}$$

Let us introduce notation to help the intuition. We write the velocity of the particle as seen by an observer in frame S as

$$\left(\frac{d\vec{r}}{dt} \right)_S = \sum_i \dot{r}_i \hat{e}_i$$

Similarly, the velocity seen in S' is just

$$\left(\frac{d\vec{r}}{dt}\right)_{S'} = \sum_i \dot{r}'_i \hat{e}'_i$$

From our previous results, we then see that

$$\left(\frac{d\vec{r}}{dt}\right)_S = \left(\frac{d\vec{r}}{dt}\right)_{S'} + \vec{\omega} \times \vec{r}$$

The difference is (obviously) the relative velocity between the two frames.

We can play the same game with the acceleration. This is a bit more complicated, but after a bit of algebra¹, we find that

$$\left(\frac{d^2\vec{r}}{dt^2}\right)_S = \left(\frac{d^2\vec{r}}{dt^2}\right)_{S'} + 2\vec{\omega} \times \left(\frac{d\vec{r}}{dt}\right)_{S'} + \dot{\vec{\omega}} \times \vec{r} + \vec{\omega} \times (\vec{\omega} \times \vec{r})$$

This relation holds the key to understanding the motion of particles in a rotating frame.

Let us see what happens to **N2**. In the inertial frame S , we have

$$m \left(\frac{d^2\vec{r}}{dt^2}\right)_S = \vec{F}$$

This means that, in S' we have

$$\left(\frac{d^2\vec{r}}{dt^2}\right)_{S'} = \vec{F} - 2m\vec{\omega} \times \left(\frac{d\vec{r}}{dt}\right)_{S'} - m\dot{\vec{\omega}} \times \vec{r} - m\vec{\omega} \times (\vec{\omega} \times \vec{r})$$

We learn that, in order to explain the motion of a particle an observer in S' must invoke the existence of three further terms on the right-hand side of **N2**. These are called *fictitious forces*. Viewed from S' , a free particle does not travel in a straight line and these fictitious forces are necessary to explain this departure from uniform motion. The three extra terms are called the Coriolis force, the centrifugal force and the Euler force, respectively. It is worth noting that all the fictitious forces are proportional to the inertial mass m . This is because they all originated from the “ma” side of **N2**. But, the gravitational force also appears to be proportional to the inertial mass. Is this evidence that gravity is also a fictitious force? In fact it is. Einstein’s theory of general relativity recasts gravity as the fictitious force that we experience due to the curvature of space and time.

Example 28: Centrifugal force

The centrifugal force is given by

$$\vec{F}_{\text{cent}} = -m\vec{\omega} \times (\vec{\omega} \times \vec{r}) = -m(\vec{\omega} \cdot \vec{r})\vec{\omega} + m\omega^2\vec{r}$$

It is easy to see that this points away from the axis of rotation. Rotating bodies want to fly out from the axis of rotation to lower their energy.

¹Kindly left as an exercise!

On Earth, the centrifugal force impacts on the (local) strength of gravity. As an example of this, consider a piece of string suspended from the ceiling. You might expect the string to point down to the centre of the Earth. But the centrifugal force due to the Earth's rotation means that this is not the case. The angle that the string makes with the line pointing to the Earth's centre depends on the latitude, θ , at which we are sitting. The effective acceleration, due to the combination of gravity and the centrifugal force, is

$$\vec{g}_{\text{eff}} = \vec{g} - \vec{\omega} \times (\vec{\omega} \times \vec{r})$$

We can resolve this in radial and angular (Southernly) directions using unit vectors \hat{r} and $\hat{\theta}$. The centrifugal force then becomes

$$\vec{F}_{\text{cent}} = |\vec{F}| \cos \theta \hat{r} + |\vec{F}| \sin \theta \hat{\theta} = m\omega^2 r \cos^2 \theta \hat{r} - m\omega^2 r \cos \theta \sin \theta \hat{\theta}$$

At the surface of the Earth, with $r = R$, this leads to

$$\vec{g}_{\text{eff}} = (-g + \omega^2 R \cos^2 \theta) \hat{r} - \omega^2 R \cos \theta \sin \theta \hat{\theta}$$

We see that, at the equator, when $\theta = 0$, the string hangs directly towards the centre of the Earth, but we would measure gravity to be somewhat weaker than expected:

$$g_{\text{eff}} = g - \omega^2 R$$

Example 29: Coriolis force

The Coriolis force is given by

$$\vec{F}_{\text{cor}} = -2m\vec{\omega} \times \vec{v}$$

where \vec{v} is the velocity measured in the rotating frame

$$\vec{v} = \left(\frac{d\vec{r}}{dt} \right)_{S'}$$

The force is velocity dependent: it is only felt by moving particles, but it is independent of the position.

We can compare the Coriolis force to one of our earlier examples, that of a charged particle moving in a magnetic field. This means that we already know what the effect of the force will be – it makes moving particles turn in circles.

On Earth, the Coriolis force is responsible for the formation of hurricanes. These rotate in an anti-clockwise direction in the Northern hemisphere and a clockwise direction in the Southern hemisphere.

10.5 Central Forces

A central force acting between two particles is a force that always acts along a line joining the two particles. Often, for convenience, one of the particles will be chosen to represent the origin of the coordinate system. This is the convention we adopt from here on.

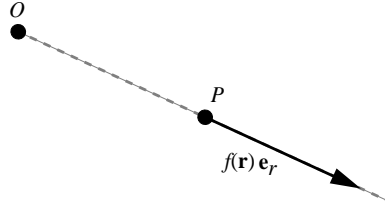


Figure 10.5: A central force $f(\vec{r})\hat{e}_r$ acting on particle P .

Examples of important central forces include:

- **Gravity:** Consider two masses m_1 and m_2 in free space. We fix the coordinate origin on mass m_1 , say, and then the position vector of m_2 with respect to m_1 is \vec{r} . The gravitational force acting on m_2 is

$$\vec{F}(\vec{r}) = -\frac{Gm_1m_2}{|\vec{r}|^2}\hat{e}_r$$

where $\hat{e}_r = \vec{r}/|\vec{r}|$ and $G = 6.6742 \times 10^{-11} \text{ Nm}^2\text{kg}^{-2}$ is the universal gravitational constant. G is a fundamental physical constant and is the same for gravitational interactions between bodies with mass.² Note that the force is always attractive.

- **Coulomb interactions between electrical charge:** Consider two point particles in free space with electric charge q_1 and q_2 . We fix the coordinate origin on charge q_1 , say, and then the position vector of q_2 with respect to q_1 is \vec{r} . The electrical force acting on q_2 is

$$\vec{F}(\vec{r}) = \frac{q_1q_2}{4\pi\epsilon_0|\vec{r}|^2}\hat{e}_r$$

where $\hat{e}_r = \vec{r}/|\vec{r}|$ and $\epsilon_0 = 8.854 \dots \times 10^{-12} \text{ Fm}^{-1}$. ϵ_0 is the electrical permittivity of free space and is another fundamental constant of nature. Note that if $q_1q_2 > 0$ (like charges) the force is repulsive. If $q_1q_2 < 0$ (unlike charges) the force is attractive.

Moving on, let us represent the force acting per unit mass as $f(\vec{r})\hat{e}_r$. **N2** can then be written as

$$m\ddot{\vec{r}} = mf(\vec{r})\hat{e}_r.$$

If we take moments about the origin, using **N2** we have

$$\begin{aligned}\vec{r} \times \ddot{\vec{r}} &= \vec{r} \times (f(\vec{r})\hat{e}_r), \\ &= f(\vec{r}) (\vec{r} \times \hat{e}_r), \\ &= |\vec{r}|f(\vec{r}) (\hat{e}_r \times \hat{e}_r), \\ \Rightarrow \vec{r} \times \ddot{\vec{r}} &= \vec{0}\end{aligned}$$

²Due to the extremely weak nature of gravity G is still one of the least accurately known fundamental constants.

Now, since

$$\begin{aligned}\frac{d}{dt}(\vec{r} \times \dot{\vec{r}}) &= (\dot{\vec{r}} \times \dot{\vec{r}}) + (\vec{r} \times \ddot{\vec{r}}) = \vec{0} + (\vec{r} \times \ddot{\vec{r}}) \\ \Rightarrow \vec{r} \times \ddot{\vec{r}} &= \frac{d}{dt}(\vec{r} \times \dot{\vec{r}}),\end{aligned}$$

we have

$$\begin{aligned}\frac{d}{dt}(\vec{r} \times \dot{\vec{r}}) &= \vec{0} \\ \Rightarrow \vec{r} \times \dot{\vec{r}} &= \vec{l}\end{aligned}$$

where \vec{l} is a constant vector.

We learn that both the velocity and the position of a particle moving under a central force always lie in a plane. The normal to this plane is the constant vector \vec{l} .

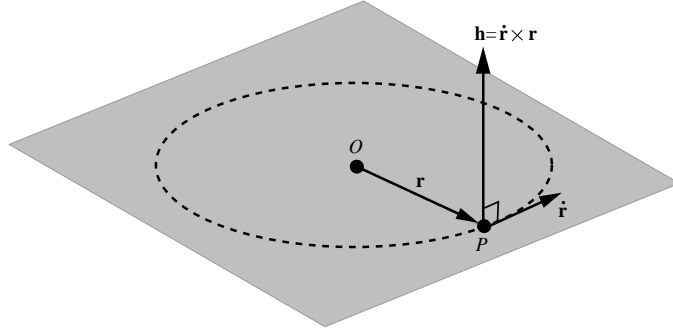


Figure 10.6: Motion under a central force is confined to a plane.

As we have seen, gravitational attraction is a central force acting between the centres of mass of two objects. This is why the Earth travels around the sun in a plane, or why a satellite moves in a planar orbit.

10.6 Conservation of Angular Momentum

Recall that the angular momentum \vec{L} of a body at position \vec{r} with linear momentum \vec{p} about the origin \mathcal{O} is

$$\vec{L} \equiv \vec{r} \times \vec{p} = m(\vec{r} \times \dot{\vec{r}}) = m\vec{l}.$$

The angular momentum of a particle acted on by a central force is clearly conserved.

Since the motion of the particle is confined to a plane and is around a central point to which the force is directed to or from, it is natural to represent the dynamics of the particle using cylindrical polar coordinates with $z = 0$, i.e., polar coordinates in the plane (the problem is effectively two dimensional). If

we write **N2** in component form in these coordinates (ignoring the constant z -component), we have

$$m\ddot{\vec{r}} = mf(\vec{r})\hat{e}_r$$

$$\left(\ddot{r} - r\dot{\theta}^2\right)\hat{e}_r + \frac{1}{r}\frac{d}{dt}\left(r^2\dot{\theta}\right)\hat{e}_\theta = f(r)\hat{e}_r$$

$$\Rightarrow \begin{cases} \left(\ddot{r} - r\dot{\theta}^2\right) = f(r) \\ \frac{1}{r}\frac{d}{dt}\left(r^2\dot{\theta}\right) = 0 \end{cases}$$

This equation may be integrated immediately to obtain

$$r^2\dot{\theta} = l$$

where l is a (scalar) constant.

This is the explicit coordinate form for the conservation of angular momentum for a unit mass. To see this we write $\vec{r} \times \dot{\vec{r}}$ in coordinate form:

$$\begin{aligned} \vec{r} \times \dot{\vec{r}} &= r\hat{e}_r \times \left(\dot{r}\hat{e}_r + r\dot{\theta}\hat{e}_\theta\right) \\ &= r\dot{r}\underbrace{(\hat{e}_r \times \hat{e}_r)}_{\vec{0}} + r^2\dot{\theta}\underbrace{(\hat{e}_r \times \hat{e}_\theta)}_{\hat{e}_z} \\ &= l\hat{e}_z = \vec{l}. \end{aligned}$$

To determine l it therefore suffices to know the position and velocity of a particle at a single point in its motion. This information will usually be given in the form of initial data, e.g., $\vec{r}(0)$ and $\dot{\vec{r}}(0)$.

10.7 Conservative Central Forces

Let us now consider the implication of the conservation of energy for systems with central forces. We assume that the central force only depends on $r(t)$, the (scalar) distance of the body from the origin at time t . Then **N2** becomes

$$m\ddot{\vec{r}} = mf(r)\hat{e}_r.$$

Moreover, such central forces are conservative³ The upshot of this is that we expect $f(r)$ to possess a potential and to generate an energy conservation law.

In order to demonstrate this observation law we need a couple of few preliminary calculations.

We have already seen that the motion under a central force is confined to a plane. Use this result to express the square of the speed of a body by introducing cylindrical polar coordinates.

³This can be checked by writing ∇ in polar coordinates and showing that the curl of $f(r)\hat{e}_r$ vanishes.

The velocity of a body in cylindrical polars is given by

$$\dot{\vec{r}} = \dot{r}\hat{e}_r + r\dot{\theta}\hat{e}_\theta.$$

The local basis vectors \hat{e}_r and \hat{e}_θ are orthogonal unit vectors. Hence, the square of the speed v of the body is given by

$$v^2 = \dot{\vec{r}} \cdot \dot{\vec{r}} = (\dot{r}\hat{e}_r + r\dot{\theta}\hat{e}_\theta) \cdot (\dot{r}\hat{e}_r + r\dot{\theta}\hat{e}_\theta) = \dot{r}^2 + r^2\dot{\theta}^2$$

Next, we use the conservation of angular momentum, which means that

$$\frac{1}{r} \frac{d}{dt} (r^2\dot{\theta}) = r\ddot{\theta} + 2\dot{r}\dot{\theta} = 0 \Rightarrow r\ddot{\theta} = -2\dot{r}\dot{\theta}$$

Hence, we have

$$\frac{d}{dt} \left(\frac{1}{2} r^2 \dot{\theta}^2 \right) = r\dot{r}\dot{\theta}^2 + r^2\dot{\theta}\ddot{\theta} = r\dot{r}\dot{\theta}^2 + r\dot{\theta} (-2\dot{r}\dot{\theta}) = -r\dot{r}\dot{\theta}^2.$$

We shall use this result shortly.

Now since f is no longer a function of \vec{r} but of r only, the \hat{e}_r component of **N2** above becomes

$$(\ddot{r} - r\dot{\theta}^2) = f(r).$$

If we multiply this equation through by \dot{r} we have

$$(\dot{r}\ddot{r} - \dot{r}r\dot{\theta}^2) = f(r)\dot{r}$$

$$\begin{aligned} \frac{d}{dt} \left\{ \frac{1}{2} (\dot{r}^2) \right\} + \overbrace{\frac{d}{dt} \left\{ \frac{1}{2} (r^2\dot{\theta}^2) \right\}}^{\text{using above result}} &= f(r) \frac{dr}{dt} \\ \frac{1}{2} \left\{ \dot{r}^2 + r^2\dot{\theta}^2 \right\} &= \int f(r) \frac{dr}{dt} dt + \text{constant} \\ \Rightarrow \frac{1}{2} m \left\{ \dot{r}^2 + r^2\dot{\theta}^2 \right\} - \int m f(r) dr &= E \end{aligned}$$

We now identify

$$\begin{aligned} \text{Kinetic Energy } T &= \frac{1}{2} m \left\{ \dot{r}^2 + r^2\dot{\theta}^2 \right\}, \\ \text{Potential Energy } V(r) &= - \int m f(r) dr, \\ \text{Total Energy} &= E \end{aligned}$$

and deduce that a conservation of energy equation

$$T + V = E$$

holds for central forces.

We can also use conservation of angular momentum $r^2\dot{\theta}$ to simplify the above expression for energy further. In particular, we can remove the explicit dependence⁴ on $\dot{\theta}$.

$$\begin{aligned}\frac{1}{2}m\left\{\dot{r}^2 + r^2\dot{\theta}^2\right\} + V(r) &= E \\ \frac{1}{2}m\left\{\dot{r}^2 + \frac{l^2}{r^2}\right\} &= E - V(r) \\ \Rightarrow \dot{r}^2 &= \frac{2}{m}[E - V(r)] - \frac{l^2}{r^2}\end{aligned}$$

This equation is a separable ordinary differential equation, which (at least in principle) can be integrated to give the position of the particle $r(t)$ explicitly. Once this is known then $r^2\dot{\theta} = l$ can be used to find $\theta(t)$. Of course, whether this is easily achievable depends on the exact form of $f(r)$.

10.8 Circular orbits and stability

We can use the above results to discuss the boundedness of the motion. For example, the existence of a potential (V) means that it is possible to establish the existence of a stable circular orbit, by analogy with the treatment for equilibrium states in statics.

Consider the central orbit energy equation

$$\frac{1}{2}m\left\{\dot{r}^2 + \frac{l^2}{r^2}\right\} + V(r) = E$$

This can be written in a form analogous to that of the the 1D equilibrium problem

$$\frac{1}{2}m\dot{r}^2 + U(r) = E$$

where

$$U(r) = \frac{1}{2}m\frac{l^2}{r^2} + V(r)$$

is an “effective potential”, since it combines the real potential $V(r)$ with an extra term. The extra l^2/r^2 piece is often referred to as the *centrifugal barrier*. It stops the particle from getting too close to the origin – there would be a heavy price to pay in “energy”.

Now, the term $\frac{1}{2}m\dot{r}^2$ looks like the usual 1D kinetic energy term. So we can pretend that this is a 1D system and treat the “effective potential” $U(r)$ as the potential in which the particle moves. This is not strictly correct, since $U(r)$ contains a term containing l , a constant that depends on the motion. However, since this term does not depend explicitly on the speed \dot{r} we can use this analogy to work out where the particle comes to rest, where the equilibrium points are, and assess their stability.

Continuing with the analogy, we know from 1D equilibrium theory that for a stable equilibrium to exist, $U(r)$ must have a minimum. In the context of $U(r)$, if such a minimum exists, it will be at a value of $r = r_0$ say. This corresponds to a circular orbit about the central point of radius r_0 .

⁴Recall that $l = r^2\dot{\theta}$.

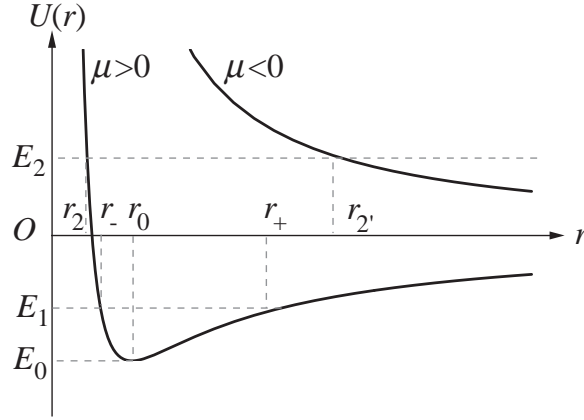


Figure 10.7: The effective potential $U(r) = \frac{ml^2}{2r^2} - \frac{m\mu}{r}$

Suppose we consider – as an example – an attractive inverse square law $f(r) = -\mu/r^2$. Relative to a point at infinity, this has a “proper” potential

$$V(r) = +m \int_{\infty}^r \frac{\mu}{\rho^2} d\rho = -\frac{m\mu}{r}.$$

This gives rise to the “effective” potential:

$$U(r) = \frac{ml^2}{2r^2} - \frac{m\mu}{r}$$

This has a minimum for $\mu > 0$ at

$$U'(r) = 0 \Rightarrow r = r_0 = \frac{l^2}{\mu}$$

Now suppose that the total energy of the system is E . Then rearranging the conservation equation we have

$$E - U(r) = \frac{1}{2}m\dot{r}^2.$$

Due to the positive quantity on the right-hand side, classically allowed motion can only take place when $E - U(r) > 0$, or $E > U(r)$. With this in mind we now examine the possibilities using figure 10.7.

- If the total energy of the system is $E = E_0 = U(r_0) = -m\mu/2l^2$, then the orbit will be a stable circle with radius r_0 .
- If the system has a higher energy E_1 (see diagram) the system will oscillate between r_- and r_+ .
- If the system has energy E_2 (see diagram) the system will possess no bounded orbit, and will lie in the range $r_2 < r < \infty$.
- If $\mu < 0$ there is no real minimum. All orbits are unbounded, e.g., if $E = E_2$, the orbit will lie in the range $r_{2'} < r < \infty$

It is also useful to ask if the orbits we determine are stable. As we are effectively dealing with a 1D problem, this is not a very tricky problem. Let us take as an example the case of circular orbits from above. We already know that we have a circular orbit if there is a solution, with $l \neq 0$, such that $\dot{r} = 0$ for all time. This means that we must have $\ddot{r} = 0$, which in turn leads to

$$U'(r) = 0$$

Not surprisingly, a circular orbit corresponds to a critical point r_0 of the effective potential. This orbit will be stable if a small perturbation brings us back to the original point. Basically, we have to sit at a minimum of the potential. That is, for stability we must have

$$U''(r_0) > 0$$

10.9 The orbit equation

So far we have not chosen to work out the actual orbital motion. Instead, we arrived at the energy conservation law by integrating **N2** once. Sometimes it is also possible to integrate **N2** directly to find the locus of the body as it moves under a central force.

$$\ddot{r} - r\dot{\theta}^2 = f(r), \quad r^2\dot{\theta} = l \Rightarrow \ddot{r} - \frac{l^2}{r^3} = f(r).$$

The solution of this equation is often facilitated by using the change of variables

$$r = \frac{1}{u}, \quad r^2\dot{\theta} = l \Rightarrow \dot{\theta} = lu^2$$

This way we obtain an ordinary differential equation for $u(\theta)$ as opposed to $r(t)$. Note that here we are solving for r as a function of θ . We are not deriving the time dependence (which in principle can be obtained from the energy equation). The procedure is as follows:

$$\dot{r} = \frac{d}{dt} \left(\frac{1}{u} \right) = -\frac{1}{u^2} \frac{du}{dt} = -\frac{1}{u^2} \frac{du}{d\theta} \frac{d\theta}{dt} = -r^2 \dot{\theta} \frac{du}{d\theta} = -l \frac{du}{d\theta}$$

$$\begin{aligned} \ddot{r} &= \frac{d}{dt} (\dot{r}) = \frac{d}{dt} \left(-l \frac{du}{d\theta} \right) = \frac{d\theta}{dt} \frac{d}{d\theta} \left(-l \frac{du}{d\theta} \right) \\ &= \dot{\theta} \frac{d}{d\theta} \left(-l \frac{du}{d\theta} \right) = -l \dot{\theta} \frac{d^2 u}{d\theta^2} = -l^2 u^2 \frac{d^2 u}{d\theta^2} \end{aligned}$$

where we have used

$$\frac{dl}{d\theta} = \frac{dl}{dt} \frac{dt}{d\theta} = 0$$

Making these two substitutions into **N2** we have

$$\ddot{r} - \frac{l^2}{r^3} = f(r) \Rightarrow -l^2 u^2 \frac{d^2 u}{d\theta^2} - l^2 u^3 = f \left(\frac{1}{u} \right)$$

or

$$\frac{d^2 u}{d\theta^2} + u = -\frac{1}{l^2 u^2} f \left(\frac{1}{u} \right)$$

This is a second order differential equation for u in terms of θ and so requires two pieces of initial/boundary data. The solution then generates $r(\theta) = 1/u(\theta)$. Alternatively, if the orbit of the particle is known experimentally as $u = u(\theta)$, this equation can be used to determine the central force $f(1/u)$ underlying the motion.

10.10 Inverse square law

Perhaps the most physically important central force is the inverse square law, which governs the gravitational attraction between masses and the Coulomb interaction of charged particles.

Let us focus on the simplest case; an attractive law

$$f(r) = -\frac{\mu}{r^2} = -\mu u^2,$$

which generates an orbit equation of the form

$$\frac{d^2u}{d\theta^2} + u = -\frac{1}{l^2 u^2} f\left(\frac{1}{u}\right) \Rightarrow \frac{d^2u}{d\theta^2} + u = \frac{\mu}{l^2}.$$

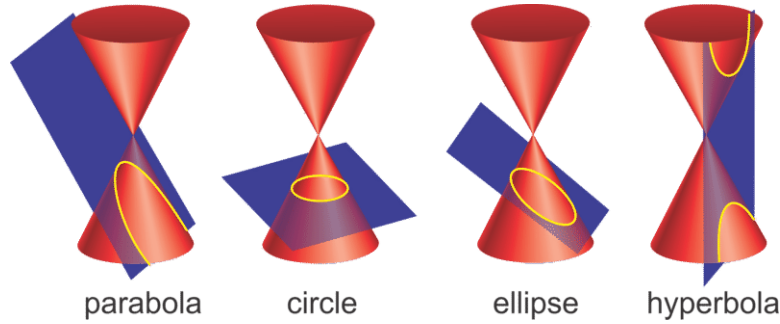


Figure 10.8: The conic sections.

This is a problem we know how to solve – it is just the constant-forcing harmonic oscillator equation. It can be solved with constant particular integral to give:

$$u(\theta) = \frac{\mu}{h^2} + A \sin \theta + B \cos \theta,$$

or in an equivalent form

$$u(\theta) = \frac{\mu}{h^2} + C \cos(\theta + D).$$

Now let us assume (actually without loss of generality for what follows) that we fix the $\theta = 0$ direction such that $D = 0$. Furthermore, we can write $l = h^2/\mu$ and rename the arbitrary constant C as e/l . Then the locus becomes:

$$u = \frac{1}{l} + \frac{e}{l} \cos \theta \Rightarrow \frac{l}{r} = 1 + e \cos \theta,$$

$$\Rightarrow r = \frac{l}{1 + e \cos \theta}.$$

This is the general equation of a **conic section**⁵.

- l is called the ‘**semi latus rectum**’⁶ and effectively determines the scale of the orbit.
- e is called the **eccentricity** of the orbit and determines whether it is circular, elliptic, parabolic, or hyperbolic. It is related to the energy of the orbiting particle.

Once we understand the motion, we can derive Kepler’s three laws of planetary motion (from 1605!):

1. Each planet moves in an ellipse with the Sun at one focus.
2. The line between the planet and the Sun sweeps out equal areas in equal times.
3. The period of the orbit is proportional to $r^{3/2}$.

⁵So we are making contact with a classic bit of geometry.

⁶Yes, this is worth sniggering at.

Chapter 11

Analytical Dynamics

We are now going to take a slightly different approach to dynamical problems. The starting point is a formalism developed by Lagrange to represent the behaviour of mechanical systems, irrespective of the number of constituent parts or chosen coordinate systems. At the heart of the approach is the notion of energy conservation. The advantage is that this new strategy can be applied to quite complicated systems (with extensions to particle physics, general relativity etc), to derive the equations of motion, without the need for physical intuition. The underlying physical laws are not changed by this approach, but are reduced to a set of mathematical relationships between standard functions and equations, abstracted from the physical context.

The ideas introduced here, including functions such as the Lagrangian or Hamiltonian, turn up in many other areas of applied mathematics and physics, from the quantum to the astronomical scale.

Analytical dynamics is also a powerful grounding in the more general (and abstract) theory of the dynamical systems which studies the evolution of arbitrary systems, not necessarily mechanical, as functions of their underlying governing parameters.

The concepts of conserved quantities naturally arise here, which ultimately can lead to studies of integrable and chaotic systems (which we will not discuss).

11.1 The Calculus of Variations

Lagrange's approach to mechanics is based on a method of deriving the equations of motion by finding the *minimum value* of an integral. This involves solving a problem in the calculus of variations. Before we consider the particular problem that arises in Lagrange's approach to mechanics we will start by considering some more general problems in the calculus of variations.

Example 30: The brachistochrone

Consider a bead sliding under the force of gravity. The bead starts being stationary at a point P and then slides down a smooth wire to a point Q . The points P and Q are fixed and we want to find the shape of the wire that minimises the time taken for the bead to get to Q .

Let the curve be $z = z(x)$ (where the z -axis points vertically upwards), and take P to have coordinates (x_0, z_0) to Q to have coordinates (x_1, z_1) . Using the initial data at P and conservation of energy we can find the time taken to go from P to Q . We have

$$T = \int_P^Q \frac{dl}{v}$$

where dl denotes a small element of length, measured along the wire. By conservation of energy:

$$mgz_0 = \frac{1}{2}mv^2 + mgz \Rightarrow v = [2g(z_0 - z)]^{1/2}$$

while

$$dl = [(dx)^2 + (dz)^2]^{1/2} = \left[1 + \left(\frac{dz}{dx}\right)^2\right]^{1/2} dx$$

We therefore obtain

$$T = \int_{x=x_0}^{x=x_1} \sqrt{\frac{1 + \dot{z}^2}{2g(z_0 - z)}} dx$$

where $\dot{z} = \frac{dz}{dx}$.

Again the time T depends upon the function $z(x)$ and is again a functional

$$T = T[z]$$

The problem is then to find $z(x)$ with $z(x_0) = z_0$ and $z(x_1) = z_1$ so that T is minimised. The general solution to the problem of finding the brachistochrone, or curve of quickest descent, was first found by J Bernoulli in 1696 who showed that it was a cycloid.

11.2 The Euler-Lagrange Equations

In essence, we have to minimise an integral that depends upon some function and its derivatives and has fixed end points. We therefore now consider the general case of a functional¹ that depends upon $x(t)$ and has the form

$$I[x] = \int_{t=t_0}^{t_1} f(x(t), \dot{x}(t), t) dt$$

where $f(\cdot, \cdot, \cdot)$ is a given function and $\dot{x} = \frac{dx}{dt}$.

The function $x(t)$ is taken to be fixed at the end points of the integral $x(t_0) = x_0$ and $x(t_1) = x_1$ and is required to be twice differentiable. We then want to find an expression for $x(t)$ which minimises the functional $I[x]$.

¹A functional is simply a function of a function.

In order to find $x(t)$ it is useful to suppose we have (somehow) found an $x(t)$ which minimises $I[x]$ and then consider what happens “near” that function. Consider some nearby function $\tilde{x}(t)$ which is ‘close to’ $x(t)$. We can construct a general expression for such nearby functions $\tilde{x}(t)$ as follows. Let $\eta(t)$ be a twice differentiable function (it has a continuous second derivative) defined for $t_0 \leq t \leq t_1$ which also satisfies the conditions

$$\eta(t_0) = 0, \quad \eta(t_1) = 0.$$

We call such a function η a variation. Then we can define a one parameter family of functions \tilde{x} close to $x(t)$ by using the variation η to define

$$\tilde{x}(t) = x(t) + \epsilon \eta(t)$$

where ϵ is a small parameter.

We now consider some fixed variation $\eta(t)$ and see how $I[\tilde{x}(t)]$ varies with ϵ . Hence we consider the function

$$\phi(\epsilon) = I[x(t) + \epsilon \eta(t)]$$

Since $x(t)$ gives a minimum for $I[x(t)]$ it follows that

$$I[x(t) + \epsilon \eta(t)] \geq I[x(t)]$$

and hence

$$\phi(\epsilon) \geq \phi(0)$$

Using results from the calculus of a function of one variable we see that in order that $\epsilon = 0$ is a minimum we must have

$$\frac{d\phi}{d\epsilon}(0) = 0$$

More generally if the above equation is true for all possible choices of variation $\eta(t)$ we say that $x(t)$ is a stationary value of $I[x(t)]$.

We now use the definition of $\phi(\epsilon)$, and the fact that $\epsilon = 0$ is a minimum of $\phi(\epsilon)$ to find equations that $x(t)$ must satisfy. We have

$$\phi(\epsilon) = \int_{t=t_0}^{t_1} f(x + \epsilon \eta, \dot{x} + \epsilon \dot{\eta}, t) dt$$

so that²

$$\begin{aligned} \frac{d\phi}{d\epsilon} &= \frac{d}{d\epsilon} \int_{t=t_0}^{t_1} f(x + \epsilon \eta, \dot{x} + \epsilon \dot{\eta}, t) dt \\ &= \int_{t=t_0}^{t_1} \frac{d}{d\epsilon} \{f(x + \epsilon \eta, \dot{x} + \epsilon \dot{\eta}, t)\} dt \\ &= \int_{t=t_0}^{t_1} \left(\frac{\partial f}{\partial \tilde{x}} \frac{d\tilde{x}}{d\epsilon} + \frac{\partial f}{\partial \dot{\tilde{x}}} \frac{d\dot{\tilde{x}}}{d\epsilon} \right) dt \\ &= \int_{t=t_0}^{t_1} \left(\eta \frac{\partial f}{\partial \tilde{x}} + \dot{\eta} \frac{\partial f}{\partial \dot{\tilde{x}}} \right) dt \end{aligned}$$

²There is no longer a way to avoid partial derivatives...

At $\epsilon = 0$ we have $\tilde{x} = x$ so that

$$\frac{d\phi}{d\epsilon}(0) = \int_{t=t_0}^{t_1} \left(\eta \frac{\partial f}{\partial x} + \dot{\eta} \frac{\partial f}{\partial \dot{x}} \right) dt$$

Integrating the $\dot{\eta}$ term by parts gives

$$\begin{aligned} \frac{d\phi}{d\epsilon}(0) &= \left[\frac{\partial f}{\partial \dot{x}} \eta \right]_{t_0}^{t_1} + \int_{t=t_0}^{t_1} \left\{ \eta \frac{\partial f}{\partial x} - \eta \frac{d}{dt} \left(\frac{\partial f}{\partial \dot{x}} \right) \right\} dt \\ &= - \int_{t=t_0}^{t_1} \left\{ \frac{d}{dt} \left(\frac{\partial f}{\partial \dot{x}} \right) - \frac{\partial f}{\partial x} \right\} \eta dt \end{aligned}$$

The first term vanishes because $\eta(t_0) = \eta(t_1) = 0$.

We have shown that if $x(t)$ is an stationary value of $I[x(t)]$ then we must have

$$\int_{t=t_0}^{t_1} \left\{ \frac{d}{dt} \left(\frac{\partial f}{\partial \dot{x}} \right) - \frac{\partial f}{\partial x} \right\} \eta dt = 0$$

for *all* possible choices of variation $\eta(t)$.

It follows that³

$$\frac{d}{dt} \left(\frac{\partial f}{\partial \dot{x}} \right) - \frac{\partial f}{\partial x} = 0$$

This is known as the Euler-Lagrange equation for $x(t)$. It is straightforward to reverse the argument to show that if $x(t)$ satisfies the Euler-Lagrange equation then $x(t)$ is an extremal for $I[x]$. We have therefore established the following result

$$x(t) \text{ is an extremal of } I[x(t)] \iff \frac{d}{dt} \left(\frac{\partial f}{\partial \dot{x}} \right) - \frac{\partial f}{\partial x} = 0$$

11.3 The Lagrangian

We are now in a position to give our first example of the application of the calculus of variations to mechanics.

³Technically we should appeal to the so-called Euler-Lagrange Lemma here...

Example 31: Recovering Newton

Consider the motion of a particle of mass m moving in 1D along the x -axis under the influence of a force $F(x)$ due to a potential $V(x)$.

We define the *Lagrangian* L of the particle to be given by

$$L = T - V = \frac{1}{2}m\dot{x}^2 - V(x)$$

(Note the minus sign that makes this different from the energy!!!)

Note also that

$$L = L(\dot{x}, x) .$$

We also define the *action* $I[x]$ to be given by

$$I[x] = \int_{t=t_0}^{t_1} L(x(t), \dot{x}(t), t) dt$$

Then an extremal value of $I[x]$ occurs if

$$\frac{d}{dt} \left(\frac{\partial L}{\partial \dot{x}} \right) - \frac{\partial L}{\partial x} = 0$$

This is *Lagrange's equation of motion*. Let us now show that this result is equivalent to Newton's second law for this problem. Using the definition of L it follows that (hold x fixed)

$$\frac{\partial L}{\partial \dot{x}} = m\dot{x}$$

and that (hold \dot{x} fixed in the derivative!)

$$\frac{\partial L}{\partial x} = -\frac{dV}{dx} = F$$

Using these in the Euler-Lagrange equations of motion gives us

$$\frac{d}{dt} (m\dot{x}) - F = 0$$

and so

$$m\ddot{x} - F = 0$$

which is just the equation of motion in Newtonian mechanics.

11.4 Adding degrees of freedom

The advantages of the Lagrangian method become more obvious if we consider a problem with several degrees of freedom⁴. We start by considering the motion of a particle moving in three dimensions under the action of a potential V . Let (x_1, x_2, x_3) be the x , y and z coordinates of the particle.

⁴Although we need to be comfortable with partial derivatives!

We use $x = x_1$, $y = x_2$ and $z = x_3$ so that the theory can be generalised later.

Then the kinetic energy is

$$T = \frac{1}{2}m(\dot{x}_1^2 + \dot{x}_2^2 + \dot{x}_3^2)$$

Since the force may be derived from a potential $V(x_1, x_2, x_3)$ the components of the force are given by

$$F_1 = -\frac{\partial V}{\partial x_1}, \quad F_2 = -\frac{\partial V}{\partial x_2} \quad \text{and} \quad F_3 = -\frac{\partial V}{\partial x_3}$$

The Lagrangian for this problem is

$$L = T - V = \frac{1}{2}m(\dot{x}_1^2 + \dot{x}_2^2 + \dot{x}_3^2) - V(x_1, x_2, x_3)$$

and the action is

$$I[x_1, x_2, x_3] = \int_{t=t_0}^{t_1} L(x_1, x_2, x_3, \dot{x}_1, \dot{x}_2, \dot{x}_3, t) dt$$

minimising the action now gives three equations

$$\frac{d}{dt} \left(\frac{\partial L}{\partial \dot{x}_i} \right) - \frac{\partial L}{\partial x_i} = 0, \quad \text{for } i = 1, 2, 3$$

Using the definition of the Lagrangian we have

$$\frac{\partial L}{\partial \dot{x}_i} = m\dot{x}_i$$

so that

$$\frac{d}{dt} \left(\frac{\partial L}{\partial \dot{x}_i} \right) = m\ddot{x}_i$$

Also

$$\frac{\partial L}{\partial x_i} = -\frac{\partial V}{\partial x_i} = F_i$$

So that the Lagrange equations of motion give

$$m\ddot{x}_i - F_i = 0, \quad \text{for } i = 1, 2, 3$$

which are simply the components of Newton's equations in three dimensions.

We have established the following result:

- Newton's equations for a particle acted on by a force due to a potential V are equivalent to the Lagrange equations of motion, which give the motion of the particle which minimises the action I .

In other words, one can derive Newton's laws of motion from the *principle of least action* (sometimes called Hamilton's principle). In fact the principle of least action is more fundamental than Newton's laws. As well as being used to derive the equations of motion for a particle it can be used to derive the equations of motion for a field. For example Maxwell's equations for electromagnetism can

be derived from the principle of least action as can Einstein's general relativistic field equations.

The big advantage of the principle of least action as far as mechanics is concerned is that it is true in *any* coordinate system, since the Lagrangian is a *scalar* that only depends on the motion of the particle not the coordinates used to describe it.

Let us illustrate this by obtaining the equations of motion for a particle moving in the plane but using polar coordinates (r, θ) to describe the motion rather than the usual Cartesian coordinates (x, y) .

The first step is to obtain a formula for the kinetic energy in plane polar coordinates. There are a number of ways of obtaining this but the most straightforward is to use the chain rule to differentiate the equations that relate Cartesian and polar coordinates.

$$x = r \cos \theta, \quad y = r \sin \theta$$

Differentiating the first of these⁵ with respect to t we get

$$\begin{aligned} \frac{dx}{dt} &= \frac{\partial x}{\partial r} \frac{dr}{dt} + \frac{\partial x}{\partial \theta} \frac{d\theta}{dt} \\ \dot{x} &= \dot{r} \cos \theta - r \dot{\theta} \sin \theta \end{aligned}$$

Similarly differentiating the second gives

$$\begin{aligned} \frac{dy}{dt} &= \frac{\partial y}{\partial r} \frac{dr}{dt} + \frac{\partial y}{\partial \theta} \frac{d\theta}{dt} \\ \dot{y} &= \dot{r} \sin \theta + r \dot{\theta} \cos \theta \end{aligned}$$

Squaring and adding these equations gives

$$\begin{aligned} \dot{x}^2 + \dot{y}^2 &= \dot{r}^2 \cos^2 \theta - 2r\dot{r}\dot{\theta} \sin \theta \cos \theta + r^2 \dot{\theta}^2 \sin^2 \theta \\ &+ \dot{r}^2 \sin^2 \theta + 2r\dot{r}\dot{\theta} \sin \theta \cos \theta + r^2 \dot{\theta}^2 \cos^2 \theta \\ &= \dot{r}^2 + r^2 \dot{\theta}^2 \end{aligned}$$

Hence in polar coordinates

$$T = \frac{1}{2} m (\dot{r}^2 + r^2 \dot{\theta}^2)$$

Using the formula for x and y in terms of r and θ also allows us to write the potential V as a function of r and θ . We may therefore write the Lagrangian for this problem as

$$L = \frac{1}{2} m (\dot{r}^2 + r^2 \dot{\theta}^2) - V(r, \theta)$$

To find the equations of motion we simply calculate the Euler-Lagrange equations for L . We start with the r -equation

$$\frac{\partial L}{\partial \dot{r}} = m \dot{r}$$

⁵This should be familiar from our discussion of cylindrical coordinates.

Hence

$$\frac{d}{dt} \left(\frac{\partial L}{\partial \dot{r}} \right) = m\ddot{r}$$

and

$$\frac{\partial L}{\partial r} = mr\dot{\theta}^2 - \frac{\partial V}{\partial r}$$

Note that this contains two terms because both T and V depend on r .

Hence the Euler-Lagrange equation for the r variable

$$\frac{d}{dt} \left(\frac{\partial L}{\partial \dot{r}} \right) - \frac{\partial L}{\partial r} = 0$$

gives

$$m\ddot{r} - mr\dot{\theta}^2 + \frac{\partial V}{\partial r} = 0$$

or

$$m(\ddot{r} - r\dot{\theta}^2) = -\frac{\partial V}{\partial r}$$

which may be shown to be the component of Newton's equation in the radial direction (by using vectors and resolving). The $r\dot{\theta}^2$ term is the 'centripetal acceleration' term from before.

To find the θ -equation we again just look at the corresponding Euler-Lagrange equation.

$$\frac{\partial L}{\partial \dot{\theta}} = mr^2\dot{\theta}$$

so that

$$\frac{d}{dt} \left(\frac{\partial L}{\partial \dot{\theta}} \right) = mr^2\ddot{\theta} + 2mr\dot{r}\dot{\theta}$$

Note that this contains two terms because both r and $\dot{\theta}$ depend upon t .

Differentiating L with respect to θ gives

$$\frac{\partial L}{\partial \theta} = -\frac{\partial V}{\partial \theta}$$

Hence the Euler-Lagrange equation for the θ variable

$$\frac{d}{dt} \left(\frac{\partial L}{\partial \dot{\theta}} \right) - \frac{\partial L}{\partial \theta} = 0$$

gives

$$mr^2\ddot{\theta} + 2mr\dot{r}\dot{\theta} = -\frac{\partial V}{\partial \theta}$$

or

$$m(r\ddot{\theta} + 2\dot{r}\dot{\theta}) = -\frac{1}{r} \frac{\partial V}{\partial \theta}$$

which may be shown to be the transverse component of Newton's equations. The $2\dot{r}\dot{\theta}$ term represent the Coriolis force (as before).

Example 32: The pendulum

As a first example, consider a simple pendulum. In this case the kinetic energy is given by

$$T = \frac{1}{2}mv^2 = \frac{1}{2}m(\dot{r}^2 + r^2\dot{\theta}^2) = \frac{1}{2}ml^2\dot{\theta}^2$$

since $r = l$ and $\dot{r} = 0$, while the potential energy is

$$V = mgh = mgl(1 - \cos \theta)$$

Hence the Lagrangian is

$$L = \frac{1}{2}ml^2\dot{\theta}^2 - mgl(1 - \cos \theta)$$

We now calculate the Lagrange equations of motion

$$\frac{\partial L}{\partial \dot{\theta}} = ml^2\dot{\theta}, \quad \frac{d}{dt} \left(\frac{\partial L}{\partial \dot{\theta}} \right) = ml^2\ddot{\theta}, \quad \frac{\partial L}{\partial \theta} = -mgl \sin \theta$$

so that $ml^2\ddot{\theta} + mgl \sin \theta = 0$ or

$$\ddot{\theta} + \frac{g}{l} \sin \theta = 0$$

One cannot solve this differential equation in closed form (without introducing so-called elliptic functions). However for small θ we have $\sin \theta \approx \theta$, so that for small oscillations the equation is approximately given by

$$\ddot{\theta} + \frac{g}{l} \theta = 0$$

writing $g/l = \omega^2$ we recognise this as simple harmonic motion with angular frequency ω so that the general solution is

$$\theta(t) = A \cos(\omega t) + B \sin(\omega t), \quad \text{where} \quad \omega = \sqrt{\frac{g}{l}}$$

So that the motion of the pendulum is described by oscillations about $\theta = 0$ with period $\tau = 2\pi\sqrt{l/g}$.

Example 33: Mass on a spring

Another example is a mass on a spring. Let x be the extension of the spring. The the force due to the spring is $F = -kx$, where k is the spring constant. Hence the potential energy stored in the spring is

$$V_{\text{spring}} = \frac{1}{2}kx^2$$

The gravitational potential energy is

$$V_{\text{grav}} = -mgx$$

So that the total potential energy is

$$V = \frac{1}{2}kx^2 - mgx$$

The kinetic energy of the mass is

$$T = \frac{1}{2}mv^2 = \frac{1}{2}m\dot{x}^2$$

Hence the Lagrangian for this problem is

$$L = \frac{1}{2}m\dot{x}^2 - \frac{1}{2}kx^2 + mgx$$

We now calculate the Lagrange equation of motion

$$\frac{\partial L}{\partial \dot{x}} = m\dot{x}, \quad \frac{d}{dt} \left(\frac{\partial L}{\partial \dot{x}} \right) = m\ddot{x}, \quad \frac{\partial L}{\partial x} = -kx + mg$$

So that

$$\frac{d}{dt} \left(\frac{\partial L}{\partial \dot{x}} \right) - \frac{\partial L}{\partial x} = 0$$

gives

$$m\ddot{x} + kx - mg = 0$$

or

$$\ddot{x} + \frac{k}{m}x = g$$

That is, we have recovered the usual result and we know that the mass oscillates about the equilibrium position $x = mg/k$ with angular frequency $\omega = \sqrt{k/m}$.

Example 34: Planetary motion

Lagrangian methods provide a very easy way of deriving the equations of motion of a small particle in the gravitational field of a much larger one, e.g. the motion of a planet around the Sun. We can assume that the planet's motion is confined to a plane containing the Sun, and use plane polar coordinates (r, ϕ) to label its position. The planet then has velocity

$$\mathbf{v} = \dot{r}\mathbf{e}_r + r\dot{\phi}\mathbf{e}_\phi$$

where a dot denotes a time derivative and $\mathbf{e}_r, \mathbf{e}_\phi$ are unit vectors in the radial and angular directions. The planet's kinetic energy is then

$$T = \frac{1}{2}m(\dot{r}^2 + r^2\dot{\phi}^2)$$

The potential energy in this case is gravitational potential energy, which Newton took to be given by the equation

$$V = -\frac{GMm}{r}$$

where M is the mass of the Sun (assumed fixed in space), m the mass of the planet, and G is Newton's gravitational constant. Note that this leads to a force \mathbf{F} given by

$$\mathbf{F} = -\nabla V = -\frac{\partial V}{\partial r}\mathbf{e}_r = -\frac{GMm}{r^2}\mathbf{e}_r,$$

This is the famous “inverse square law”.

The Lagrangian is then

$$L = T - V = \frac{1}{2}m(\dot{r}^2 + r^2\dot{\phi}^2) + \frac{GMm}{r}$$

The radial Euler-Lagrange equation

$$\frac{d}{dt}\frac{\partial L}{\partial \dot{r}} = \frac{\partial L}{\partial r} \quad \longrightarrow \quad \frac{d}{dt}(m\dot{r}) = m\dot{r}\dot{\phi}^2 - \frac{GMm}{r^2} \Rightarrow \ddot{r} = r\dot{\phi}^2 - \frac{GM}{r^2}$$

while the angular Euler-Lagrange equation

$$\frac{d}{dt}\frac{\partial L}{\partial \dot{\phi}} = \frac{\partial L}{\partial \phi} \quad \longrightarrow \quad \frac{d}{dt}(mr^2\dot{\phi}) = 0$$

These are two coupled ordinary differential equations which can then be solved to give $r(t)$ and $\phi(t)$, although the solution cannot be written in simple form. Note that the angular equation is particularly simple: the particle's angular momentum is given by (we will use J for the angular momentum to avoid confusion with the Lagrangian)

$$\mathbf{J} = \mathbf{r} \times \mathbf{p} = r\mathbf{e}_r \times m(\dot{r}\mathbf{e}_r + r\dot{\phi}\mathbf{e}_\phi) = mr^2\dot{\phi}\mathbf{e}_r \times \mathbf{e}_\phi$$

A little thought shows that for points in the $z = 0$ plane where our planet resides, $\mathbf{e}_r \times \mathbf{e}_\phi = \mathbf{e}_z$, so that

$$\mathbf{J} = mr^2\dot{\phi}\mathbf{e}_z \quad \Rightarrow \quad J = mr^2\dot{\phi}$$

i.e. the angular Euler-Lagrange equation is simply $J = \text{constant}$ (as expected from the discussion of central forces). The particle's energy is:

$$E = T + V = \frac{1}{2}m(\dot{r}^2 + r^2\dot{\phi}^2) - \frac{GMm}{r}$$

Eliminating $\dot{\phi}$ in favour of J leads to

$$\hat{E} = \frac{1}{2} \left(\dot{r}^2 + \frac{\hat{J}^2}{r^2} \right) - \frac{GM}{r}$$

where $\hat{E} = E/m$ and $\hat{J} = J/m$, the energy and angular momentum of the particle per unit mass.