Advanced Classical Physics, Autumn 2013

Arttu Rajantie

December 2, 2013

Preface

These are the lecture notes for the third-year Advanced Classical Physics course in the 2013-14 academic year at Imperial College London. They are based on the notes which I inherited from the previous lecturer Professor Angus MacKinnon.

The notes are designed to be self-contained, but there are also some excellent textbooks, which I want to recommend as supplementary reading. The core textbooks are

- Classical Mechanics (5th Edition), Kibble & Berkshire (Imperial College Press 2004),
- Introduction to Electrodynamics (3rd Edition), Griffiths (Pearson 2008),

and these books may also be useful:

- Classical Mechanics (2nd Edition), McCall (Wiley 2011),
- Classical Mechanics, Gregory (Cambridge University Press 2006),
- Classical Mechanics (3rd Edition), Goldstein, Poole & Safko (Addison-Wesley 2002),
- Mechanics (3rd Edition), Landau & Lifshitz (Elsevier 1976),
- Classical Electrodynamics (3rd Edition), Jackson (Wiley 1999),
- The Classical Theory of Fields, Landau & Lifshitz (Elsevier 1975).

The course assumes Mechanics, Relativity and Electromagnetism as background knowledge. Being a theoretical course, it also makes heavy use of most aspects of the compulsory mathematics courses. Mathematical Methods is also useful, but it is not a formal prerequisite, and all necessary concepts are introduced as part of this course, although in a less general way.

Arttu Rajantie

Contents

Chapter 1

Rotating Frames

1.1 Angular Velocity[1]

In order to describe rotation, you need to know its speed and the direction of the rotation axis. The speed is a scalar quantity and it is given by the angular frequency $\omega = 2\pi/T$, where T is the period of rotation, and the axis is a direction in space, so it can be represented by a unit vector $\hat{\mathbf{n}}$.

It is natural and useful to combine these to an angular velocity vector $\boldsymbol{\omega} = \omega\hat{\mathbf{n}}$. The sign of the vector $\boldsymbol{\omega}$ is determined by the *right-hand-rule*: If you imagine gripping the axis of rotation with the fingers of your right hand, your thumb will point to the direction of $\boldsymbol{\omega}$.

The angular velocity vector is called an *axial* vector (or sometimes a *pseudo-vector*), which means that is has different symmetry properties from a normal (*polar*) vector. Consider the effect of reflection in a plane containing the vector, e.g. a vector in the $\hat{\mathbf{z}}$ direction reflected in the $(y-z)$ plane. A *polar* vector is unchanged under such an operation, whereas an *axial* vector changes sign, as the direction of rotation is reversed.

For example, for the rotation of the earth (against the background of the stars), the angular velocity ω takes the value

$$\omega = \frac{2\pi}{86164\mathrm{s}} = 7.292 \times 10^{-5}\,\mathrm{s}^{-1}\,. \tag{1.1.1}$$

The angular momentum vector $\boldsymbol{\omega}$ points up at the North Pole.

Consider now a point $\mathbf{r}$ on the the rotating body, e.g. Blackett Lab. at latitude 51.5° N. This point is moving tangentially eastwards with a speed $v = \omega r \sin\theta$, where r is the distance from the origin (assumed to be on the axis of rotation (e.g. the centre of the Earth) and θ is the angle between the vectors $\mathbf{r}$ and ω (i.e., for Blackett, $\theta = 90^\circ - 51.5^\circ = 38.5^\circ$). Hence the velocity of the point $\mathbf{r}$ may be written

$$\frac{\mathrm{d}\mathbf{r}}{\mathrm{d}t} = \boldsymbol{\omega} \times \mathbf{r}\,. \tag{1.1.2}$$

We note here that geographers tend to measure latitude from the equator whereas the angle θ in spherical polar co-ordinates is defined from the pole. Thus the geographical designation 51.5° N corresponds to $\theta = 38.5^\circ$.

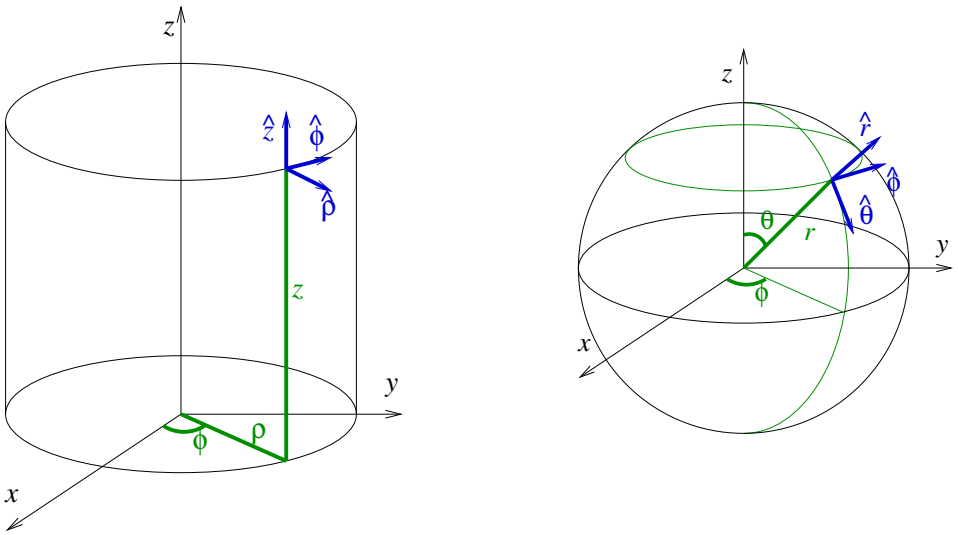

Figure 1.1: Cylindrical (left) and spherical (right) coordinate systems

1.2 Transformation of Vectors

Actually the result (1.1.2) is valid for any vector fixed in the rotating body, not just for position vectors. So, in general we may write

$$\frac{\mathrm{d}\mathbf{A}}{\mathrm{d}t} = \boldsymbol{\omega} \times \mathbf{A}\,. \tag{1.2.1}$$

In particular we can consider the case of a set of orthogonal unit vectors, $\hat{\boldsymbol{\imath}}$, $\hat{\boldsymbol{\jmath}}$, $\hat{\mathbf{k}}$ fixed in the body, chosen in accordance with the right-hand rule (index finger points in the direction of $\hat{\boldsymbol{\imath}}$, middle finger in the direction of $\hat{\boldsymbol{\jmath}}$ and thumb in the direction of $\hat{\mathbf{k}}$), which we can use as a basis of a rotating coordinate system. It is often convenient to choose $\hat{\boldsymbol{\imath}}$ to point radially away from the rotation axis and $\hat{\mathbf{k}}$ to be in the direction $\boldsymbol{\omega}$, so that $\hat{\boldsymbol{\jmath}}$ points in the direction of motion. This forms the basis of *cylindrical coordinates* (ρ, ϕ, z) and are often denoted by $\hat{\boldsymbol{\imath}} = \hat{\boldsymbol{\rho}}$, $\hat{\boldsymbol{\jmath}} = \hat{\boldsymbol{\phi}}$, $\hat{\mathbf{k}} = \hat{\mathbf{z}}$.

On the other hand, for motion on the surface for a rotating sphere such as the Earth, a convenient choice is to take $\hat{\boldsymbol{\imath}}$ pointing south, $\hat{\boldsymbol{\jmath}}$ pointing east and $\hat{\mathbf{k}}$ pointing up (i.e. away from the centre). These form the basis of *spherical coordinates* (r, θ, ϕ), with $\hat{\boldsymbol{\imath}} = \hat{\boldsymbol{\theta}}$, $\hat{\boldsymbol{\jmath}} = \hat{\boldsymbol{\phi}}$, $\hat{\mathbf{k}} = \hat{\mathbf{r}}$.

We have to be careful to distinguish between the point of view of an observer on the rotating object and one in a fixed (*inertial*) frame of reference observing the situation from outside. The basis vectors $\hat{\boldsymbol{\imath}}$, $\hat{\boldsymbol{\jmath}}$, $\hat{\mathbf{k}}$ are fixed from the point of view of the rotating observer to the inertial observer. We shall adopt the convention of using subscripts I and R to donate quantities in the inertial and rotating frames respectively (N.B. Kibble & Berkshire use a different convention).

Given a set of basis vectors $\hat{\boldsymbol{\imath}}$, $\hat{\boldsymbol{\jmath}}$, $\hat{\mathbf{k}}$, we can write any vector $\mathbf{A}$ as

$$\mathbf{A} = A_x\hat{\boldsymbol{\imath}} + A_y\hat{\boldsymbol{\jmath}} + A_z\hat{\mathbf{k}}\,. \tag{1.2.2}$$

We want to write down an expression which relates the rates of change of this vector in the two frames. We first note that a scalar quantity cannot depend on the choice of frame and that A_x, A_y and A_z may

[1] Kibble & Berkshire, chapter 5

each be considered as such scalars. Hence

$$\left.\frac{\mathrm{d}A_x}{\mathrm{d}t}\right|_{\mathrm{I}} = \left.\frac{\mathrm{d}A_x}{\mathrm{d}t}\right|_{\mathrm{R}} \qquad \text{etc.} \tag{1.2.3}$$

so that the differences between the vector $\mathbf{A}$ in the 2 frames must be solely related to the difference in the basis vectors. Hence

$$\begin{aligned}\left.\frac{\mathrm{d}\mathbf{A}}{\mathrm{d}t}\right|_{\mathrm{I}} &= \left(\frac{\mathrm{d}A_x}{\mathrm{d}t}\hat{\boldsymbol{\imath}} + \frac{\mathrm{d}A_y}{\mathrm{d}t}\hat{\boldsymbol{\jmath}} + \frac{\mathrm{d}A_z}{\mathrm{d}t}\hat{\mathbf{k}}\right) + \left(A_x\frac{\mathrm{d}\hat{\boldsymbol{\imath}}}{\mathrm{d}t} + A_y\frac{\mathrm{d}\hat{\boldsymbol{\jmath}}}{\mathrm{d}t} + A_z\frac{\mathrm{d}\hat{\mathbf{k}}}{\mathrm{d}t}\right) \\ &= \left(\frac{\mathrm{d}A_z}{\mathrm{d}t}\hat{\boldsymbol{\imath}} + \frac{\mathrm{d}A_y}{\mathrm{d}t}\hat{\boldsymbol{\jmath}} + \frac{\mathrm{d}A_z}{\mathrm{d}t}\hat{\mathbf{k}}\right) + \left(A_x(\boldsymbol{\omega}\times\hat{\boldsymbol{\imath}}) + A_y(\boldsymbol{\omega}\times\hat{\boldsymbol{\jmath}}) + A_z(\boldsymbol{\omega}\times\hat{\mathbf{k}})\right). \end{aligned} \tag{1.2.4}$$

The expression in the first brackets on the right-hand-side is precisely the time derivative measured by the rotating observer, so this is the relation we wanted: It relates the time derivatives measured by inertial and rotating observers. We can write it compactly as in the compact form

$$\left.\frac{\mathrm{d}\mathbf{A}}{\mathrm{d}t}\right|_{\mathrm{I}} = \left.\frac{\mathrm{d}\mathbf{A}}{\mathrm{d}t}\right|_{\mathrm{R}} + \boldsymbol{\omega}\times\mathbf{A}\,. \tag{1.2.5}$$

1.3 Equation of Motion

Applying Eq. (1.2.5) to the position of the particle $\mathbf{r}$, its velocity may be written as

$$\mathbf{v}_{\mathrm{I}} = \left.\frac{\mathrm{d}\mathbf{r}}{\mathrm{d}t}\right|_{\mathrm{I}} = \left.\frac{\mathrm{d}\mathbf{r}}{\mathrm{d}t}\right|_{\mathrm{R}} + \boldsymbol{\omega}\times\mathbf{r} = \mathbf{v}_{\mathrm{R}} + \boldsymbol{\omega}\times\mathbf{r}. \tag{1.3.1}$$

Now consider an object subject to a force $\mathbf{F}$. Newton's second law applies in the inertial frame, so we have

$$m\left.\frac{\mathrm{d}^2\mathbf{r}}{\mathrm{d}t^2}\right|_{\mathrm{I}} = \mathbf{F}\,. \tag{1.3.2}$$

To write this equation in the rotating frame, we differentiate the velocity $\mathbf{v}_{\mathrm{I}}$ using Eq. (1.2.5) again,

$$\begin{aligned}\left.\frac{\mathrm{d}^2\mathbf{r}}{\mathrm{d}t^2}\right|_{\mathrm{I}} &= \left.\frac{\mathrm{d}\mathbf{v}_{\mathrm{I}}}{\mathrm{d}t}\right|_{\mathrm{I}} \\ &= \left.\frac{\mathrm{d}\mathbf{v}_{\mathrm{R}}}{\mathrm{d}t}\right|_{\mathrm{I}} + \boldsymbol{\omega}\times\left.\frac{\mathrm{d}\mathbf{r}}{\mathrm{d}t}\right|_{\mathrm{I}} \\ &= \left.\frac{\mathrm{d}\mathbf{v}_{\mathrm{R}}}{\mathrm{d}t}\right|_{\mathrm{R}} + \boldsymbol{\omega}\times\mathbf{v}_{\mathrm{R}} + \boldsymbol{\omega}\times(\mathbf{v}_{\mathrm{R}} + \boldsymbol{\omega}\times\mathbf{r}) \\ &= \left.\frac{\mathrm{d}^2\mathbf{r}}{\mathrm{d}t^2}\right|_{\mathrm{R}} + 2\boldsymbol{\omega}\times\left.\frac{\mathrm{d}\mathbf{r}}{\mathrm{d}t}\right|_{\mathrm{R}} + \boldsymbol{\omega}\times(\boldsymbol{\omega}\times\mathbf{r})\,. \end{aligned} \tag{1.3.3}$$

Alternatively we can write this very concisely as

$$\begin{aligned}\left.\frac{\mathrm{d}^2\mathbf{r}}{\mathrm{d}t^2}\right|_{\mathrm{I}} &= \left(\left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{\mathrm{R}} + \boldsymbol{\omega}\times\right)^2\mathbf{r} \\ &= \left.\frac{\mathrm{d}^2\mathbf{r}}{\mathrm{d}t^2}\right|_{\mathrm{R}} + 2\boldsymbol{\omega}\times\left.\frac{\mathrm{d}\mathbf{r}}{\mathrm{d}t}\right|_{\mathrm{R}} + \boldsymbol{\omega}\times(\boldsymbol{\omega}\times\mathbf{r})\,. \end{aligned} \tag{1.3.4}$$

We now rearrange (1.3.3) and combine it with (1.3.2) to obtain an equation of motion for the particle in the rotating frame

$$m \left.\frac{\mathrm{d}^2\mathbf{r}}{\mathrm{d}t^2}\right|_{\mathrm{R}} = \mathbf{F} - 2m\boldsymbol{\omega} \times \mathbf{v}_{\mathrm{R}} - m\boldsymbol{\omega} \times (\boldsymbol{\omega} \times \mathbf{r}) \ . \tag{1.3.5}$$

The second term on the right-hand-side of (1.3.5) is the *Coriolis force* while the final term, pointing away from the axis, is the *centrifugal* force. They are called *fictitious forces*, because they do not represent real physical interactions but appear only because of the choice of the coordinate system. Fictitious forces are always proportional to the mass of the particle, so that the corresponding acceleration is independent of mass. Of course, this is also true for gravity, and in fact general relativity describes gravity as a fictitious force.

1.4 Coriolis Force

Equation (1.3.5) shows that, due to the Coriolis force, an object moving at velocity $\mathbf{v}$ in the rotating frame, experiences apparent acceleration

$$\mathbf{a}_{\mathrm{Cor}} = -2\boldsymbol{\omega} \times \mathbf{v}. \tag{1.4.1}$$

For example, imagine a car travelling north along Queen's Gate at $50\,\mathrm{km/h}$. To calculate the Coriolis acceleration it experiences, let us choose a set of orthogonal basis vectors that rotate with the Earth, for example $\hat{\boldsymbol{\imath}}$ pointing east, $\hat{\boldsymbol{\jmath}}$ north and $\hat{\mathbf{k}}$ up. With this choice the angular velocity vector has components

$$\boldsymbol{\omega} = (0, \omega\sin\theta, \omega\cos\theta) \ , \tag{1.4.2}$$

and the car has velocity $\mathbf{v} = v\hat{\boldsymbol{\jmath}}$. The Coriolis acceleration is, therefore,

$$\begin{aligned}
\mathbf{a}_{\mathrm{Cor}} &= -2\boldsymbol{\omega} \times \mathbf{v} = -2 \times \begin{vmatrix} \hat{\boldsymbol{\imath}} & \hat{\boldsymbol{\jmath}} & \hat{\mathbf{k}} \\ 0 & \omega\sin\theta & \omega\cos\theta \\ 0 & v & 0 \end{vmatrix} = 2\omega v\cos\theta\,\hat{\boldsymbol{\imath}} \\
&\approx 2 \times (7.292 \times 10^{-5}\,\mathrm{s}^{-1}) \times \left(\frac{50 \times 10^3\,\mathrm{m}}{3600\,\mathrm{s}}\right) \times \cos(38.5^\circ)\,\hat{\boldsymbol{\imath}} \\
&\approx 1.5\,\mathrm{mm/s^2}\ \text{eastwards}\,,
\end{aligned} \tag{1.4.3}$$

which is equivalent to a velocity change of $\approx 9\,\mathrm{cm/s}$ after $1\,\mathrm{min}$. In this context it's not a big effect and can safely be ignored. There are other contexts in which it is anything but negligible, however.

1.5 Centrifugal Force

According to Eq. (1.3.5), the acceleration due to the centrifugal force is

$$\mathbf{a}_{\mathrm{cf}} = -\boldsymbol{\omega} \times (\boldsymbol{\omega} \times \mathbf{r})\,. \tag{1.5.1}$$

. To calculate this, it is useful to note the general identity for the triple cross product,

$$\mathbf{a} \times (\mathbf{b} \times \mathbf{c}) = (\mathbf{a} \cdot \mathbf{c})\,\mathbf{b} - (\mathbf{a} \cdot \mathbf{b})\,\mathbf{c}. \tag{1.5.2}$$



Figure 1.2: Gaspard-Gustave de Coriolis (1792–1843) (left), Léon Foucault (1819–68) (centre) and Sir Joseph Larmor (1857–1942) (right)

Using the same coordinates as in Section 1.4, with $\mathbf{r} = r\hat{\mathbf{k}}$, we find

$$\begin{aligned} \mathbf{a}_{\rm cf} &= (\boldsymbol{\omega}\cdot\boldsymbol{\omega})\mathbf{r} - (\boldsymbol{\omega}\cdot\mathbf{r})\boldsymbol{\omega} = \omega^2 r\left(0, -\sin\theta\cos\theta, \sin^2\theta\right) \\ &= \omega^2 r\sin\theta(0, -\cos\theta, \sin\theta). \end{aligned} \tag{1.5.3}$$

The acceleration points away from the rotation axis, and for example in London, it has the strength

$$a_{\rm cf} = \omega^2 r\sin\theta \approx (7.3\times10^{-5}\,{\rm s}^{-1})^2 \times 6.4\times10^6\,{\rm m}\times\sin(38.5^\circ) \approx 0.02\,{\rm m/s}^2. \tag{1.5.4}$$

Because this force is independent of velocity, we cannot distinguish it locally from the gravitational force, and therefore it acts essentially as a small correction to it, changing not only the apparent strength but also the direction of the gravitational force. In other situations the centrifugal force can be very important, especially when rotation speeds are high.

It is useful to consider the special case of a particle at rest in the rotating frame, so that the left-hand-side of Eq. (1.3.5) vanishes and $\mathbf{v}_{\rm R} = 0$ as well. In that case, Eq. (1.3.5) implies that there has to be a real physical force

$$\mathbf{F} = m\boldsymbol{\omega}\times(\boldsymbol{\omega}\times\mathbf{r})\,. \tag{1.5.5}$$

This is known as the *centripetal force*, and it is the total net force acting on the particle to keep it fixed in the rotating frame. Because it is a real force, it has to due to some type of physical interaction, such as gravitational or electrostatic force, a support force provided by, e.g., a rope, or usually a combination of several different physical forces.

1.6 Examples

1.6.1 Weather

Probably the most important effect attributed to the Coriolis effect is in meteorology: Winds don't flow from areas of high pressure to those of low pressure but instead tend to flow round the minima and maxima of pressure, giving rise to cyclones and anticyclones respectively. A simple way to understand this is to consider a simple uniform pressure gradient in the presence of a Coriolis force, giving an equation of motion such as

$$\frac{{\rm d}^2\mathbf{r}}{{\rm d}t^2} = -\boldsymbol{\nabla}p - 2\boldsymbol{\omega}\times\mathbf{v}\,. \tag{1.6.1}$$

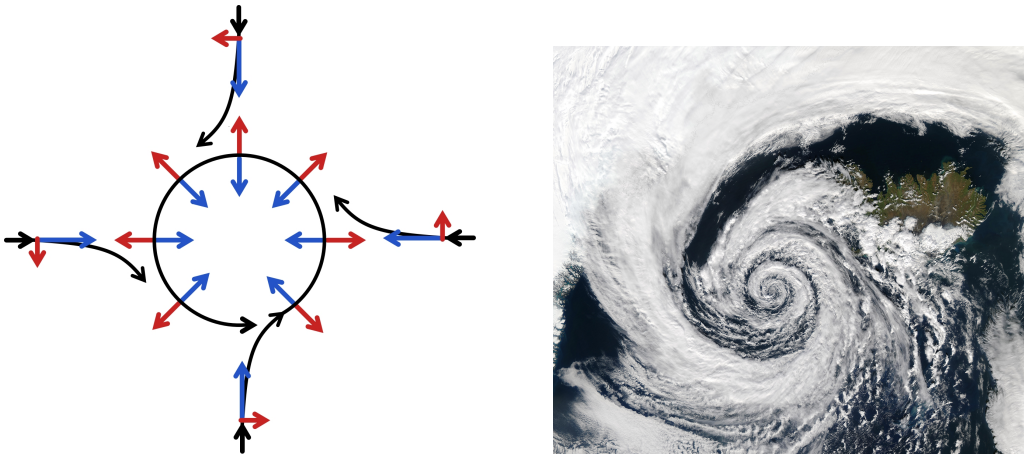

Figure 1.3: Left: schematic representation of flow around a low-pressure area in the Northern hemisphere. The pressure-gradient force is represented by blue arrows, the Coriolis acceleration (always perpendicular to the velocity) by red arrows. Right: this low pressure system over Iceland spins counter-clockwise due to balance between the Coriolis force and the pressure gradient force.

The general solution of such a problem is complicated. However, if we confine ourselves to 2–dimensions and ignore the component of $\boldsymbol{\omega}$ parallel to the surface, just as we did in our discussion of the Foucault pendulum, we can always find a solution of (1.6.1) with a constant velocity such that the 2 terms on the right cancel. In such a solution $\boldsymbol{\nabla} p$ must be perpendicular to $\mathbf{v}$, as a cross product is always perpendicular to both vectors. Hence (1.6.1) has a solution in which the velocity is perpendicular to the pressure gradient. In such a system the wind always follows the isobars (lines of constant pressure), a pressure minimum is not easily filled and a cyclone (or anticyclone) is stable. This is illustrated in Fig. 1.3a.

1.6.2 Foucault's Pendulum

Consider a pendulum which is free to move in any direction and is sufficiently long and heavy that it will swing freely for several hours. Ignoring the vertical component both of the pendulum's motion and of the Coriolis force, the equations of motion for the bob (in the coordinate system described above) are

$$\ddot{x} = -\frac{g}{\ell}x + 2\omega\cos\theta\dot{y}\,, \tag{1.6.2}$$

$$\ddot{y} = -\frac{g}{\ell}y - 2\omega\cos\theta\dot{x}\,, \tag{1.6.3}$$

or, using the complex number trick from Section 1.6.3 with $\tilde{r} = x + \mathrm{i}y$,

$$\frac{\mathrm{d}^2\tilde{r}}{\mathrm{d}t^2} + 2\mathrm{i}\Omega\frac{\mathrm{d}\tilde{r}}{\mathrm{d}t} + \omega_0^2\tilde{r} = 0\,, \tag{1.6.4}$$

where $\Omega = \omega\cos\theta$ and $\omega_0^2 = g/\ell$. Using standard methods for second order differential equations we obtain the general solution

$$\tilde{r} = A\mathrm{e}^{-\mathrm{i}(\Omega-\omega_1)t} + B\mathrm{e}^{-\mathrm{i}(\Omega+\omega_1)t}\,, \tag{1.6.5}$$

where $\omega_1^2 = \omega_0^2 + \Omega^2$. In particular, if the pendulum is released from the origin with velocity $(v_0, 0)$, we have $A = -B = v_0/2\mathrm{i}\omega_1$, so that[2]

$$\tilde{r} = \frac{v_0}{\omega_1} \mathrm{e}^{-\mathrm{i}\Omega t} \sin \omega_1 t, \tag{1.6.6}$$

which means in terms of the original variables x and y,

$$\begin{aligned} x &= \frac{v_0}{\omega_1} \cos \Omega t \sin \omega_1 t \,, \\ y &= -\frac{v_0}{\omega_1} \sin \Omega t \sin \omega_1 t \,. \end{aligned} \tag{1.6.7}$$

We can also write this in terms of polar co-ordinates, (ρ, ϕ), as

$$\rho = \frac{v_0}{\omega_1} \sin \omega_1 t \qquad \phi = -\Omega t \quad .$$

As $\Omega \ll \omega_0$, the period of oscillation is much less than a day, the result is easy to interpret: the pendulum swings with a basic angular frequency ω_1 ($\approx \omega_0$) but the plane of oscillation rotates with angular frequency Ω. At the pole, $\theta = 0$ the plane of the pendulum apparently rotates once a day. In other words, the plane doesn't rotate at all but the Earth rotates under it once a day. On the other hand, at the equator, $\theta = \frac{\pi}{2}$ and $\Omega = 0$ the plane of the pendulum is stable. In South Kensington

$$\Omega = \omega \cos\theta = (7.292 \times 10^{-5}\,\mathrm{s}^{-1}) \times \cos(38.5^\circ) \Rightarrow T = 30.58\,\mathrm{hr}\,. \tag{1.6.8}$$

Note that this is the time for a complete rotation of the plane of the pendulum through 360°. However, after it has rotated through 180° it would be hard to tell the difference between that and the starting position.

A working Foucault pendulum may be seen in the Science Museum.

1.6.3 Particle in Magnetic Field

The equation of motion for a charged particle in a magnetic field takes the form

$$m\frac{\mathrm{d}^2\mathbf{r}}{\mathrm{d}t^2} = q\mathbf{v} \times \mathbf{B}, \tag{1.6.9}$$

which we can write as

$$\frac{\mathrm{d}\mathbf{v}}{\mathrm{d}t} = -\left(\frac{q}{m}\mathbf{B}\right) \times \mathbf{v}. \tag{1.6.10}$$

If we identify $\boldsymbol{\omega}$ with $q\mathbf{B}/m$, we note that this has same form as Eq. (1.2.5) for a velocity which is constant in the inertial frame. Hence we should expect the motion of a particle in a magnetic field to be similar to motion of a free particle in a rotating frame with $\mathbf{omega} = q\mathbf{B}/m$, namely circular motion with angular frequency ω. In this case this is called the *cyclotron* motion and $\omega = qB/m$ is the cyclotron frequency.

We can check this by solving the equation of motion. We first choose a coordinate system such that $\boldsymbol{\omega}$ (or equivalently $\mathbf{B}$) is in the $\hat{\mathbf{k}}$ direction. Then, since $\boldsymbol{\omega} \times \hat{\mathbf{k}} = 0$, the z component of $\mathbf{v}$ is constant, and the x and y components satisfy

$$\frac{\mathrm{d}v_x}{\mathrm{d}t} = +\omega v_y, \quad \frac{\mathrm{d}v_y}{\mathrm{d}t} = -\omega v_x. \tag{1.6.11}$$

[2] N.B. There is an error in the example given in K & B p 117.

Define a complex variable $\tilde{v} = v_x + \mathrm{i}v_y$, these can be written as a single equation

$$\frac{\mathrm{d}\tilde{v}}{\mathrm{d}t} = -\mathrm{i}\omega\tilde{v}, \tag{1.6.12}$$

which is easy to solve,

$$\tilde{v} = \tilde{v}_0 \exp(-\mathrm{i}\omega t)\,. \tag{1.6.13}$$

Finally we obtain v_x and v_y by taking the real and imaginary parts of $\tilde{v}$.

We can take this analogy further by generalising Eq. (1.6.10) to the full Lorentz force, including an electric field $\mathbf{E}$ to obtain

$$m\frac{\mathrm{d}\mathbf{v}}{\mathrm{d}t} = q\mathbf{v}\times\mathbf{B} + q\mathbf{E}\,. \tag{1.6.14}$$

where we immediately see that $\frac{q}{m}\mathbf{E}$ is analogous to the rate of change of the velocity in the inertial frame. If we consider now the simple case in which $\mathbf{E}$ is in the x–direction and $\mathbf{B}$ in the z–direction, we can write the x and y components of Eq. (1.6.14) in the form

$$\begin{aligned} m\frac{\mathrm{d}v_x}{\mathrm{d}t} &= qv_yB_z + qE_x \\ m\frac{\mathrm{d}v_y}{\mathrm{d}t} &= -qv_xB_z\,. \end{aligned} \tag{1.6.15}$$

Using the simple transformation $v_y' = v_y - (E_x/B_z)$, Eq. (1.6.15) becomes

$$\begin{aligned} m\frac{\mathrm{d}v_x}{\mathrm{d}t} &= qv_y'B_z \\ m\frac{\mathrm{d}v_y'}{\mathrm{d}t} &= -qv_xB_z\,, \end{aligned} \tag{1.6.16}$$

which is the same as Eq. (1.6.10). Hence the complete solution is a circular motion with an additional drift with speed E_x/B_z in the y–direction, perpendicular to both $\mathbf{E}$ and $\mathbf{B}$.

More generally, whenever we find that when the time derivative of any vector is of the form (1.2.5), we can immediately tell that it rotates with the corresponding angular velocity $\boldsymbol{\omega}$.

1.6.4 Larmor Effect

It is sometimes useful to consider a rotating frame, not because the system is itself rotating, but because it helps to simplify the mathematics. In this sense it is similar to choosing an appropriate coordinate system.

Consider a particle of charge q moving around a fixed point charge $-q'$ in a uniform magnetic field $\mathbf{B}$. The equation of motion is

$$m\frac{\mathrm{d}^2\mathbf{r}}{\mathrm{d}t^2} = -\frac{k}{r^2}\hat{\mathbf{r}} + q\frac{\mathrm{d}\mathbf{r}}{\mathrm{d}t}\times\mathbf{B}\,, \tag{1.6.17}$$

where $k = qq'/4\pi\epsilon_0$.

Rewriting (1.6.17) in a rotating frame, we obtain

$$\left.\frac{\mathrm{d}^2\mathbf{r}}{\mathrm{d}t^2}\right|_{\mathrm{R}} + 2\boldsymbol{\omega}\times\left.\frac{\mathrm{d}\mathbf{r}}{\mathrm{d}t}\right|_{\mathrm{R}} + \boldsymbol{\omega}\times(\boldsymbol{\omega}\times\mathbf{r}) = -\frac{k}{mr^2}\hat{\mathbf{r}} + \frac{q}{m}\left(\left.\frac{\mathrm{d}\mathbf{r}}{\mathrm{d}t}\right|_{\mathrm{R}} + \boldsymbol{\omega}\times\mathbf{r}\right)\times\mathbf{B}\,. \tag{1.6.18}$$

If we choose $\boldsymbol{\omega} = -(q/2m)\,\mathbf{B}$, the terms in the velocity fall out and we are left with

$$\left.\frac{\mathrm{d}^2\mathbf{r}}{\mathrm{d}t^2}\right|_{\mathrm{R}} = -\frac{k}{mr^2}\hat{\mathbf{r}} + \left(\frac{q}{2m}\right)^2 \mathbf{B}\times(\mathbf{B}\times\mathbf{r})\ . \tag{1.6.19}$$

In a weak magnetic field we may ignore terms in B^2, such as the last term in (1.6.19), so that we are left with an expression which is identical to that of the system without the magnetic field. What does this mean?

The solution of (1.6.19) is an ellipse (as shown by Newton, Kepler, etc.). Hence, the solution in the inertial frame must be the same ellipse but rotating with the angular frequency $\boldsymbol{\omega} = -(q/2m)\,\mathbf{B}$. If this is smaller than the period of the ellipse then the effect is that the major axis of the ellipse slowly rotates. Such a behaviour is known as *precession.* We shall see other examples of this later.

Note that there are some confusing diagrams on the internet and in textbooks which purport to illustrate this precession. There are two special cases to consider: firstly when the magnetic field is perpendicular to the plane of the ellipse. In this case the major axis of the ellipse rotates about the focus, while remaining in the same plane. In the second special case the magnetic field is in the plane of the ellipse. Here the ellipse rotates about an axis through the focus and perpendicular to the major axis. For a circular orbit, the first case the orbit would remain circular, but with slightly different periods clockwise and anti-clockwise, whereas the second would behave like a plate rotating while standing on its edge.

Chapter 2

Rigid Bodies

2.1 Many-Body Systems[1]

Let us consider a system consisting N particles, which we label by an integer $a = 1, \ldots, N$ (or other letters $b, c, \ldots$ at the start of the alphabet). We denote their positions by $\mathbf{r}_a$ and masses by m_a.

The net force acting on each particle, is the sum of the forces due to each other particle in the system as well as any external forces. Therefore Newton's second law for particle a has the form

$$m_a \ddot{\mathbf{r}}_a = \sum_{b \neq a} \mathbf{F}_{ab} + \mathbf{F}_a^{\text{ext}}, \tag{2.1.1}$$

where $\mathbf{F}_{ab}$ is the force on particle a due to particle b, and $\mathbf{F}_a^{\text{ext}}$ is the external force acting on particle a. Note that the external force is generally dependent on the particle's position, velocity etc. and is therefore different for each particle, which is why it has the index a.

Instead of trying to solve the motion of each particle, let us first look at the motion of the system as a whole. For that purpose, it is useful to define the total mass

$$M = \sum_a m_a, \tag{2.1.2}$$

and the centre or mass

$$\mathbf{R} = \frac{1}{M} \sum_a m_a \mathbf{r}_a. \tag{2.1.3}$$

We also define the total momentum of the system as the sum of the momenta of the individual particles,

$$\mathbf{P} = \sum_a \mathbf{p}_a = \sum_a m_a \dot{\mathbf{r}} = M \dot{\mathbf{R}}. \tag{2.1.4}$$

Differentiating this with respect to time gives

$$\dot{\mathbf{P}} = \sum_a m_a \ddot{\mathbf{r}} = \sum_{ab} \mathbf{F}_{ab} + \sum_a \mathbf{F}_a^{\text{ext}} = \sum_a \mathbf{F}_a^{\text{ext}}, \tag{2.1.5}$$

where the sum of the inter-particle forces $\mathbf{F}_{ab}$ vanishes because of Newton's third law $\mathbf{F}_{ba} = -\mathbf{F}_{ab}$. Therefore the rate of change of the total momentum is given by the total external force. In particular, the total momentum $\mathbf{P}$ is conserved in isolated systems, i.e., when there are no external forces.

[1] Kibble & Berkshire, chapter 9

Similarly, we define the total angular momentum $\mathbf{L}$ as the sum of angular momenta $\mathbf{l}_a = m_a \mathbf{r}_a \times \dot{\mathbf{r}}_a$ of the individual particles,

$$\mathbf{L} = \sum_a \mathbf{l}_a = \sum_a m_a \mathbf{r}_a \times \dot{\mathbf{r}}_a \,. \tag{2.1.6}$$

(N.B. Kibble & Berkshire use $\mathbf{J}$ for the angular momentum. We shall stick to the more conventional $\mathbf{L}$ here.) Its rate of change is given by

$$\begin{aligned} \dot{\mathbf{L}} &= \sum_a m_a \dot{\mathbf{r}}_a \times \dot{\mathbf{r}}_a + \sum_a m_a \mathbf{r}_a \times \ddot{\mathbf{r}}_a = \sum_a m_a \mathbf{r}_a \times \ddot{\mathbf{r}}_a \\ &= \sum_a \mathbf{r}_a \times \left(\sum_b \mathbf{F}_{ab} + \mathbf{F}_a^{\text{ext}} \right). \end{aligned} \tag{2.1.7}$$

Using Newton's third law, we can write this as

$$\dot{\mathbf{L}} = \frac{1}{2} \sum_{ab} (\mathbf{r}_a - \mathbf{r}_b) \times \mathbf{F}_{ab} + \sum_a \mathbf{r}_a \times \mathbf{F}_a^{\text{ext}}. \tag{2.1.8}$$

In general, the first term on the right-hand-side is non-zero, but it vanishes if we assume that the inter-particle forces are *central*, which means that the force $\mathbf{F}_{ab}$ between particles i and j is in the direction of their separation vector $(\mathbf{r}_a - \mathbf{r}_b)$. In that case we have

$$\dot{\mathbf{L}} = \sum_a \mathbf{r}_a \times \mathbf{F}_a^{\text{ext}} \equiv \boldsymbol{\tau}, \tag{2.1.9}$$

where the right-hand-side is known as the *torque*. The torque is zero and the angular momentum is conserved if the system is isolated or if the external forces all point to the origin (i.e., $\mathbf{F}_a^{\text{ext}} || \mathbf{r}_a$).

Note that there are some forces that are not central, such as the electromagnetic force between moving charges, and in that case $\mathbf{L}$ is not conserved. (In fact, the electromagnetic field can carry angular momentum, and when it is included the total angular momentum is still conserved.)

It is often useful to separate the coordinates $\mathbf{r}_a$ into centre of mass and relative contributions

$$\mathbf{r}_a = \mathbf{R} + \mathbf{r}_a^* \tag{2.1.10}$$

such that, by definition,

$$\sum_a m_a \mathbf{r}_a^* = 0 \,. \tag{2.1.11}$$

Substituting this into Eq. (2.1.6) gives

$$\begin{aligned} \mathbf{L} &= \sum_a m_a \left(\mathbf{R} + \mathbf{r}_a^* \right) \times \left(\dot{\mathbf{R}} + \dot{\mathbf{r}}_a^* \right) \\ &= \left(\sum_a m_a \right) \mathbf{R} \times \dot{\mathbf{R}} + \left(\sum_a m_a \mathbf{r}_a^* \right) \times \dot{\mathbf{R}} + \mathbf{R} \times \left(\sum_a m_a \dot{\mathbf{r}}_a^* \right) + \sum_a m_a \mathbf{r}_a^* \times \dot{\mathbf{r}}_a^* \\ &= M \mathbf{R} \times \dot{\mathbf{R}} + \mathbf{L}^* \qquad \text{where} \qquad \mathbf{L}^* = \sum_a m_a \mathbf{r}_a^* \times \dot{\mathbf{r}}_a^* \,, \end{aligned} \tag{2.1.12}$$

as the other terms are zero due to Eq. (2.1.11). The angular momentum can be separated into a centre of mass part, $M\mathbf{R} \times \dot{\mathbf{R}}$, and the angular momentum about the centre of mass, $\mathbf{L}^*$.

The rate of change of the relative angular momentum $\mathbf{L}^*$ may be written as the sum of the moments of the particles about the centre of mass due to external forces alone

$$\begin{aligned}\dot{\mathbf{L}}^* &= \dot{\mathbf{L}} - \frac{\mathrm{d}}{\mathrm{d}t}\left(M\mathbf{R}\times\dot{\mathbf{R}}\right) = \sum_a \mathbf{r}_a \times \mathbf{F}_a^{\mathrm{ext}} - M\mathbf{R}\times\ddot{\mathbf{R}} \\ &= \sum_a (\mathbf{r}_a - \mathbf{R}) \times \mathbf{F}_a^{\mathrm{ext}} = \sum_a \mathbf{r}_a^* \times \mathbf{F}_a^{\mathrm{ext}}. \end{aligned} \tag{2.1.13}$$

This means that we can often study the centre-of-mass motion and the relative motion separately from each other. In particular, if the external forces are position-independent, the relative angular momentum evolves independently of the centre-of-mass motion.

Likewise, the total kinetic energy separates into the kinetic energy of the centre of mass and the kinetic energy relative to the centre of mass,

$$T = \frac{1}{2}\sum_a m_a \dot{\mathbf{r}}_a^2 = \frac{1}{2}\sum_a m_a \left(\dot{\mathbf{R}} + \dot{\mathbf{r}}_a^*\right)\cdot\left(\dot{\mathbf{R}} + \dot{\mathbf{r}}_a^*\right) = \frac{1}{2}M\dot{\mathbf{R}}^2 + \frac{1}{2}\sum_a m_a \dot{\mathbf{r}}_a^{*2}. \tag{2.1.14}$$

The above results apply generally to many-body systems, but for the rest of this Section we will focus on a special class of them known as *rigid bodies*. These are many-body systems in which all distances $|\mathbf{r}_a - \mathbf{r}_b|$ between particles are fixed. The whole system can still move and rotate. In reality, a rigid body a mathematical idealisation because it requires infinitely strong forces between particles, but in many cases it is a very good approximation.

2.2 Rotation about a Fixed Axis

We first consider the case in which the body is free to rotate about a fixed axis. Cylindrical polar coordinates (ρ, ϕ, z) are ideally suited for this situation (see Fig. 1.1). If we choose the z axis as the rotation axis, then for every particle a the coordinate z_a and ρ_a are fixed and only the angular coordinate ϕ_a changes as $\dot{\phi}_a = \omega$. Then we can write the z component of the angular momentum (2.1.6) as

$$L_z = \sum_a m_a \rho_a \left(\rho_a \dot{\phi}\right) = \sum_a m_a \rho_a^2 \omega = I\omega\,, \tag{2.2.1}$$

where $\rho\dot{\phi}$ is the tangential velocity and $I = \sum_a m_a\rho_a^2$ is the *moment of inertia* about the axis. As I is obviously constant we can write its rate of change as

$$\dot{L}_z = I\dot{\omega} = \sum_a \rho_a F_{a\phi}, \tag{2.2.2}$$

where $F_{a\phi}$ is the component of the external force $\mathbf{F}_a^{\mathrm{ext}}$ in the $\hat{\phi}$ direction.

Similarly we can write the kinetic energy in terms of I and ω as

$$T = \sum_a \tfrac{1}{2} m_a \left(\rho_a \dot{\phi}_a\right)^2 = \tfrac{1}{2} I\omega^2\,. \tag{2.2.3}$$

Note the similarity of these expressions to the corresponding linear ones where $m \mapsto I$ and $v \mapsto \omega$

$$p = mv \qquad \leftrightarrow \qquad L = I\omega \tag{2.2.4}$$

$$T = \tfrac{1}{2}mv^2 \qquad \leftrightarrow \qquad T = \tfrac{1}{2}I\omega^2\,. \tag{2.2.5}$$

Of course, there is no reason why the axis should be through the centre of mass, and for example in a pendulum it is not. If we define the origin to be on the axis then we can define $\mathbf{R}$ as the distance of the centre of mass from that axis. In general the axis would be free to move, so in order for it to remain fixed, there must be a support force $\mathbf{Q}$ that prevents it from moving. From brevity, we refer to it as the "force at axis". Denoting the sum of all other external forces by $\mathbf{F}$, we can write Newton's second law as

$$\dot{\mathbf{P}} = M\ddot{\mathbf{R}} = \mathbf{Q} + \mathbf{F}\,. \tag{2.2.6}$$

Using $\dot{\mathbf{R}} = \boldsymbol{\omega} \times \mathbf{R}$ we can write

$$\ddot{\mathbf{R}} = \dot{\boldsymbol{\omega}} \times \mathbf{R} + \boldsymbol{\omega} \times \dot{\mathbf{R}} = \dot{\boldsymbol{\omega}} \times \mathbf{R} + \boldsymbol{\omega} \times (\boldsymbol{\omega} \times \mathbf{R})\,. \tag{2.2.7}$$

The first of these is the tangential acceleration and the second one is the centripetal force which keeps the centre of mass on a circular trajectory. From these equations we can determine the support force $\mathbf{Q}$ required to keep the axis fixed.

2.2.1 Compound Pendulum

As an example, let us consider a *compound pendulum*, which is a rigid body attached to a pivot and subject to a gravitational force. We take the z–axis to be the axis of rotation, which is now horizontal, and $\hat{\mathbf{x}}$ to be pointing downwards. Then the pendulum is subject to an external gravitational force $\mathbf{F} = (Mg, 0, 0)$ acting through its centre of mass. In terms of the unit vectors $\hat{\boldsymbol{\rho}}$ and $\hat{\boldsymbol{\phi}}$, we can write this as

$$\mathbf{F} = Mg\cos\phi\hat{\boldsymbol{\rho}} - Mg\sin\phi\hat{\boldsymbol{\phi}}. \tag{2.2.8}$$

Thus the equation of motion (2.2.2) is

$$I\ddot{\phi} = -MgR\sin\phi, \tag{2.2.9}$$

and the energy conservation equation is

$$E = T + V = \tfrac{1}{2}I\dot{\phi}^2 - MgR\cos\phi = \text{constant}\,. \tag{2.2.10}$$

For small amplitudes, $\phi \ll 1$, Eq. (2.2.9) reduces to the equation for a simple harmonic oscillator,

$$\ddot{\phi} = -\frac{MgR}{I}\phi, \tag{2.2.11}$$

with period $T = 2\pi\sqrt{I/MgR}$.

Rewriting Eq. (2.2.7) in polar coordinates and noting that only ϕ actually changes we can calculate the net force on the system and hence the support force $\mathbf{Q}$ at the axis,

$$\dot{\mathbf{P}} = M\ddot{\mathbf{R}} = MR\ddot{\phi}\,\hat{\boldsymbol{\phi}} - MR\dot{\phi}^2\,\hat{\boldsymbol{\rho}} \tag{2.2.12}$$

$$\Rightarrow \mathbf{Q} = \dot{\mathbf{P}} - \mathbf{F} = \left(-Mg\cos\phi - MR\dot{\phi}^2\right)\hat{\boldsymbol{\rho}} + \left(Mg\sin\phi + MR\ddot{\phi}\right)\hat{\boldsymbol{\phi}} \tag{2.2.13}$$

$$= -\left[Mg\cos\phi\left(1 + \frac{2MR^2}{I}\right) + \frac{2MR}{I}E\right]\hat{\boldsymbol{\rho}} + Mg\sin\phi\left(1 - \frac{MR^2}{I}\right)\hat{\boldsymbol{\phi}}\,, \tag{2.2.14}$$

where, in the final step, we have substituted from Eqs. (2.2.10) and (2.2.9) to eliminate $\dot{\phi}$ and $\ddot{\phi}$. Note that, in contrast with a simple pendulum (for which $I = MR^2$), the force $\mathbf{Q}$ is not in the radial direction $\hat{\boldsymbol{\rho}}$.

2.2.2 Centre of Percussion

As a further example, consider a compound pendulum which is initially at rest. An external force $\mathbf{F}$ in the angular direction $\hat{\boldsymbol{\phi}}$ is then applied at distance d from the pivot point for a short period of time. We want to calculate the force $\mathbf{Q}$ needed to keep the axis fixed.

To make this more concrete, you can think of the pendulum as a tennis racket which you are holding in your hand, so that your hand acts as the pivot point. A ball hits the racket at distance d from your hand and exerts a force $\mathbf{F}$ on the racket. We want to calculate the support force $\mathbf{Q}$ which your hand has to provide to remain stationary, or equivalently the impact you will feel with your hand.

Because initially the racket is not rotating, $\dot{\phi} = 0$, and therefore Eqs. (2.2.6) and (2.2.7) become

$$\begin{aligned} \mathbf{Q} + \mathbf{F} &= \dot{\mathbf{P}} = M\ddot{\mathbf{R}} = MR\ddot{\phi}\hat{\boldsymbol{\phi}}, \\ I\ddot{\phi} &= dF_\phi, \end{aligned} \tag{2.2.15}$$

from which we find

$$Q_\phi = MR\ddot{\phi} - F_\phi = \frac{MRd}{I}F_\phi - F_\phi = \left(\frac{MRd}{I} - 1\right)F_\phi. \tag{2.2.16}$$

If the distance at which the ball hits the racket is $d = I/MR$, the linear and rotational motion balance each other and the pivot point does not feel any impact. This point is known as the *centre of percussion*. In sport it is also called the "sweet spot", because you hit the ball but feel no impact with your hand.

2.3 Inertia Tensor

In general, the angular momentum vector

$$\mathbf{L} = \sum_a m_a \mathbf{r}_a \times \dot{\mathbf{r}}_a = \sum_a m_a \mathbf{r}_a \times (\omega \times \mathbf{r}_a) \tag{2.3.1}$$

is not parallel to the angular velocity ω.

Let us use the Cartesian coordinates and write the position vector as

$$\mathbf{r} = \begin{pmatrix} x \\ y \\ z \end{pmatrix}. \tag{2.3.2}$$

For simplicity, we first assume that ω is in the z direction, so that

$$\boldsymbol{\omega} = \omega\hat{\mathbf{k}} = \begin{pmatrix} 0 \\ 0 \\ \omega \end{pmatrix}. \tag{2.3.3}$$

Then we have

$$\mathbf{r} \times (\boldsymbol{\omega} \times \mathbf{r}) = (\mathbf{r}\cdot\mathbf{r})\boldsymbol{\omega} - (\mathbf{r}\cdot\boldsymbol{\omega})\mathbf{r} = (x^2+y^2+z^2)\boldsymbol{\omega} - z\omega\mathbf{r} = \begin{pmatrix} -xz\omega \\ -yz\omega \\ (x^2+y^2)\omega \end{pmatrix}. \tag{2.3.4}$$

Using this in Eq. (2.3.1), we find the components of the angular momentum vector $\mathbf{L}$,

$$\begin{aligned} L_x &= -\sum_a m_a x_a z_a \omega, \\ L_y &= -\sum_a m_a y_a z_a \omega, \\ L_z &= \sum_a m_a \left(x_a^2 + y_a^2\right) \omega \,. \end{aligned} \tag{2.3.5}$$

We can summarise these by writing

$$L_x = I_{xz}\omega, \qquad L_y = I_{yz}\omega, \qquad L_z = I_{zz}\omega, \tag{2.3.6}$$

where

$$I_{xz} = -\sum_a m_a x_a z_a, \qquad I_{yz} = -\sum_a m_a y_a z_a, \qquad I_{zz} = \sum_a m_a \left(x_a^2 + y_a^2\right) . \tag{2.3.7}$$

I_{zz} is the moment of inertia as previously defined. I_{xz} and I_{yz} are sometimes known as *products of inertia.*

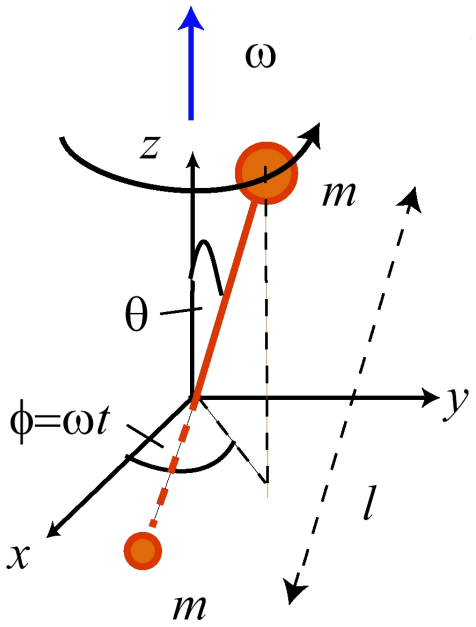

Figure 2.1:

As an example of a simple system for which the angular momentum is not parallel the angular velocity, consider a rigid rod with equal masses on either end (a dumbbell) inclined at an angle θ to the axis of rotation. If the masses are at $\pm\mathbf{r}$ then the total angular momentum is

$$\mathbf{L} = m\mathbf{r} \times \dot{\mathbf{r}} + m(-\mathbf{r}) \times (-\dot{\mathbf{r}}) = 2m\mathbf{r} \times (\boldsymbol{\omega} \times \mathbf{r}) \,, \tag{2.3.8}$$

which is clearly in a direction perpendicular to $\mathbf{r}$. Because it is also perpendicular to the vector $(\boldsymbol{\omega} \times \mathbf{r})$, it has to lie on the plane spanned by $\boldsymbol{\omega}$ and $\mathbf{r}$, and therefore it is rotating around the axis $\boldsymbol{\omega}$ together with the rod.

Using Eq. (2.3.5), we can write the components of the angular momentum as

$$\mathbf{L} = \begin{pmatrix} -2mxz\omega \\ -2myz\omega \\ 2m(x^2+y^2)\omega \end{pmatrix} = 2m\rho\omega \begin{pmatrix} -z\cos\phi \\ -z\sin\phi \\ \rho \end{pmatrix}. \tag{2.3.9}$$

For completeness, let us write down the angular momentum vector $\mathbf{L}$ for a general angular velocity $\boldsymbol{\omega}$. We have

$$\begin{aligned} \mathbf{L} &= \sum_a \left[(\mathbf{r}_a\cdot\mathbf{r}_a)\boldsymbol{\omega} - (\mathbf{r}_a\cdot\boldsymbol{\omega})\mathbf{r}_a\right] \\ &= \sum_a m_a \left[(x_a^2+y_a^2+z_a^2)\begin{pmatrix}\omega_x\\ \omega_y\\ \omega_z\end{pmatrix} - (x_a\omega_x + y_a\omega_y + z_a\omega_z)\begin{pmatrix}x_a\\ y_a\\ z_a\end{pmatrix}\right] \\ &= \sum_a m_a \begin{pmatrix} (y_a^2+z_a^2)\omega_x - x_a y_a\omega_y - x_a z_a\omega_z \\ -x_a y_a\omega_x + (x_a^2+z_a^2)\omega_y - y_a z_a\omega_z \\ -x_a z_a\omega_x - y_a z_a\omega_y + (x_a^2+y_a^2)\omega_z \end{pmatrix}. \end{aligned} \tag{2.3.10}$$

Using linear algebra, we can write this as a product of a matrix and a vector

$$\mathbf{L} = \sum_a m_a \begin{pmatrix} y_a^2+z_a^2 & -x_a y_a & -x_a z_a \\ -x_a y_a & x_a^2+z_a^2 & -y_a z_a \\ -x_a z_a & -y_a z_a & x_a^2+y_a^2 \end{pmatrix} \begin{pmatrix}\omega_x\\ \omega_y\\ \omega_z\end{pmatrix}, \tag{2.3.11}$$

or more concisely

$$\mathbf{L} = \bar{\bar{\mathbf{I}}}\cdot\boldsymbol{\omega}, \tag{2.3.12}$$

where the three-by-three matrix

$$\bar{\bar{\mathbf{I}}} = \begin{pmatrix} I_{xx} & I_{xy} & I_{xz} \\ I_{yx} & I_{yy} & I_{yz} \\ I_{zx} & I_{zy} & I_{zz} \end{pmatrix} = \sum_a m_a \begin{pmatrix} y_a^2+z_a^2 & -x_a y_a & -x_a z_a \\ -x_a y_a & x_a^2+z_a^2 & -y_a z_a \\ -x_a z_a & -y_a z_a & x_a^2+y_a^2 \end{pmatrix} \tag{2.3.13}$$

is known as the *inertia tensor*. In general, a tensor is a geometric object that describes a linear relation between two or more vectors vectors. In this case, the inertia tensor describes the linear relation between $\boldsymbol{\omega}$ and $\mathbf{L}$, and can be represented by a three-by-three matrix. Just like the components of a vector, the elements of the matrix $\bar{\bar{\mathbf{I}}}$ change under rotations. For more details, see Appendix A.9 in Kibble&Berkshire.

Finally, it is often convenient to work in the component notation. Labelling the coordinates x, y and z by $i, j \in \{1,2,3\}$, we can write the components of the inertia tensor in a compact form as

$$I_{ij} = \sum_a m_a \left(\mathbf{r}_a^2\delta_{ij} - r_{ai}r_{aj}\right), \tag{2.3.14}$$

where δ_{ij} is the Kronecker delta (that is, $\delta_{ij}=1$ if $i=j$ and $\delta_{ij}=0$ if $i\neq j$), and r_{ai} is the ith component of the position vector $\mathbf{r}_a$. In the component notation, Eq. (2.3.12) becomes

$$L_i = \sum_j I_{ij}\omega_j. \tag{2.3.15}$$

2.4 Principal Axes of Inertia

We can make use of our knowledge of the properties of matrices to understand the meaning of the inertia tensor $\overline{\overline{\mathbf{I}}}$. We note that $\overline{\overline{\mathbf{I}}}$ is symmetric, $I_{xy} = I_{yx}$, so that the eigenvalues of $\overline{\overline{\mathbf{I}}}$ are real. We denote these eigenvalues by I_1, I_2 and I_3 and call them the *principal moments of inertia*. The corresponding eigenvectors, which we denote by $\hat{\mathbf{e}}_1$, $\hat{\mathbf{e}}_2$ and $\hat{\mathbf{e}}_3$ are called the *principal axes of inertia*. They are orthogonal to each other, and we choose them to be unit vectors. By definition, the eigenvectors satisfy

$$\overline{\overline{\mathbf{I}}} \cdot \hat{\mathbf{e}}_i = I_i \hat{\mathbf{e}}_i, \tag{2.4.1}$$

where again $i \in \{1, 2, 3\}$. This also means that is the angular velocity $\boldsymbol{\omega}$ is parallel to a principal axis, then the angular momentum $\mathbf{L}$ is parallel to it.

It is convenient to work in a coordinate system based on the principal axes, and write

$$\boldsymbol{\omega} = \sum_i \omega_i \hat{\mathbf{e}}_i. \tag{2.4.2}$$

The angular momentum is then

$$\mathbf{L} = \sum_i I_i \omega_i \hat{\mathbf{e}}_i. \tag{2.4.3}$$

It is important to note that the principal axes rotate with the body. They therefore represent a rotating frame of reference (see Chapter 1).

Using the identity $(\mathbf{a} \times \mathbf{b}) \cdot \mathbf{c} = \mathbf{a} \cdot (\mathbf{b} \times \mathbf{c})$, the kinetic energy can be expressed as

$$\begin{aligned} T &= \sum_a \tfrac{1}{2} m_a \dot{\mathbf{r}}_a \cdot \dot{\mathbf{r}}_a = \sum_a \tfrac{1}{2} m_a \left(\boldsymbol{\omega} \times \mathbf{r}_a\right) \cdot \left(\boldsymbol{\omega} \times \mathbf{r}_a\right) = \sum_a \tfrac{1}{2} m_a \boldsymbol{\omega} \cdot \left[\mathbf{r}_a \times \left(\boldsymbol{\omega} \times \mathbf{r}_a\right)\right] \\ &= \frac{1}{2} \boldsymbol{\omega} \cdot \mathbf{L} = \frac{1}{2} \boldsymbol{\omega} \cdot \overline{\overline{\mathbf{I}}} \cdot \boldsymbol{\omega} = \frac{1}{2} \sum_i I_i \omega_i^2. \end{aligned} \tag{2.4.4}$$

The principal axes can always be found by diagonalising the inertia tensor $\overline{\overline{\mathbf{I}}}$, but calculations become easier if one already knows their directions because then one can choose them as the coordinate axes. It is therefore useful to know that any symmetry axis is always a principal axis, and than the direction normal to any symmetry plane is also a principal axis.

If two of the principal moments of inertia are equal, say $I_1 = I_2$, we say that the body is a *symmetric body*. In this case, any linear combination of the $\hat{\mathbf{e}}_1$ and $\hat{\mathbf{e}}_2$, so any two orthogonal directions on the plane spanned by $\hat{\mathbf{e}}_1$ and $\hat{\mathbf{e}}_2$ can be chosen as the principal axis. Note that although a system with an axis of cylindrical symmetry, e.g. a cylinder or a cone, would certainly be a symmetric body in this sense, it is not necessary. In fact any system with a more than 2-fold rotational symmetry would suffice, e.g. a triangular prism, or the two principal moments could be equal just by chance in spite of the body have no geometrical symmetry. In the case of a symmetric body, Eq. (2.4.3) becomes

$$\mathbf{L} = I_1 \left(\omega_1 \hat{\mathbf{e}}_1 + \omega_2 \hat{\mathbf{e}}_2\right) + I_3 \omega_3 \hat{\mathbf{e}}_3 \,. \tag{2.4.5}$$

If all 3 moments of inertia are equal, we say the body is *totally symmetric*. Again, this can happen either by symmetry, as in a sphere, cube, regular tetrahedron or any of the five regular solids, or by coincidence. In the case of a totally symmetric body, we have $\mathbf{L} = I\boldsymbol{\omega}$ and $\mathbf{L}$ is always in the same direction as $\boldsymbol{\omega}$. In that case the choice of the 3 principal axes is completely arbitrary, as long as they are mutually perpendicular.

2.5 Calculation of Moments of Inertia

2.5.1 Shift of Origin

It is often useful to be able to relate the moments of inertia about different pivots, e.g. when a body is pivoted around a point other than its centre of mass. We write the position vector $\mathbf{r}_a$ has the sum of the centre-of-mass position $\mathbf{R}$ and the position relative to the centre-of-mass $\mathbf{r}_a^*$, i.e., $\mathbf{r}_a = \mathbf{R} + \mathbf{r}_a^*$. Then, by definition,

$$\sum_a m_a \mathbf{r}_a^* = 0. \tag{2.5.1}$$

Therefore we can write the components of the inertia tensor (2.3.14) as

$$\begin{aligned} I_{ij} &= \sum_a m_a \left[\left(\mathbf{R}^2 + \mathbf{r}_a^* \right)^2 \delta_{ij} - (R_i + r_{ai}^*)(R_j + r_{aj}^*) \right] \\ &= \sum_a m_a \left[\mathbf{R}^2 \delta_{ij} + (\mathbf{r}_a^*)^2 \delta_{ij} - R_i R_j - r_{ai}^* r_{aj}^* \right] = M \left(\mathbf{R}^2 \delta_{ij} - R_i R_j \right) + I_{ij}^*, \end{aligned} \tag{2.5.2}$$

where

$$I_{ij}^* = \sum_a m_a \left[(\mathbf{r}_a^*)^2 \delta_{ij} - r_{ai}^* r_{aj}^* \right]. \tag{2.5.3}$$

If we know the inertia tensor with respect to the centre of mass $\overline{\overline{\mathbf{I}}}^*$, we can use these relations to easily calculate with respect to any origin we want. This is known as the *parallel axes theorem.* Note that the principal axes about a general point are not necessarily parallel to those about the centre of mass, unless the point itself lies on one of the principal axes.

2.5.2 Continuous Solid

Generally we have a continuous solid rather than a group of point particles. In this case the sums become integrals and the masses, m_a become densities, $\rho(\mathbf{r})$, so that we have

$$I_{ij} = \int \rho(\mathbf{r}) \left(r^2 \delta_{ij} - r_i r_j \right) \, \mathrm{d}^3 r. \tag{2.5.4}$$

2.5.3 Routh's Rule

Let us now imagine that the coordinate axes have been chosen to agree with the principal axes. We can then see from Eq. (2.5.4) that we can split the principal moments of inertia such that

$$I_1^* = K_y + K_z, \qquad I_2^* = K_x + K_z, \qquad I_3^* = K_x + K_y, \tag{2.5.5}$$

where

$$K_i = \int_V \rho r_i^2 \, \mathrm{d}^3 r. \tag{2.5.6}$$

It is now useful to ask how the principal moments change if we rescale (i.e. stretch or squeeze) the body in the directions the principal axes. To do this in practice, let us first consider the original body $\tilde{V}$ assuming that we know the constants $\tilde{K}_i$ defined by by Eq. (2.5.6). The rescaled body V is obtained by rescaling the coordinates as $r_i = a_i \tilde{r}_i$ for each $i \in \{1, 2, 3\}$. The constants K_i in Eq.(2.5.6) change to

$$K_i = \int_V \rho {r_i}^2 \, \mathrm{d}x \, \mathrm{d}y \, \mathrm{d}z = a_1 a_2 a_3 a_i^2 \int_{\tilde{V}} \rho \tilde{r}_i^2 \, \mathrm{d}\tilde{x} \, \mathrm{d}\tilde{y} \, \mathrm{d}\tilde{z} = a_1 a_2 a_3 a_i^2 \tilde{K}_i. \tag{2.5.7}$$

We can also note that the total mass of the body V is

$$M = \int_V \rho \,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z = a_1 a_2 a_3 \int_{\tilde{V}} \rho \mathrm{d}\tilde{x}\,\mathrm{d}\tilde{y}\,\mathrm{d}\tilde{z} = a_1 a_2 a_3 \tilde{M}, \tag{2.5.8}$$

where $\tilde{M}$ is the mass of the original body. Hence $K_i \propto a_i^2 M$, and we can write

$$K_i = \lambda_i a_i^2 M, \tag{2.5.9}$$

where λ_z is a dimensionless number, which is the same for all bodies of the same general type. Hence we have *Routh's rule* which states that

$$\begin{aligned} I_1^* &= M\left(\lambda_y a_y^2 + \lambda_z a_z^2\right), \\ I_2^* &= M\left(\lambda_x a_x^2 + \lambda_z a_z^2\right), \\ I_3^* &= M\left(\lambda_x a_x^2 + \lambda_y a_y^2\right). \end{aligned} \tag{2.5.10}$$

By checking the standard bodies we obtain the following values for the coefficients: $\lambda = \frac{1}{3}$ for 'rectangular' axes, $\lambda = \frac{1}{4}$ for 'elliptical' axes and $\lambda = \frac{1}{5}$ for 'ellipsoidal' ones. This covers most special cases. For example, a sphere is an ellipsoid with $a_x = a_y = a_z = a$ and each principal moment of inertia is $\frac{2}{5}Ma^2$, whereas a cube is a parallelepiped with $a_x = a_y = a_z = a$ and $I = \frac{2}{3}Ma^2$.

For a cylinder we have λ_x and λ_y elliptical and λ_z rectangular. This nomenclature can be confusing as it refers to the symmetry of the corresponding integrals and not to symmetry about the axes. For a cylinder with $a_x = a_y \neq a_z$ we have

$$I_1^* = I_2^* = M\left(\tfrac{1}{4}a_x^2 + \tfrac{1}{3}a_z^2\right) \qquad I_3^* = M\left(\tfrac{1}{4}a_x^2 + \tfrac{1}{4}a_x^2\right) = \tfrac{1}{2}Ma_x^2\,, \tag{2.5.11}$$

and therefore a flat circular plate, i.e. a cylinder with $a_z = 0$, has

$$I_1^* = I_2^* = \tfrac{1}{4}Ma_x^2 \qquad I_3^* = \tfrac{1}{2}Ma_x^2\,. \tag{2.5.12}$$

Conversely, a thin rod is a cylinder with $a_x = a_y = 0$ and

$$I_1^* = I_2^* = \tfrac{1}{3}Ma_z^2 \qquad I_3^* = 0\,. \tag{2.5.13}$$

2.6 Effect of Small Force

Suppose a body is rotating about a principal axis such that $\boldsymbol{\omega} = \omega\hat{\mathbf{e}}_3$ and $\mathbf{L} = I_3\omega\hat{\mathbf{e}}_3$. Then

$$\dot{\mathbf{L}} = I_3\dot{\boldsymbol{\omega}} = 0\,, \tag{2.6.1}$$

the axis will remain fixed in space and the angular velocity will be constant. Note that this would not be true if $\boldsymbol{\omega}$ were not a principal axis.

Suppose now that the axis is fixed at the origin and a small force $\mathbf{F}$ is applied to the axis at point $\mathbf{r}$. Then the equation of motion becomes

$$\dot{\mathbf{L}} = \mathbf{r} \times \mathbf{F}\,. \tag{2.6.2}$$

The body will acquire a small component of angular velocity perpendicular to its axis. However, if the force is small, this will be small compared with the angular velocity of rotation about the axis. We may then neglect the angular momentum components normal to the axis and again write

$$\dot{\mathbf{L}} = I_3\dot{\boldsymbol{\omega}} = \mathbf{r} \times \mathbf{F}\,. \tag{2.6.3}$$

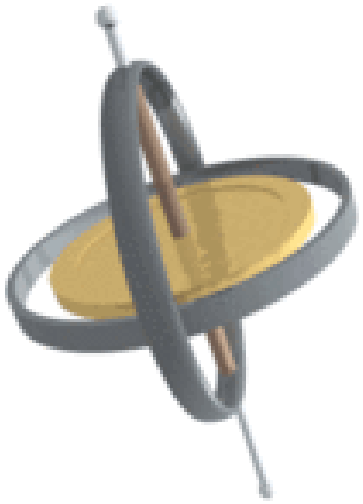

Figure 2.2:

Since $\mathbf{r} \times \mathbf{F}$ is perpendicular to $\boldsymbol{\omega}$ ($\mathbf{r}$ is parallel to $\boldsymbol{\omega}$) the magnitude of ω does not change ($\mathrm{d}\omega^2/\mathrm{d}t = 2\boldsymbol{\omega} \cdot \dot{\boldsymbol{\omega}} = 0$). Its direction does change, however, in the direction of $\mathbf{r} \times \mathbf{F}$ and hence perpendicular to the applied force $\mathbf{F}$.

As an example, consider a child's spinning top. In general, the rotation axis is not exactly vertical. We consider the point as which the top touches the ground as the pivot point, and use it as our origin. There is a gravitational force $\mathbf{F} = -Mg\hat{\mathbf{k}}$, acting at the centre of mass at position $\mathbf{R} = R\hat{\mathbf{e}}_3$. Eq. (2.6.3) gives

$$\begin{aligned} I_3\omega \frac{\mathrm{d}\hat{\mathbf{e}}_3}{\mathrm{d}t} &= -MgR\,\hat{\mathbf{e}}_3 \times \hat{\mathbf{k}} \\ \Rightarrow \frac{\mathrm{d}\hat{\mathbf{e}}_3}{\mathrm{d}t} &= \left(\frac{MgR}{I_3\omega}\right) \hat{\mathbf{k}} \times \hat{\mathbf{e}}_3. \end{aligned} \tag{2.6.4}$$

This has the same form as Eq. (1.2.1), i.e.,

$$\frac{\mathrm{d}\hat{\mathbf{e}}_3}{\mathrm{d}t} = \boldsymbol{\Omega} \times \hat{\mathbf{e}}_3, \tag{2.6.5}$$

Which means that the principal axis $\hat{\mathbf{e}}_3$ rotates around the vertical direction $\hat{\mathbf{k}}$ with angular velocity

$$\boldsymbol{\Omega} = \frac{MgR}{I_3\omega}\hat{\mathbf{k}}\,. \tag{2.6.6}$$

The analysis is only valid when $\Omega \ll \omega$ or when $MgR \ll I_3\omega^2$; the potential energy associated with the tilt is much smaller than the kinetic energy of the rotation. The system is very similar to Larmor precession (see section 1.6.4). The expression for Ω tells us a great deal about this system. Note that Ω is inversely proportional to both the moment of inertia I_3 and the angular frequency ω. This implies that to minimise the precession and hence to improve the stability of the system we have to choose both to be large: we require a fat rapidly spinning body.

This is the basis of the gyroscope: the high stability of such a rapidly rotating body makes it ideal for use in navigation, especially near the poles where a compass is almost useless. It can also be used, e.g. , to provide an "artificial horizon" when flying blind, either in cloud or at night.

2.7 Rotation about a Principal Axis

As the principal axes are fixed in the body we are really dealing with a rotating frame. We here return to the notation used in Section 1 to distinguish between the inertial and rotating frames. The rate of change of the angular momentum in the inertial frame is

$$\left.\frac{\mathrm{d}\mathbf{L}}{\mathrm{d}t}\right|_{\mathrm{I}} = \sum_a \mathbf{r}_a \times \mathbf{F}_a = \boldsymbol{\tau}\,. \tag{2.7.1}$$

Eq. (Rot:eq:7) relates this to the rate of change measured in the rotating frame,

$$\left.\frac{\mathrm{d}\mathbf{L}}{\mathrm{d}t}\right|_{\mathrm{I}} = \left.\frac{\mathrm{d}\mathbf{L}}{\mathrm{d}t}\right|_{\mathrm{R}} + \boldsymbol{\omega} \times \mathbf{L}\,. \tag{2.7.2}$$

On the other hand, because in the rotating frame the principal axes and principal moments are fixed, we have

$$\left.\frac{\mathrm{d}\mathbf{L}}{\mathrm{d}t}\right|_{\mathrm{R}} = I_1\dot{\omega}_1\hat{\mathbf{e}}_1 + I_2\dot{\omega}_2\hat{\mathbf{e}}_2 + I_3\dot{\omega}_3\hat{\mathbf{e}}_3\,, \tag{2.7.3}$$

and, therefore,

$$\left.\frac{\mathrm{d}\mathbf{L}}{\mathrm{d}t}\right|_{\mathrm{R}} + \boldsymbol{\omega} \times \mathbf{L} = \boldsymbol{\tau}. \tag{2.7.4}$$

Calculating the cross product

$$\begin{aligned}\boldsymbol{\omega} \times \mathbf{L} &= \begin{vmatrix} \hat{\mathbf{e}}_1 & \hat{\mathbf{e}}_2 & \hat{\mathbf{e}}_3 \\ \omega_1 & \omega_2 & \omega_3 \\ I_1\omega_1 & I_2\omega_2 & I_3\omega_3 \end{vmatrix} \\ &= (I_3 - I_2)\omega_2\omega_3\hat{\mathbf{e}}_1 + (I_1 - I_3)\omega_1\omega_3\hat{\mathbf{e}}_2 + (I_2 - I_1)\omega_1\omega_2\hat{\mathbf{e}}_3,\end{aligned} \tag{2.7.5}$$

we find the Euler equations

$$\begin{aligned} I_1\dot{\omega}_1 + (I_3 - I_2)\,\omega_2\omega_3 &= \tau_1\,, \\ I_2\dot{\omega}_2 + (I_1 - I_3)\,\omega_3\omega_1 &= \tau_2\,, \\ I_3\dot{\omega}_3 + (I_2 - I_1)\,\omega_1\omega_2 &= \tau_3\,. \end{aligned} \tag{2.7.6}$$

In principle these equations could be solved to give $\boldsymbol{\omega}(t)$. In practice, however, we often don't have the force expressed in a useful form to do this and, in any case, it is easier to solve this system using Lagrangian methods (see chapter 3).

For the moment we concentrate on studying the stability of the motion in the absence of external forces ($\boldsymbol{\tau} = 0$). Suppose that the object is rotating about the principal axis $\hat{\mathbf{e}}_3$ and that $\omega_1 = \omega_2 = 0$ then it is obvious from Eq. (2.7.6) that the object will continue indefinitely to rotate about $\hat{\mathbf{e}}_3$. On the other hand let us suppose that the motion deviates slightly from this such that ω_1 and ω_2 are much smaller than ω_3. We may therefore ignore any terms which are quadratic in ω_1 and ω_2 so that, from the third line in Eq. (2.7.6), we have $\dot{\omega}_3 = 0$ and ω_3 is constant.

We look for solutions of the form[2]

$$\omega_1 = a_1\mathrm{e}^{\gamma t} \qquad \omega_2 = a_2\mathrm{e}^{\gamma t} \tag{2.7.7}$$

[2]Those doing computational physics will note the similarity between this analysis and the stability analysis considered there.

where a_1, a_2 and γ are constants. Substituting this into Eq. (2.7.6) gives

$$I_1\gamma a_1 + (I_3 - I_2)\,\omega_3 a_2 = 0 \tag{2.7.8}$$
$$I_2\gamma a_2 + (I_1 - I_3)\,\omega_3 a_1 = 0\,, \tag{2.7.9}$$

which is a 2×2 eigenvalue problem with a solution

$$\gamma^2 = \frac{(I_3 - I_2)\,(I_1 - I_3)}{I_1 I_2}\omega_3^2\,. \tag{2.7.10}$$

We note that $\omega_3^2/I_1 I_2$ is always positive. Hence, if I_3 is the smallest or the largest of the 3 moments of inertia γ^2 is negative. In that case γ is imaginary and the motion is oscillatory. Hence its amplitude does not change, and we say that the rotation is stable.

However, if I_3 is the middle of the three moments then γ^2 is positive and γ is real. There are two independent solutions with opposite signs of γ, and in general the solution is a linear combination of them. However, at late times ($t \gg 1/\gamma$) the solution with a positive exponent dominates. Hence ω_1 and ω_2 tend to grow exponentially and the motion about $\hat{\mathbf{e}}_3$ is unstable: any small deviation from rotation about $\hat{\mathbf{e}}_3$ will tend to grow.

You can test this by trying to spin an appropriately dimensioned object, such as a book or a tennis racket. It is much easier to spin it around the axis with the smallest or the largest moment of inertia, but not the middle one.

2.8 Euler's Angles

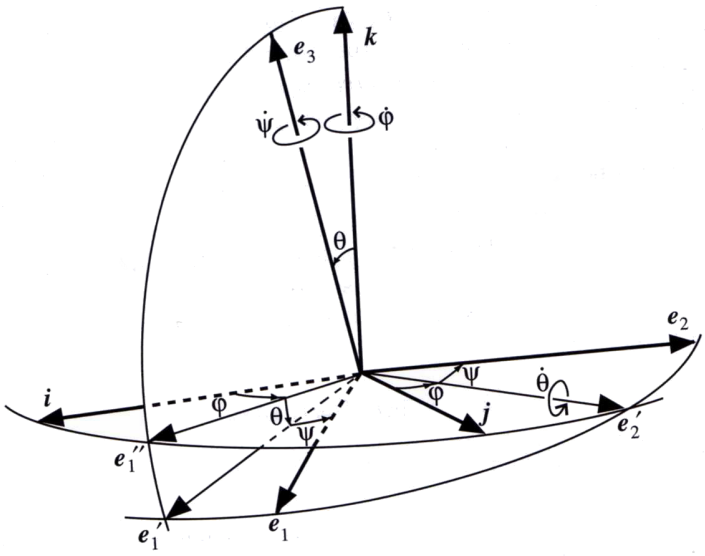

Figure 2.3:

In order to describe the orientation of a solid body we require 3 angles. The conventional way to do this is to define angles (ϕ, θ, ψ) these is known as *Euler's Angles,*, which are illustrated in Fig. 2.3. Note however that there are several different conventions for Euler's Angles. We shall stick to the one used by Kibble & Berkshire, known as the y–convention. The meaning of the angles is, essentially,

that ϕ and θ are the usual spherical coordinates expressing the direction of the principal axis $\hat{\mathbf{e}}_3$, and ψ expresses the orientation of the object about this axis.

Let us construct the angles in detail. We can obviously express the orientation of the body by giving the orientations of the three principal axes, i.e., by a triplet of orthogonal unit vectors $(\hat{\mathbf{e}}_1, \hat{\mathbf{e}}_2, \hat{\mathbf{e}}_3)$. To show that we can parameterise these with the three Euler angles, let us start with the orientation $(\hat{\boldsymbol{\imath}}, \hat{\boldsymbol{\jmath}}, \hat{\mathbf{k}})$, which means that the principal axes are aligned with the axes of our original Cartesian coordinate system. As illustrated in Fig. 2.3, we then carry out three steps:

- We first rotate by ϕ about the $\hat{\mathbf{k}}$ axis. The changes the directions of the first two principal axes, and we denote the new directions by $\hat{\mathbf{e}}_1''$ and $\hat{\mathbf{e}}_2'$. Thus, the orientation of the principal axes changes as $(\hat{\boldsymbol{\imath}}, \hat{\boldsymbol{\jmath}}, \hat{\mathbf{k}}) \to (\hat{\mathbf{e}}_1'', \hat{\mathbf{e}}_2', \hat{\mathbf{k}})$.
- Secondly we rotate by θ about the second principal axis $\hat{\mathbf{e}}_2'$. This changes the directions of the first and third principal axes to $\hat{\mathbf{e}}_1'$ and $\hat{\mathbf{e}}_3$, so the orientation of the body changes as $(\hat{\mathbf{e}}_1'', \hat{\mathbf{e}}_2', \hat{\mathbf{k}}) \to (\hat{\mathbf{e}}_1', \hat{\mathbf{e}}_2', \hat{\mathbf{e}}_3)$.
- Finally we rotate by ψ about the third principal axis $\hat{\mathbf{e}}_3$, to bring the first principal axis to direction $\hat{\mathbf{e}}_1$ and the second principal axis to $\hat{\mathbf{e}}_2$, i.e., $(\hat{\mathbf{e}}_1', \hat{\mathbf{e}}_2', \hat{\mathbf{e}}_3) \to (\hat{\mathbf{e}}_1, \hat{\mathbf{e}}_2, \hat{\mathbf{e}}_3)$.

Using these three rotations we can reach any orientation $(\hat{\mathbf{e}}_1, \hat{\mathbf{e}}_2, \hat{\mathbf{e}}_3)$ we want, and therefore the orientation of the body is fully parameterised by the three Euler angles.

Because the three angles (ϕ, θ, ψ) correspond to rotations about the axes $\hat{\mathbf{k}}$, $\hat{\mathbf{e}}_2'$ and $\hat{\mathbf{e}}_3$, respectively. Note that these axes are not mutually perpendicular. We can, nevertheless, use them to express the angular velocity $\boldsymbol{\omega}$ in terms of Euler angles as

$$\boldsymbol{\omega} = \dot{\phi}\hat{\mathbf{k}} + \dot{\theta}\hat{\mathbf{e}}_2' + \dot{\psi}\hat{\mathbf{e}}_3 \,. \tag{2.8.1}$$

For a symmetric system such as a gyroscope we can choose $\hat{\mathbf{e}}_3$ as the symmetry axis and, as $I_1 = I_2$, any two mutually perpendicular axes as the other two. In this case the most convenient are $\hat{\mathbf{e}}_1'$ and $\hat{\mathbf{e}}_2'$ as two of the axes are already used in Eq. (2.8.1). We can therefore use that $\hat{\mathbf{k}} = -\sin\theta\,\hat{\mathbf{e}}_1' + \cos\theta\,\hat{\mathbf{e}}_3$ to obtain

$$\boldsymbol{\omega} = -\dot{\phi}\sin\theta\,\hat{\mathbf{e}}_1' + \dot{\theta}\hat{\mathbf{e}}_2' + \left(\dot{\psi} + \dot{\phi}\cos\theta\right)\hat{\mathbf{e}}_3 \,, \tag{2.8.2}$$

where the unit vectors are mutually perpendicular and, for a symmetric body, principal axes.

Using Eq. (2.8.2), we can express the angular momentum and kinetic energy as

$$\mathbf{L} = -I_1\dot{\phi}\sin\theta\,\hat{\mathbf{e}}_1' + I_1\dot{\theta}\hat{\mathbf{e}}_2' + I_3\left(\dot{\psi} + \dot{\phi}\cos\theta\right)\hat{\mathbf{e}}_3 \tag{2.8.3}$$

$$T = \tfrac{1}{2}I_1\dot{\phi}^2\sin^2\theta + \tfrac{1}{2}I_1\dot{\theta}^2 + \tfrac{1}{2}I_3\left(\dot{\psi} + \dot{\phi}\cos\theta\right)^2 . \tag{2.8.4}$$

To find equations of motion we could either translate this into Cartesian coordinates, $\hat{\boldsymbol{\imath}}, \hat{\boldsymbol{\jmath}}, \hat{\mathbf{k}}$, or try to write the equations in terms of the Euler angles. Either way is difficult. It is much easier to use Lagrangian methods (see Chapter 3).

In the meantime we can consider the free motion, with no forces. In this case $\mathbf{L}$ is a constant. We therefore choose the vector $\hat{\mathbf{k}}$ to be in the direction of $\mathbf{L}$ such that

$$\mathbf{L} = L\hat{\mathbf{k}} = -L\sin\theta\,\hat{\mathbf{e}}_1' + L\cos\theta\hat{\mathbf{e}}_3 \,. \tag{2.8.5}$$

This must be equal to Eq. (2.8.3) so that by equating components we can write

$$I_1\dot{\phi}\sin\theta = L\sin\theta \tag{2.8.6}$$
$$I_1\dot{\theta} = 0 \tag{2.8.7}$$
$$I_3\left(\dot{\psi}+\dot{\phi}\cos\theta\right) = L\cos\theta \tag{2.8.8}$$

From Eq. (2.8.7) we deduce that θ is constant. As long as $\sin\theta \neq 0$, Eq. (2.8.6) implies that that $\dot{\phi}$ is constant, too,

$$\dot{\phi} = \frac{L}{I_1}, \tag{2.8.9}$$

and hence, from Eq. (2.8.8), we find that $\dot{\psi}$ is also a constant,

$$\dot{\psi} = L\cos\theta\left(\frac{1}{I_3}-\frac{1}{I_1}\right). \tag{2.8.10}$$

We conclude therefore that the axis $\hat{\mathbf{e}}_3$ rotates around $\mathbf{L}$ at a constant rate $\dot{\phi}$ and at an angle θ to it. In addition the body spins about the axis $\hat{\mathbf{e}}_3$ at a constant rate $\dot{\psi}$. The angular velocity vector $\boldsymbol{\omega}$ deduced from Eq. (2.8.2) is

$$\boldsymbol{\omega} = -\dot{\phi}\sin\theta\,\hat{\mathbf{e}}_1' + \left(\dot{\psi}+\dot{\phi}\cos\theta\right)\hat{\mathbf{e}}_3 \tag{2.8.11}$$

which describes a cone around the direction of $\mathbf{L}$.

Note that this appears very similar to precession (see Section 2.6): ψ is the angle of rotation of the gyroscope around its axis, θ is the angle between the gyroscope axis and the angular momentum and ϕ is the angle which describes the precession around this direction. However, here we are describing free rotation with no external forces involved.

Chapter 3

Lagrangian Mechanics

3.1 Action Principle

In this chapter we will see how the familiar laws of mechanics can be expressed and understood from a very different point of view, which is known as the Lagrangian formulation of mechanics. This is in many ways more elegant than the Newtonian formulation, and it is particularly useful when moving to quantum mechanics. For example, quantum field theories are usually studied in a Lagrangian framework.

The idea is similar to Fermat's principle in optics, according to which light follows the shortest optical path, i.e., the path of shortest time to reach its destination. As a reminder, let us see how Snell's law

$$\frac{\sin\theta_2}{\sin\theta_1} = \frac{n_1}{n_2}, \tag{3.1.1}$$

which tells how a light ray bends at the interface of two materials with refractive indices n_1 and n_2.

Consider a light ray from point (x_a, y_a) to (x_b, y_b). There is a horizontal interface at y, and between y_a and y, the refractive index is n_1 and between y and y_b it is n_2. We now assume that the light follows a straight line from (x_a, y_a) to a point (x, y) on the interface, and then from (x, y) to (x_b, y_b). The only unknown is therefore x. Because of the speed of light in medium is c/n, the optical path length is

$$S(x) = \int_a^b dt = \int_a^b \frac{n}{c} dl = \frac{n_1}{c}\sqrt{(x-x_a)^2 + (y-y_a)^2} + \frac{n_2}{c}\sqrt{(x_b-x)^2 + (y_b-y)^2}. \tag{3.1.2}$$

According to Fermat's principle, we need to find the minimum of this quantity. At the minimum, the derivative with respect to x vanishes, so

$$0 = \frac{\partial S}{\partial x} = \frac{n_1}{c}\frac{x-x_a}{\sqrt{(x-x_a)^2 + (y-y_a)^2}} - \frac{n_2}{c}\frac{x_b-x}{\sqrt{(x_b-x)^2 + (y_b-y)^2}} = \frac{n_1}{c}\sin\theta_1 - \frac{n_2}{c}\sin\theta_2, \tag{3.1.3}$$

from which Snell's law (3.1.1) follows immediately.

Fermat had proposed his principle in 1657, and it motivated Maupertuis to suggest in 1746 that also matter particles would obey an analogous variational principle. He postulated that there exists a quantity called action, which the trajectory of the matter particle would minimise. This idea was later refined by Lagrange and Hamilton, who developed it into its current form, in which the action S is

defined as an integral over a function $L(x, \dot{x}, t)$ known as the Lagrangian, as

$$S[x] = \int_{t_a}^{t_b} L(x(t), \dot{x}(t), t)dt. \tag{3.1.4}$$

The Lagrangian is a function of the position x and velocity $\dot{x}$ of the particle, and it may also have some explicit time dependence. We will see later that for conservative systems, the Lagrangian is simply the difference of the kinetic energy T and the potential energy V, i.e., $L = T - V$.

Because the action S is given by an integral over time, it depends on the position and velocity at all times, i.e., on the whole trajectory of the particle. It is therefore a function from the space of functions $x(t)$ to real numbers, and we indicate that by having the argument (i.e. function x) in square brackets. Such function of functions are called functionals.

Given a Lagrangian L, the dynamics is determined by Hamilton's principle (or action principle), which states that *to move from position x_a at time t_a to position x_b at time t_b, the particle follows the trajectory that minimises the action $S[x]$*. In other words, the actual physical trajectory is the function x that minimises the action subject to the boundary conditions $x(t_a) = x_a$ and $x(t_b) = x_b$.

To find this minimising function $x(t)$, we want to calculate the derivative of the action $S[x]$ with respect to the function $x(t)$ and set it to zero. Functional derivatives such as this are studied in the branch of mathematics known as functional analysis. However, he we adopt a slightly simpler approach and consider small variations of the trajectory. This is known as *variational calculus*.

Let us assume that $x(t)$ is the function that minimises the action, and consider a slightly perturbed trajectory

$$\tilde{x}(t) = x(t) + \delta x(t), \tag{3.1.5}$$

where we assume that the perturbation is infinitesimally small and vanishes at the endpoints,

$$\delta x(t_a) = \delta x(t_b) = 0. \tag{3.1.6}$$

This perturbation changes the action by

$$\begin{aligned}
\delta S &= S[x + \delta x] - S[x] \\
&= \int_{t_a}^{t_b} \left[L(x(t) + \delta x(t), \dot{x}(t) + \delta\dot{x}(t), t) - L(x(t), \dot{x}(t), t)\right] dt \\
&= \int_{t_a}^{t_b} \left[\frac{\partial L}{\partial x}\delta x(t) + \frac{\partial L}{\partial \dot{x}}\delta\dot{x}(t)\right] dt \\
&= \int_{t_a}^{t_b} \left[\frac{\partial L}{\partial x}\delta x(t) + \frac{\partial L}{\partial \dot{x}}\frac{d\delta x(t)}{dt}\right] dt \\
&= \int_{t_a}^{t_b} \frac{\partial L}{\partial x}\delta x(t)dt + \left[\frac{\partial L}{\partial \dot{x}}\delta x(t)\right]_{t_a}^{t_b} - \int_{t_a}^{t_b} \left[\frac{d}{dt}\frac{\partial L}{\partial \dot{x}}\right] \delta x(t)dt,
\end{aligned} \tag{3.1.7}$$

where we Taylor expanded to linear order in $\delta x(t)$ and integrated the second term by parts. Because of the boundary conditions (3.1.6), the substitution term vanishes, and we have

$$\delta S = \int_{t_a}^{t_b} \left[\frac{\partial L}{\partial x} - \frac{d}{dt}\frac{\partial L}{\partial \dot{x}}\right] \delta x(t)dt. \tag{3.1.8}$$

For $x(t)$ to be minimum, the variation of the action (3.1.8) has to vanish for any function $\delta x(t)$. You can see this by noting that if $\delta S < 0$ for any perturbation $\delta x(t)$, then $S[x + \delta x] < S[x]$.

Correspondingly, if $\delta S > 0$, then $S[x - \delta x] < S[x]$. In either case, we have found a function that has lower action than $x(t)$. Therefore $x(t)$ can only be the minimum if $\delta S = 0$.

The only way we can have $\delta S = 0$ for every perturbation $\delta x(t)$ is that the expression inside the brackets in Eq. (3.1.8) vanishes, i.e.,

$$\frac{d}{dt}\frac{\partial L}{\partial \dot{x}} - \frac{\partial L}{\partial x} = 0. \tag{3.1.9}$$

This is known as the *Euler-Lagrange equation*, and it is the equation of motion in the Lagrangian formulation of mechanics. When using Eq. (3.1.9), it is very important to understand the difference between the partial (∂) and total (d) derivatives.

To check that Eq. (3.1.9) really describes the same physics as Newtonian mechanics, let us consider a simple example of a particle in a one-dimensional potential $V(x)$. Because the system is conservative, the Lagrangian is

$$L = T - V = \frac{1}{2}m\dot{x}^2 - V(x), \tag{3.1.10}$$

and the Euler-Lagrange equation is

$$\frac{d}{dt}\frac{\partial L}{\partial \dot{x}} - \frac{\partial L}{\partial x} = \frac{d}{dt}(m\dot{x}) + \frac{dV}{dx} = m\ddot{x} + \frac{dV}{dx} = 0. \tag{3.1.11}$$

This is nothing but Newton's second law

$$m\ddot{x} = -\frac{dV}{dx}. \tag{3.1.12}$$

It is interesting to note that although Newtonian and Lagrangian formulations of mechanics are mathematically equivalent and describe the same physics, their starting point is very different. Newton's laws describe the evolution of the system as an initial value problem: We know the position and velocity of the particle at the initial time, $x(t_a)$ and $\dot{x}(t_a)$, and we then use Newton's laws to determine the evolution $x(t)$ at later times $t > t_a$.

In contrast, the Lagrangian formulation describes the same physics as a boundary value problem. We know the initial and final positions of the particle $x(t_a)$ and $x(t_b)$, and we use the action principle to determine $x(t)$ for $t_a < t < t_b$, i.e., how the particle travels from one to the other. In particular, we cannot choose the initial velocity because it is determined by the final destination of the particle through the action principle. This may appear very non-local in time because the behaviour of the particle in the far future determines its motion at the current time. However, because because the two formulations are equivalent, this apparent non-locality in time does not actually affect the physics. For example, it is not possible to use it to send information back in time. In practice, it is usually easier to solve initial value problems, and therefore one usually uses the Lagrangian formulation to set up the problem and derive the equations of motion but then solves them as an initial value problem.

In many ways the Lagrangian formulation is closer to quantum mechanics, which does not allow one to determine the initial position and velocity of the particle either. Furthermore, the principle that the particle chooses one out of all possible trajectories resembles the double slit experiment in quantum mechanics, with the key difference that in the quantum case one has to sum over all possible trajectories rather than just selecting one. This correspondence turns out to be fully accurate and becomes obvious in the path integral formulation in quantum mechanics.

3.2 Generalised Coordinates

One attractive aspect of the Lagrangian formulation is that it is independent of the variables that are used to describe the state of the system. This is because the minimum of the function does not depend on the coordinate system, and the same applies to a functional such as the action S. Therefore, in contrast with Newtonian mechanics, we do not have to use the Cartesian position coordinates, and the Euler-Lagrange equation still has the same form (3.1.9). Instead, we are free to choose whichever set of variables we want to parameterise the state of the system, and which are then called *generalised coordinates* and usually denoted by q. They can be position coordinates, but also angles etc.

Usually we need more than one generalised coordinate, which we label by index i, so that we have some number N generalised coordinates q_i, with $i = 1, \ldots, N$. Each coordinate satisfies the corresponding Euler-Lagrange equation

$$\frac{d}{dt}\frac{\partial L}{\partial \dot{q}_i} - \frac{\partial L}{\partial q_i} = 0. \tag{3.2.1}$$

As the first example of generalised coordinates, let us consider a simple pendulum that has a mass m at the end of a light rod of fixed length l. The angle of the pendulum from the vertical position is θ, which we choose as the generalised coordinate $q = \theta$. The Lagrangian is

$$L = T - V = \frac{1}{2}ml^2\dot{\theta}^2 + mgl\cos\theta. \tag{3.2.2}$$

The Euler-Lagrange equation is

$$\frac{d}{dt}\frac{\partial L}{\partial \dot{\theta}} - \frac{\partial L}{\partial \theta} = \frac{d}{dt}\left(ml^2\dot{\theta}\right) + mgl\sin\theta = ml^2\ddot{\theta} + mgl\sin\theta = 0, \tag{3.2.3}$$

from which we find

$$\ddot{\theta} = -\frac{g}{l}\sin\theta. \tag{3.2.4}$$

As a slightly more complex example, let us consider the motion of a particle in a central potential $V(r)$ in three dimensions. We use the spherical coordinates (r, θ, ϕ) as the generalised coordinates. The Lagrangian is

$$L = T - V = \frac{1}{2}m\left(\dot{r}^2 + r^2\dot{\theta}^2 + r^2\dot{\phi}^2\sin^2\theta\right) - V(r). \tag{3.2.5}$$

The Euler-Lagrange equation for r is

$$\frac{d}{dt}\frac{\partial L}{\partial \dot{r}} - \frac{\partial L}{\partial r} = \frac{d}{dt}\left(m\dot{r}\right) - mr\left(\dot{\theta}^2 + \dot{\phi}^2\sin^2\theta\right) + \frac{dV}{dr} = 0, \tag{3.2.6}$$

which gives

$$m\ddot{r} = mr\left(\dot{\theta}^2 + \dot{\phi}^2\sin^2\theta\right) - \frac{dV}{dr}. \tag{3.2.7}$$

The second term on the right hand side is the force due to the potential, and the first term is the centrifugal force.

The Euler-Lagrange equation for θ is

$$\frac{d}{dt}\frac{\partial L}{\partial \dot{\theta}} - \frac{\partial L}{\partial \theta} = \frac{d}{dt}\left(mr^2\dot{\theta}\right) - mr^2\dot{\phi}^2\sin\theta\cos\theta = 0. \tag{3.2.8}$$

Finally, because ϕ does appear in the Lagrangian (except as a time derivative $\dot{\phi}$), the Euler-Lagrange equation for ϕ is

$$\frac{d}{dt}\frac{\partial L}{\partial \dot{\phi}} - \frac{\partial L}{\partial \phi} = \frac{d}{dt}\frac{\partial L}{\partial \dot{\phi}} = 0. \tag{3.2.9}$$

This means that the quantity

$$\frac{\partial L}{\partial \dot{\phi}} = mr^2 \sin^2\theta \dot{\phi} \tag{3.2.10}$$

is conserved. We note that this is simply L_z, the z component of the angular momentum vector $\mathbf{L} = m\mathbf{r} \times \dot{\mathbf{r}}$.

We can easily see that this is, in fact, a very general result. For any generalised coordinate q_i, we define the generalised momentum p_i by

$$p_i = \frac{\partial L}{\partial \dot{q}_i}. \tag{3.2.11}$$

The Euler-Lagrange equation implies that whenever the Lagrangian does not depend on q_i, then the corresponding generalised momentum p_i is conserved. This is a simple example of a more general result known as Noether's theorem, which we will come back to later.

As a very simple example, let us consider the Cartesian position coordinate x as our generalised coordinate. With the Lagrangian $L = \frac{1}{2}m\dot{x}^2 - V(x)$, the generalised momentum is simply the conventional momentum $p = \partial L/\partial \dot{x} = m\dot{x}$.

3.3 Precession of a Symmetric Top

Let us now consider an example of the use of Lagrangian mechanics to solve a real problem: a symmetric top. The kinetic energy is given in terms of Euler angles by Eq. (2.8.4) whereas the gravitational potential energy is $V = MgR\cos\theta$. This leaves us with a Lagrangian

$$L = \tfrac{1}{2}I_1\dot{\phi}^2\sin^2\theta + \tfrac{1}{2}I_1\dot{\theta}^2 + \tfrac{1}{2}I_3\left(\dot{\psi} + \dot{\phi}\cos\theta\right)^2 - MgR\cos\theta\,. \tag{3.3.1}$$

The Euler-Lagrange equation (3.2.1) for θ is

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(I_1\dot{\theta}\right) = I_1\dot{\phi}^2\sin\theta\cos\theta - I_3\left(\dot{\psi} + \dot{\phi}\cos\theta\right)\dot{\phi}\sin\theta + MgR\sin\theta\,. \tag{3.3.2}$$

The Lagrangian function (3.3.1) does not contain the other two Euler angles ϕ and ψ so the generalised momenta $p_\phi = \partial L/\partial\dot{\phi}$ and $p_\psi = \partial L/\partial\dot{\psi}$ are constant

$$\frac{\mathrm{d}}{\mathrm{d}t}\left[I_1\dot{\phi}\sin^2\theta + I_3\left(\dot{\psi} + \dot{\phi}\cos\theta\right)\cos\theta\right] = 0 \tag{3.3.3}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}\left[I_3\left(\dot{\psi} + \dot{\phi}\cos\theta\right)\right] = 0\,. \tag{3.3.4}$$

Note that comparison of Eqs. (3.3.4) and (2.8.2) tells us that

$$\omega_3 = \dot{\psi} + \dot{\phi}\cos\theta = \text{constant}. \tag{3.3.5}$$

We are interested in the situation of steady precession at a constant angle θ. In this case we conclude from Eqs. (3.3.3) and (3.3.4) that $\dot{\phi}$ and $\dot{\psi}$ are constant. Hence the axis of the top precesses

around the vertical with a constant angular velocity, which we denote by Ω, i.e., $\dot{\phi} = \Omega$. Because we are looking for a solution with fixed θ, the left side of Eq. (3.3.2) must vanish, and we obtain

$$I_1\Omega^2\cos\theta - I_3\omega_3\Omega + MgR = 0\,, \tag{3.3.6}$$

which we can solve for Ω. The general solution is

$$\Omega = \frac{I_3\omega_3 \pm \sqrt{I_3^2\omega_3^2 - 4I_1MgR\cos\theta}}{2I_1\cos\theta}\,. \tag{3.3.7}$$

This only has real roots if

$$\omega_3^2 \geq \omega_c^2 \equiv \frac{4I_1MgR\cos\theta}{I_3^2}\,. \tag{3.3.8}$$

If the top is spinning more slowly than this, there is no solution with constant θ. Instead, the top starts to wobble.

For a rapidly spinning top, $\omega_3 \gg \omega_c$, we can expand the square root in Eq. (3.3.7) to obtain

$$\Omega \approx \frac{I_3\omega_3 \pm I_3\omega_3\left(1 - 2\frac{I_1MgR\cos\theta}{I_3^2\omega_3^2}\right)}{2I_1\cos\theta} \to \begin{cases} MgR/I_3\omega_3 & -\text{sign} \\ I_3\omega_3/I_1\cos\theta & +\text{sign} \end{cases} \tag{3.3.9}$$

The first of these is the precession frequency calculated in (2.6.6), while the second is the precession of a free system discussed in section 2.8. Note the absence of any contribution from gravity in the second expression.

Note that if $\theta > \frac{1}{2}\pi$, the top is hanging *below* its point of support, and there is no limit on ω_3. In particular, for $\omega_3 = 0$, we find the possible angular velocities of a compound pendulum swinging in a circle

$$\Omega = \pm\sqrt{\frac{MgR}{I_1\,|\cos\theta|}} \tag{3.3.10}$$

3.4 Constraints

Consider a system of N particles in three dimensions. To specify the position of each particle, you need $3N$ generalised coordinates. However, in many cases the coordinates are not all independent but subject to some constraints, such as the rigidity conditions (see Section 2.2), which reduce the number of generalised coordinates required. For a rigid body, the original $3N$ coordinates may be reduced to six generalised coordinates: three translational, such as the coordinates X, Y, Z of the centre of mass, and three rotational, such as the Euler angles, ϕ, θ, ψ.

We will assume that the constraints can be written in the form $f(x_1, \ldots, x_{3N}, t) = 0$. Constraints like that are called *holonomic*. For a rigid body, the constraints are of this form: The distance between each pair of particles i and j is fixed to a constant d_{ij}, and therefore one has

$$(\mathbf{r}_i - \mathbf{r}_j)^2 - d_{ij}^2 = 0. \tag{3.4.1}$$

Another example is motion on the surface of a sphere of radius R, for which the constraint is

$$f(x, y, z) = z^2 + y^2 + z^2 - R^2 = 0. \tag{3.4.2}$$

Sometimes one has to deals with non-holonomic constraints, for example if the constraint depends on velocities, but these are more complicated to handle, and we will not discuss them in this course.

In principle, solving each constraint equation eliminates one coordinate. If one has initially N coordinates x_i (with $i = 1, \ldots, N$) and C constraints, solving them will allow one to express the original coordinates in terms of $M = N - C$ generalised coordinates q_j, with $j = 1, \ldots, M$,

$$x_i = x_i(q_1, \ldots, q_M, t), \tag{3.4.3}$$

with possibly explicit time-dependence if the constraints are time-dependent. In that case the system is called *forced*, otherwise it is *natural*.

If one can solve the constraints and find the explicit relations (3.4.3), one can then write the Lagrangian in terms of the generalised coordinates q_j and solve the Euler-Lagrange equation. Substituting this solution to Eq. (3.4.3) then gives the solution in terms of the original coordinates.

An alternative approach, which is sometimes useful, is to implement the constraints using *Lagrange multipliers*. Starting with a Lagrangian $L(x_1, \ldots, x_N)$ and a constraint function $f(x_1, \ldots, x_N)$, we define a new Lagrangian L'

$$L'(x_1, \ldots, x_N, \lambda) = L(x_1, \ldots, x_N) + \lambda f(x_1, \ldots, x_N), \tag{3.4.4}$$

which is a function of the original coordinates and a Lagrange multiplier λ. If we now treat λ as the $(N+1)$th coordinate, its Euler-Lagrange equation is

$$\frac{d}{dt}\frac{\partial L'}{\partial \dot{\lambda}} - \frac{\partial L'}{\partial \lambda} = -f(x_1, \ldots, x_N) = 0, \tag{3.4.5}$$

and therefore it satisfies the constraint automatically. The Euler-Lagrange equations for the original coordinates are

$$\frac{d}{dt}\frac{\partial L}{\partial \dot{x}_i} - \frac{\partial L}{\partial x_i} - \lambda(t)\frac{\partial f}{\partial x_i} = 0, \tag{3.4.6}$$

where the extra term can be interpreted as the (generalised) force that has to be applied to the system to enforce the constraint.

As an example, consider a mass m hanging from a rope that is wrapped around a pulley of radius R and moment of inertia I. Using the vertical position z of the mass, and the angle θ of the pulley as the coordinates, the Lagrangian is

$$L = T - V = \frac{1}{2}m\dot{z}^2 + \frac{1}{2}I\dot{\theta}^2 - mzg. \tag{3.4.7}$$

If the rope does not slip, the pulley has to rotate as the mass moves, and this imposes a constraint $R\dot{\theta} = -\dot{z}$. Choosing the origin appropriately, we can write this as a holonomic constraint

$$f(\theta, z) = R\theta + z = 0. \tag{3.4.8}$$

Introducing the Lagrange multiplier λ, the new Lagrangian is

$$L' = L + \lambda(R\theta + z) = \frac{1}{2}m\dot{z}^2 + \frac{1}{2}I\dot{\theta}^2 - mzg + \lambda(R\theta + z). \tag{3.4.9}$$

The Euler-Lagrange equations are

$$\begin{aligned} \frac{d}{dt}m\dot{z} + mg - \lambda &= 0 \quad \text{for } z, \\ \frac{d}{dt}I\dot{\theta} - \lambda R &= 0 \quad \text{for } \theta, \\ -(R\theta + z) &= 0 \quad \text{for } \lambda. \end{aligned} \tag{3.4.10}$$

The third equation implements the constraint $\theta = -z/R$, and substituting this to the first two gives

$$\begin{aligned} m\ddot{z} + mg - \lambda &= 0, \\ -\frac{I}{R}\ddot{z} - \lambda R &= 0. \end{aligned} \tag{3.4.11}$$

Solving this pair of equations for $\ddot{z}$ and λ, we obtain

$$\begin{aligned} \left(m + \frac{I}{R^2}\right)\ddot{z} + mg &= 0, \\ \lambda &= \frac{mg}{1 + mR^2/I}. \end{aligned} \tag{3.4.12}$$

The first line shows that the moment of inertia I of the pulley gives the mass extra inertia. The second line gives the force that rope has to apply in order to enforce the constraint. This is just the tension of the rope.

3.5 Normal Modes

3.5.1 Orthogonal Coordinates

Instead of rigid constraints, let us now consider a situation where the constraints are flexible so that the particles can move around their equilibrium positions. We assume that the system is described by N generalised coordinates q_i. We also assume that it is *natural*, which means that the kinetic energy is a quadratic homogeneous function of the generalised velocities. We can then write it as

$$T = \frac{1}{2}\sum_{ij} a_{ij}(q_1, \ldots, q_N)\dot{q}_i\dot{q}_j, \tag{3.5.1}$$

where the coefficients a_{ij} can depend on the coordinates q_i but not on velocities $\dot{q}_i$. They can chosen to be symmetric ($a_{ji} = a_{ij}$) without any loss of generality.

The coordinates are said to be *orthogonal* if there are no cross terms, i.e., $a_{ij} = 0$ if $i \neq j$. Then the kinetic energy is simply

$$T = \frac{1}{2}\sum_{i} a_{ii}(q_1, \ldots, q_N)\dot{q}_i^2. \tag{3.5.2}$$

We can always make our coordinates orthogonal by using the *Gram-Schmidt* procedure. For example, if $N = 2$, the general form of Eq. (3.5.1) is

$$T = \tfrac{1}{2}a_{11}\dot{q}_1^2 + a_{12}\dot{q}_1\dot{q}_2 + \tfrac{1}{2}a_{22}\dot{q}_2^2\,. \tag{3.5.3}$$

Defining a new coordinate

$$q_1' = q_1 + \frac{a_{12}}{a_{11}}q_2\,, \tag{3.5.4}$$

the kinetic energy becomes

$$T = \tfrac{1}{2}a_{11}\dot{q}_1'^2 + \tfrac{1}{2}a_{22}'\dot{q}_2^2 \qquad \text{with} \quad a_{22}' = a_{22} - \frac{a_{12}^2}{a_{11}}\,. \tag{3.5.5}$$

Furthermore, if we rescale the coordinates in Eq. (3.5.2) by

$$q_i' = \sqrt{a_{ii}}q_i, \tag{3.5.6}$$

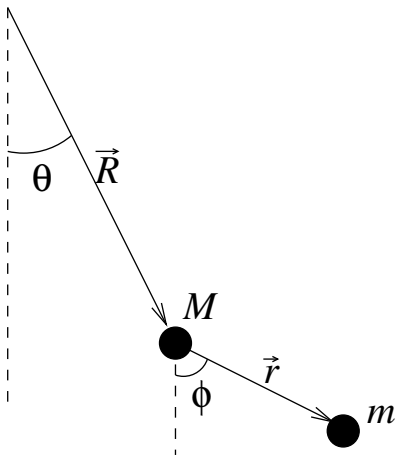

Figure 3.1: Double pendulum

the kinetic energy becomes

$$T = \frac{1}{2}\sum_i \dot{q}_i'^2. \tag{3.5.7}$$

Therefore we can always assume that the kinetic energy has this form.

As an example, consider a double pendulum, with a second pendulum hanging from the first. The kinetic energy is

$$T = \frac{1}{2}M\dot{\mathbf{R}}^2 + \frac{1}{2}m\left(\dot{\mathbf{R}} + \dot{\mathbf{r}}\right)^2 = \tfrac{1}{2}MR^2\dot{\theta}^2 + \tfrac{1}{2}m\left[R^2\dot{\theta}^2 + r^2\dot{\phi}^2 + 2Rr\dot{\theta}\dot{\phi}\cos(\theta-\phi)\right] \tag{3.5.8}$$

where M, R and θ refer to the upper pendulum and m, r and ϕ to the lower. Note that the kinetic energy of the lower pendulum depends not only on ϕ but also on the motion of the upper pendulum to which it is attached. For small values of θ and ϕ we can set the cosine term to one, so that the kinetic energy is

$$T = \tfrac{1}{2}(M+m)R^2\dot{\theta}^2 + \tfrac{1}{2}mr^2\dot{\phi}^2 + mRr\dot{\theta}\dot{\phi}. \tag{3.5.9}$$

However, because of the last term, the coordinates are not orthogonal. In this case it is obvious that we could get an orthogonal set by simply considering the displacement of the 2 bobs. For small angles this gives

$$x = R\theta \qquad y = R\theta + r\phi\,, \tag{3.5.10}$$

and the kinetic energy, T, becomes

$$T = \tfrac{1}{2}M\dot{x}^2 + \tfrac{1}{2}m\dot{y}^2\,. \tag{3.5.11}$$

Finally, we can reduce this to the standard form (3.5.7) using

$$q_1 = \sqrt{M}x \qquad q_2 = \sqrt{m}y\,. \tag{3.5.12}$$

3.5.2 Small Oscillations

We now consider the potential energy, V. If T is given by (3.5.7), the Lagrangian is

$$L = T - V = \frac{1}{2}\sum_i \dot{q}_i'^2 - V(q_1,\ldots,q_N). \tag{3.5.13}$$

The Euler-Lagrange equations are then simply

$$\ddot{q}_i = -\frac{\partial V}{\partial q_i}\,, \tag{3.5.14}$$

for all i.

If we now assume that the amplitude of the oscillations is small, we can Taylor expand the potential around the origin, which was chosen to correspond to the equilibrium state. To quadratic order we have

$$V(q_1, \ldots, q_N) = V_0 + \sum_i b_i q_i + \tfrac{1}{2}\sum_{ij} k_{ij} q_i q_j + O(q^3), \tag{3.5.15}$$

where V_0, b_i and k_{ij} are constants.

Because we can subtract a constant from the potential without changing the equations of motion (3.5.14), we are free to choose $V_0 = 0$. Because we are assuming that the equilibrium state is $q_i = 0$ for all i, we find that $b_i = 0$. Therefore, we only need to consider the quadratic term

$$V = \tfrac{1}{2}\sum_{ij} k_{ij} q_i q_j, \tag{3.5.16}$$

Note that we can also choose k_{ij} to be symmetric, i.e., $k_{ji} = k_{ij}$ without any loss of generality.

With the potential (3.5.16), the Euler-Lagrange equation (3.5.14) is simply

$$\ddot{q}_i = -\sum_j k_{ij} q_j, \tag{3.5.17}$$

which can be written in matrix form as

$$\frac{\mathrm{d}^2}{\mathrm{d}t^2}\begin{pmatrix} q_1 \\ q_2 \\ \vdots \\ q_N \end{pmatrix} = -\begin{pmatrix} k_{11} & k_{12} & \cdots & k_{1N} \\ k_{21} & k_{22} & \cdots & k_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ k_{N1} & k_{N2} & \cdots & k_{NN} \end{pmatrix}\begin{pmatrix} q_1 \\ q_2 \\ \vdots \\ q_N \end{pmatrix}. \tag{3.5.18}$$

By defining an N-dimensional coordinate vector $\mathbf{q} = (q_1, \ldots, q_N)$ and an $N \times N$ matrix $\overline{\overline{\mathbf{k}}}$ with elements k_{ij}, we can also write it more compactly as

$$\ddot{\mathbf{q}} = -\overline{\overline{\mathbf{k}}} \cdot \mathbf{q}. \tag{3.5.19}$$

In the same notation, the Lagrangian is

$$L = \frac{1}{2}\dot{\mathbf{q}} \cdot \dot{\mathbf{q}} - \frac{1}{2}\mathbf{q} \cdot \overline{\overline{\mathbf{k}}} \cdot \mathbf{q}. \tag{3.5.20}$$

In our double pendulum example, the potential is

$$\begin{aligned} V(\theta, \phi) &= MgR(1 - \cos\theta) + mg\left[R(1 - \cos\theta) + r(1 - \cos\phi)\right] \\ &\approx \frac{1}{2}(M + m)gR\theta^2 + \frac{1}{2}mgr\phi^2 \\ &= \frac{M + m}{2R}gx^2 + \frac{mg}{2r}(y - x)^2 \\ &= \frac{1}{2}\left(1 + \frac{m}{M}\frac{R + r}{r}\right)\frac{g}{R}q_1^2 + \frac{g}{2r}q_2^2 - \sqrt{\frac{m}{M}}\frac{g}{r}q_1 q_2. \end{aligned} \tag{3.5.21}$$

The coefficient matrix is therefore

$$\overline{\overline{\mathbf{k}}} = \begin{pmatrix} \left(1 + \frac{m}{M}\frac{R+r}{r}\right)\frac{g}{R} & -\sqrt{\frac{m}{M}}\frac{g}{r} \\ -\sqrt{\frac{m}{M}}\frac{g}{r} & \frac{g}{r} \end{pmatrix}. \tag{3.5.22}$$

3.5.3 Eigenvalue Problem

Eq. (3.5.19) is a set of N coupled linear second-order equation. Therefore we should find $2N$ linearly independent solutions.

We look for solutions of the form

$$\mathbf{q}(t) = \mathbf{A}e^{i\omega t}, \tag{3.5.23}$$

where $\mathbf{A}$ and ω are constants. We will check later that we have found all $2N$ solutions.

Substituting the Ansatz (3.5.23) into Eq. (3.5.19) gives

$$-\omega^2 \mathbf{A}e^{i\omega t} = -\overline{\overline{\mathbf{k}}} \cdot \mathbf{A}e^{i\omega t}, \tag{3.5.24}$$

which is equivalent to

$$\overline{\overline{\mathbf{k}}} \cdot \mathbf{A} = \omega^2 \mathbf{A}. \tag{3.5.25}$$

This has the form of the eigenvalue equation: It shows that $\mathbf{A}$ is an eigenvector of the matrix $\overline{\overline{\mathbf{k}}}$ with eigenvalue ω^2.

We know from linear algebra that the eigenvalues of a matrix are given by solutions of the characteristic equation

$$\det(\overline{\overline{\mathbf{k}}} - \omega^2 \overline{\overline{\mathbf{1}}}) \equiv \begin{vmatrix} k_{11} - \omega^2 & k_{12} & \cdots & k_{1N} \\ k_{21} & k_{22} - \omega^2 & \cdots & k_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ k_{N1} & k_{N2} & \cdots & k_{NN} - \omega^2 \end{vmatrix} = 0. \tag{3.5.26}$$

Once we have found the eigenvalue ω^2, we can substitute it to Eq. (3.5.25) to find the eigenvector.

Because $\overline{\overline{\mathbf{k}}}$ is a symmetric $N \times N$ matrix, it has N real eigenvalues ω_α^2, where $\alpha = 1, \ldots, N$, corresponding to N eigenvectors $\mathbf{A}_\alpha$, which we can choose to be orthonormal,

$$\mathbf{A}_\alpha \cdot \mathbf{A}_\beta = \delta_{\alpha\beta}. \tag{3.5.27}$$

For each eigenvalue ω_α^2, there are two independent solutions,

$$\mathbf{q}_\alpha^+(t) = \mathbf{A}_\alpha e^{i\omega_\alpha t} \quad \text{and} \quad \mathbf{q}_\alpha^-(t) = \mathbf{A}_\alpha e^{-i\omega_\alpha t}. \tag{3.5.28}$$

For N eigenvalues, this takes the total number of linearly independent solutions to $2N$, proving that we have the complete solution.

The frequencies ω_α are real because the eigenvalues ω_α^2 have to be non-negative: If we had $\omega_\alpha^2 < 0$, then for the corresponding eigenvector $\mathbf{A}_\alpha$ and small ϵ we would have

$$V(\epsilon \mathbf{A}_\alpha) = \frac{1}{2}\epsilon^2 \mathbf{A}_\alpha \cdot \overline{\overline{\mathbf{k}}} \cdot \mathbf{A}_\alpha = \frac{1}{2}\epsilon^2 \omega_\alpha^2 < 0, \tag{3.5.29}$$

meaning that $\mathbf{q} = 0$ could not be the minimum of the potential as we assumed.

The individual solutions $\mathbf{q}_\alpha^\pm$ are complex, so physical solutions have to be real linear combinations of them,

$$\begin{aligned} \mathbf{q}_\alpha(t) &= a_+ \mathbf{q}_\alpha^+(t) + a_- \mathbf{q}_\alpha^-(t) = \left(a_+ e^{i\omega_\alpha t} + a_- e^{-i\omega_\alpha t}\right) \mathbf{A}_\alpha \\ &= \left[(a_+ + a_-)\cos\omega_\alpha t + i(a_+ - a_-)\sin\omega_\alpha t\right] \mathbf{A}_\alpha. \end{aligned} \tag{3.5.30}$$

Because we want a real solution, the coefficients $a_1 \equiv (a_+ + a_-)$ and $a_2 = i(a_+ - a_-)$ have to be real, and we can equally well write

$$\mathbf{q}_\alpha(t) = (a_1 \cos\omega_\alpha t + a_2 \sin\omega_\alpha t)\, \mathbf{A}_\alpha, \tag{3.5.31}$$

which we can also write a pure cosine term with a phase shift,

$$\mathbf{q}_\alpha(t) = c\cos(\omega_\alpha t + \phi)\mathbf{A}_\alpha, \tag{3.5.32}$$

where c and ϕ are real constants that have to be determined from the initial conditions.

Finally, the general solution is a linear combination of solutions of the form (3.5.32),

$$\mathbf{q}(t) = \sum_{\alpha=1}^{N} c_\alpha \cos(\omega_\alpha t + \phi_\alpha)\mathbf{A}_\alpha. \tag{3.5.33}$$

The modes of vibration of the system, i.e., the individual solutions (3.5.32) are known as *normal modes.* For a forced system there are resonances at the frequencies of the normal modes.

It is often useful to use the normal modes to define a set of generalised coordinates known as the *normal coordinates*, which we denote by $\tilde{q}_\alpha$. They are defined by expressing the original coordinates q_i in terms of the eigenvectors $\mathbf{A}_\alpha$ as

$$\mathbf{q} = \sum_\alpha \tilde{q}_\alpha \mathbf{A}_\alpha. \tag{3.5.34}$$

Substituting this to Eq. (3.5.20) we find that the Lagrangian becomes simply

$$L = \sum_\alpha \left(\frac{1}{2}\dot{\tilde{q}}_\alpha^2 - \frac{1}{2}\omega_\alpha^2 \tilde{q}_\alpha^2 \right). \tag{3.5.35}$$

The Euler-Lagrange equations are

$$\ddot{\tilde{q}}_\alpha + \omega_\alpha^2 \tilde{q}_\alpha = 0, \tag{3.5.36}$$

which means that each normal coordinate $\tilde{q}_\alpha$ oscillates independently of all others with its own normal frequency ω_α.

3.6 Continuous Systems

In addition to mechanical systems consisting of a finite number of degrees of freedom, one is often also interested in continuous systems, for example waves propagation in continuous media or field theories describing particle physics or electromagnetism.

To see how continuous systems are described in the Lagrangian formulation, consider a stretched string. We assume that in equilibrium the string is stretched to length ℓ_0 and has tension k. The displacement of the string from its equilibrium position is given by the continuous function $y(x,t)$, where $x \in \{0, \ell_0\}$. To describe the time and space derivatives, we use the notation

$$\dot{y} = \frac{\partial y}{\partial t}, \quad y' = \frac{\partial y}{\partial x}. \tag{3.6.1}$$

The Lagrangian is still given by the difference of the kinetic and potential energies, $L = T - V$. The kinetic energy can be calculated by considering an infinitesimal segment of length dx. The mass

of such a segment is $dm = \mu dx$ where the constant μ is the mass per unit length. The velocity of the segment is simply $\dot{y}$, and therefore the kinetic energy of the infinitesimal segment is

$$dT = \frac{1}{2}\mu\, dx\, \dot{y}^2. \tag{3.6.2}$$

Integrating over the who distance ℓ_0, we find the total kinetic energy

$$T = \int_0^{\ell_0} dx\, \frac{1}{2}\mu\, \dot{y}^2. \tag{3.6.3}$$

The potential energy is V of the string is due to its tension k,

$$V = k(\ell - \ell_0), \tag{3.6.4}$$

where ℓ is the length of the displaced string. Again, this can be calculated by considering an infinitesimal segment of length dx. According to Pythagoras theorem, the length of the segment is

$$d\ell = \sqrt{dx^2 + dy^2} = \sqrt{1 + y'^2}dx \approx \left(1 + \frac{1}{2}y'^2\right) dx, \tag{3.6.5}$$

where we have assumed that the displacement is small and smooth so that $y' \ll 1$, and Taylor expanded to quadratic order. The length of the string is then obtained by summing over all the infinitesimal segments

$$\ell = \int_0^{\ell_0} dx \left(1 + \frac{1}{2}y'^2\right) = \ell_0 + \int_0^{\ell_0} dx \frac{1}{2}y'^2, \tag{3.6.6}$$

and therefore the potential energy is

$$V = \int_0^{\ell_0} dx \frac{1}{2}ky'^2. \tag{3.6.7}$$

We can now write the whole Lagrangian,

$$L = T - V = \int_0^{\ell_0} dx \left(\frac{1}{2}\mu\dot{y}^2 - \frac{1}{2}ky'^2\right). \tag{3.6.8}$$

The integrand is called the Lagrangian density and denoted by $\mathcal{L}$, i.e., $L = \int dx\mathcal{L}$, where

$$\mathcal{L} = \frac{1}{2}\mu\dot{y}^2 - \frac{1}{2}ky'^2. \tag{3.6.9}$$

The action is then given by an integral over both time and space,

$$S = \int dt \int dx \mathcal{L}(\dot{y}, y', y). \tag{3.6.10}$$

Variation of the action is

$$\delta S = \int dt \int dx \left[\frac{\partial \mathcal{L}}{\partial \dot{y}}\delta\dot{y} + \frac{\partial \mathcal{L}}{\partial y'}\delta y' + \frac{\partial \mathcal{L}}{\partial y}\delta y\right] = \int dt \int dx \left[-\frac{d}{dt}\frac{\partial \mathcal{L}}{\partial \dot{y}} - \frac{d}{dx}\frac{\partial \mathcal{L}}{\partial y'} + \frac{\partial \mathcal{L}}{\partial y}\right]\delta y. \tag{3.6.11}$$

The action principle $\delta S = 0$ therefore leads to the Euler-Lagrange equation

$$\frac{d}{dt}\frac{\partial \mathcal{L}}{\partial \dot{y}} + \frac{d}{dx}\frac{\partial \mathcal{L}}{\partial y'} - \frac{\partial \mathcal{L}}{\partial y} = 0. \tag{3.6.12}$$

Substituting the Lagrangian density (3.6.9) for the string, we find the equation of motion

$$\frac{d}{dt}\left(\mu\dot{y}\right) + \frac{d}{dx}\left(-ky'\right) = \mu\ddot{y} - ky'' = 0, \tag{3.6.13}$$

which is the wave equation, as one would expect.

Chapter 4

Hamiltonian Mechanics

4.1 Hamilton's Equations[1]

In the Lagrangian formulation of mechanics, the state of the system is descibed by a set of N generalised coordinates q_i and and their time derivatives $\dot{q}_i$. For brevity we shall write it in the form $L(q, \dot{q})$ and let the q and $\dot{q}$ stand for the whole set.

In Eq. (3.2.11) we defined the generalised momentum $p_i(q, \dot{q}) = \partial L/\partial \dot{q}_i$. The basic idea of the Hamiltonian formulation of mechanics is to use coordinates and momenta (q, p) rather than coordinates and velocities $(q, \dot{q})$ to parameterise the system. The space of momenta and coordinates is known as the *phase space*, so the Hamiltonian describes the evolution in phase space. Although the two formulations are mathematically and physically equivalent, the Hamiltonian formulation provides new insight especially when moving to quantum mechanics which is usually described in the Hamiltonian formulation.

To move from velocities to momenta, one has to invert Eq. (3.2.11) to find the velocities in terms of coordinates and momenta,

$$\dot{q}_i = \dot{q}_i(q, p). \tag{4.1.1}$$

Having done this, we define the *Hamiltonian function* H by Legendre transformation

$$H(q, p) = \sum_{i=1}^{N} p_i \dot{q}_i(q, p) - L\left(q, \dot{q}(q, p)\right). \tag{4.1.2}$$

Note that H is a function of coordinates q_i and momenta p_i, whereas the Lagrangian L is a function of coordinates q_i and velocities $\dot{q}_i$. This is analogous to transformations between different thermodynamic potentials in statistical physics. We will see that we can obtain the equations of motion of the system from the Hamiltonian, and therefore it determines the dynamics of the system.

As a simple example, consider a particle in a one-dimensional potential $V(x)$. The Lagrangian is

$$L - \tfrac{1}{2} m\dot{x}^2 - V(x). \tag{4.1.3}$$

In this case, the generalised momentum p is just the usual momentum

$$p = \frac{\partial L}{\partial \dot{x}} = m\dot{x}, \tag{4.1.4}$$

[1] Kibble & Berkshire, chapter 12

which we can easily invert to find $\dot{x} = p/m$. Therefore the Hamiltonian is

$$H(x,p) = p\dot{x} - L(x,\dot{x}) = \frac{p^2}{m} - \frac{p^2}{2m} + V(x) = \frac{p^2}{2m} + V(x). \tag{4.1.5}$$

This Hamiltonian is, of course, familiar from quantum mechanics.

In order to see how the equations of motion arise from the Hamiltonian, let us calculate its derivatives with respect to momenta and coordinates. Using Eqs. (4.1.2) and (3.2.11), we find

$$\frac{\partial H}{\partial p_j} = \dot{q}_j + \sum_i p_i \frac{\partial \dot{q}_i}{\partial p_j} - \sum_i \frac{\partial L}{\partial \dot{q}_i}\frac{\partial \dot{q}_i}{\partial p_j} = \dot{q}_j, \tag{4.1.6}$$

and similarly for the derivative with respect to the coordinate q_j,

$$\frac{\partial H}{\partial q_j} = \sum_i p_i \frac{\partial \dot{q}_i}{\partial q_j} - \frac{\partial L}{\partial q_j} - \sum_i \frac{\partial L}{\partial \dot{q}_i}\frac{\partial \dot{q}_i}{\partial q_j} = -\frac{\partial L}{\partial q_j} = -\frac{d}{dt}\frac{\partial L}{\partial \dot{q}_j} = -\dot{p}_j, \tag{4.1.7}$$

where we used the Euler-Lagrange equation (3.2.1) in the last step. These results are known as *Hamilton's equations*,

$$\begin{aligned} \dot{q}_i &= \frac{\partial H}{\partial p_i}, \\ \dot{p}_i &= -\frac{\partial H}{\partial q_i}. \end{aligned} \tag{4.1.8}$$

Together, they determine the evolution of coordinates and momenta, and they are therefore the equations of motion in the Hamiltonian formulation.

We can also calculate the time derivative of the Hamiltonian,

$$\begin{aligned} \frac{dH}{dt} &= \frac{\partial H}{\partial t} + \sum_i \frac{\partial H}{\partial q_i}\dot{q}_i + \sum_i \frac{\partial H}{\partial p_i}\dot{p}_i \\ &= \frac{\partial H}{\partial t} + \sum_i \left(\frac{\partial H}{\partial q_i}\frac{\partial H}{\partial p_i} - \frac{\partial H}{\partial p_i}\frac{\partial H}{\partial q_i} \right) = \frac{\partial H}{\partial t}. \end{aligned} \tag{4.1.9}$$

In particular this means that if the Hamiltonian has no explicit time dependence (i.e. $\partial H/\partial t = 0$), then the Hamiltonian is conserved.

To understand better the physical meaning of the Hamiltonian, let us assume that the system is *natural* (see Section 3.5.1), so that the kinetic energy can be written as

$$T = \frac{1}{2}\sum_{ij} a_{ij}\dot{q}_i\dot{q}_j. \tag{4.1.10}$$

The Lagrangian is $L = T - V$, and therefore the momenta are

$$p_i = \frac{\partial L}{\partial \dot{q}_i} = \sum_j a_{ij}\dot{q}_j. \tag{4.1.11}$$

Using this, we can write the Hamiltonian as

$$H = \sum_i p_i\dot{q}_i - L = \sum_{ij} a_{ij}\dot{q}_i\dot{q}_j - \frac{1}{2}\sum_{ij} a_{ij}\dot{q}_i\dot{q}_j + V(q) = \frac{1}{2}\sum_{ij} a_{ij}\dot{q}_i\dot{q}_j + V(q) = T + V. \tag{4.1.12}$$

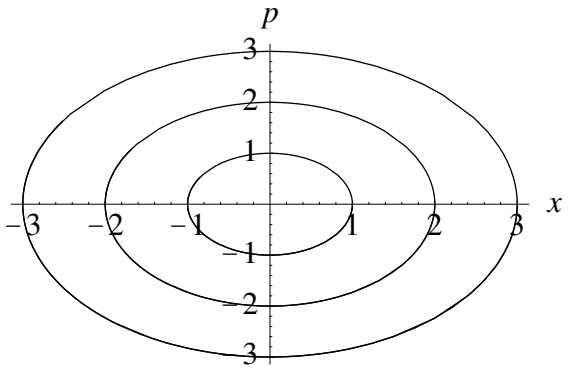

Figure 4.1: Phase space trajectories of the harmonic oscillator.

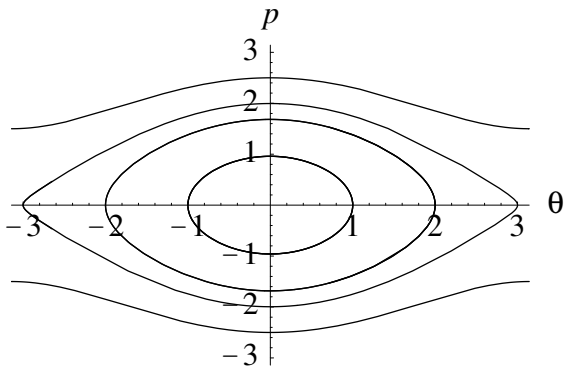

Figure 4.2: Phase space trajectories of a non-linear pendulum.

This demonstrates that the Hamiltonian is nothing but the total energy of the system, and which also explains why it is conserved.

As a simple example, let us consider a particle in a harmonic potential. The Hamiltonian is

$$H = \frac{p^2}{2m} + \frac{1}{2}kx^2. \tag{4.1.13}$$

Hamilton's equations are

$$\begin{aligned} \dot{x} &= \frac{\partial H}{\partial p} = \frac{p}{m}, \\ \dot{p} &= -\frac{\partial H}{\partial x} = -kx. \end{aligned} \tag{4.1.14}$$

The evolution of the system can be desribed by a trajectory in phase space (x, p). In the harmonic case, the solutions are ellipses (see Fig. 4.1). Note that because Hamilton's equations are first-order, the phase space trajectories cannot cross, but they can have fixed points at which the derivatives vanish. One can often deduce the qualitative properties of the solutions by simply finding the fixed points and considering the behaviour near them. In the Harmonic case, there is one fixed point at $x = p = 0$.

A less trivial example is a non-linear simple pendulum. The Lagrangian is

$$L = \frac{1}{2}mR^2\dot{\theta}^2 + mgR\cos\theta. \tag{4.1.15}$$

The generalised momentum is

$$p_\theta = \frac{\partial L}{\partial \dot{\theta}} = mR^2\dot{\theta}, \tag{4.1.16}$$

from which we find

$$\dot{\theta} = \frac{p_\theta}{mR^2}. \tag{4.1.17}$$

Therefore the Hamiltonian is

$$H = p_\theta\dot{\theta} - L = \frac{p_\theta^2}{mR^2} - \frac{1}{2}\frac{p_\theta^2}{mR^2} - mgR\cos\theta = \frac{1}{2}\frac{p_\theta^2}{mR^2} - mgR\cos\theta. \tag{4.1.18}$$

Hamilton's equations are

$$\dot{\theta} = \frac{\partial H}{\partial p_\theta} = \frac{p_\theta}{mR^2}, \quad \dot{p}_\theta = -\frac{\partial H}{\partial \theta} = -mgR\sin\theta. \tag{4.1.19}$$

These equations have fixed points at $p_\theta = 0$ and $\theta = n\pi$ where n is an arbitrary integer. Even integers correspond to the minimum-energy state in which the pendulum is pointing down, and the odd integers are unstable states in which the pendulum points exactly up. By considering the qualitative motion we can draw a phase space diagram of the trajectories (Fig. 4.2) without having to actually solve the equations.

4.2 Poisson Brackets

Suppose now that we want to study the behaviour of some other quantity which depends on p and q such that its value changes as p and q change under control of the Hamiltonian. Let $F(q, p, t)$ be such a quantity. The *total* time derivative of F is

$$\begin{aligned}\frac{\mathrm{d}F}{\mathrm{d}t} &= \frac{\partial F}{\partial t} + \sum_{i=1}^{N}\left(\frac{\partial F}{\partial q_i}\dot{q}_i + \frac{\partial F}{\partial p_i}\dot{p}_i\right) \\ &= \frac{\partial F}{\partial t} + \sum_{i=1}^{N}\left(\frac{\partial F}{\partial q_i}\frac{\partial H}{\partial p_i} - \frac{\partial F}{\partial p_i}\frac{\partial H}{\partial q_i}\right) \\ &= \frac{\partial F}{\partial t} + \{F, H\}\end{aligned} \tag{4.2.1}$$

where the quantity

$$\{F, H\} = \sum_{i=1}^{N}\left(\frac{\partial F}{\partial q_i}\frac{\partial H}{\partial p_i} - \frac{\partial F}{\partial p_i}\frac{\partial H}{\partial q_i}\right) \tag{4.2.2}$$

is known as a *Poisson bracket.* Compare this with the equation for an operator in quantum mechanics (in the Heisenberg picture)

$$\frac{\mathrm{d}\hat{F}}{\mathrm{d}t} = \frac{\partial \hat{F}}{\partial t} + \frac{1}{\mathrm{i}\hbar}[\hat{F}, \hat{H}], \tag{4.2.3}$$

where $[\hat{F}, \hat{H}] = \hat{F}\hat{H} - \hat{H}\hat{F}$ is the *commutator* of $\hat{F}$ and $\hat{H}$.

Let us note some basic properties of Poisson brackets. They are antisymmetric, $\{F, H\} = -\{H, F\}$. If F has no explicit time–dependence, $\frac{\partial F}{\partial t} = 0$, and the Poisson bracket $\{F, H\}$ is zero, then F is independent of time; F represents a conserved quantity. The Poisson brackets for coordinates and momenta are simply

$$\{q_i, p_j\} = \delta_{ij} \qquad \{q_i, q_j\} = 0 \qquad \{p_i, p_j\} = 0\,, \tag{4.2.4}$$

which resemble the canonical commutation relations in quantum mechanics. We can also write Hamilton's equations in terms of Poisson brackets as

$$\dot{p}_i = \{p_i, H\} \qquad \dot{q}_i = \{q_i, H\}\ . \tag{4.2.5}$$

4.3 Symmetries and Conservation Laws

We have already noted in Sec. 3.2 a relationship between symmetry and conservation laws. If the Lagrangian does not depend on coordinate q_i, then the corresponding momentum p_i is conserved. In terms of the Hamiltonian, we have the corresponding result that if H does not depend on q_i, then

$$\dot{p}_i = -\frac{\partial H}{\partial q_i} = 0, \tag{4.3.1}$$

so p_i is conserved.

So far, this result is only useful if we have chosen the coordinate system to reflect the symmetry. For example, if we have a two-dimensional problem with rotation symmetry, and we express it in polar coordinates (r, θ), the Hamiltonian does not depend on the polar angle θ, and we find that the corresponding momentum is conserved.

However, if we had chosen to use Cartesian coordinates (x, y), we could not use the above result. In this case, the Hamiltonian would be

$$H = \frac{p_x^2 + p_y^2}{2m} + V\left(\sqrt{x^2 + y^2}\right). \tag{4.3.2}$$

In order to treat a situation like this, we can consider an infinitesimal rotation. When we rotate the system by an infinitesimal angle $\delta\theta$, the coordinates change as

$$\begin{pmatrix} x \\ y \end{pmatrix} \to \begin{pmatrix} x + \delta x \\ y + \delta y \end{pmatrix}, \quad \text{where} \quad \begin{pmatrix} \delta x \\ \delta y \end{pmatrix} = \begin{pmatrix} -y\delta\theta \\ x\delta\theta \end{pmatrix}, \tag{4.3.3}$$

and correspondingly the momenta change as

$$\begin{pmatrix} p_x \\ p_y \end{pmatrix} \to \begin{pmatrix} p_x + \delta p_x \\ p_y + \delta p_y \end{pmatrix}, \quad \text{where} \quad \begin{pmatrix} \delta p_x \\ \delta p_y \end{pmatrix} = \begin{pmatrix} -p_y\delta\theta \\ p_x\delta\theta \end{pmatrix}. \tag{4.3.4}$$

If we find that the Hamiltonian does not change under such a rotation, we say that the Hamiltonian is *invariant*.

More generally, consider a transformation generated by some function $G(q, p)$ under which the coordinates and momenta change as

$$\delta q_i = \frac{\partial G}{\partial p_i}\delta\lambda, \quad \delta p_i = -\frac{\partial G}{\partial q_i}\delta\lambda, \tag{4.3.5}$$

where $\delta\lambda$ is an infinitesimal parameter. For the above rotation, the generator is

$$G = xp_y - yp_x = L_z, \tag{4.3.6}$$

i.e., the z component of the angular momentum.

Under this transformation, the Hamiltonian changes as

$$\begin{aligned} \frac{dH}{d\lambda} &= \sum_i \left(\frac{\partial H}{\partial q_i}\frac{\partial q_i}{\partial \lambda} + \frac{\partial H}{\partial p_i}\frac{\partial p_i}{\partial \lambda} \right) \\ &= \sum_i \left(\frac{\partial H}{\partial q_i}\frac{\partial G}{\partial p_i} - \frac{\partial H}{\partial p_i}\frac{\partial G}{\partial q_i} \right) = \{H, G\}. \end{aligned} \tag{4.3.7}$$

This means that the Hamiltonian is invariant under the transformation if $\{H, G\} = 0$. But if this is the case, then G is conserved because then

$$\frac{dG}{dt} = \{G, H\} = 0. \tag{4.3.8}$$

Therefore, we have found a more general form of Noether's theorem: If the Hamiltonian is invariant under a continuous transformation, then the generator G of the transformation is a conserved charge.

Familiar examples of this are rotations, for which the conserved charge is the angular momentum $\mathbf{L}$ as we saw above and translations, for which the conserved charge is momentum $\mathbf{p}$. To see the latter, consider $G = p_x$. Then we have

$$\delta x = \frac{\partial G}{\partial p_x}\delta\lambda = \delta\lambda, \quad \delta p_x = -\frac{\partial G}{\partial x}\delta\lambda = 0, \tag{4.3.9}$$

so this is, indeed, the generator of translations.

Conservation of energy (or, equivalently, the Hamiltonian H) can also be understood in this way. Choosing $G = H$ as the generator, the coordinates and momenta transform as

$$\begin{aligned} \delta q_i &= \frac{\partial H}{\partial p_i}\delta\lambda = \dot{q}_i\delta\lambda, \\ \delta p_i &= -\frac{\partial H}{\partial q_i}\delta\lambda = \dot{p}_i\delta\lambda. \end{aligned} \tag{4.3.10}$$

This means just time translation $t \to t + \delta\lambda$, so energy conservation is a direct consequence of time translation invariance.

These are deep results because they mean that the conservation of angular momentum, momentum and energy, which we derived earlier using the equations of motion, and actually very general results and are valid whenever the Hamiltonian has the corresponding symmetries, irrespective of it precise form.

4.4 Canonical Transformations

In addition to the coordinate transformations considered in the previous section, the Hamiltonian formulation allows even more freedom in how to parameterise the system. To see this, note that the time evolution of a function $F(q, p; t)$ depends only on the Poisson bracket (4.2.2)

$$\frac{dF}{dt} = \frac{\partial F}{\partial t} + \{F, H\}. \tag{4.4.1}$$

Therefore, if one finds another set of coordinates Q and momenta P that gives the same Poisson brackets, i.e.,

$$\{f, g\}_{Q,P} = \{f, g\}_{q,p} \tag{4.4.2}$$

for every pair of functions f and g, then they will be equally valid variables to parameterise the system. A transformation

$$\begin{aligned} q &\rightarrow Q(q,p), \\ p &\rightarrow P(p,q), \end{aligned} \tag{4.4.3}$$

that takes the original coordinates q and momenta p to the new ones satisfying Eq. (4.4.2) is known a *canonical transformation*. In fact, it turns out that it is enough to check that the Poisson brackets (4.2.4) for coordinates and momenta are unchanged under the transformation, i.e.,

$$\{Q_i, Q_j\}_{q,p} = \{P_i, P_j\}_{q,p} = 0, \quad \{Q_i, P_j\}_{q,p} = \delta_{ij}. \tag{4.4.4}$$

Any set of variables that satisfy these conditions are called *canonical conjugates*.

As an example, consider the Hamiltonian

$$H = \frac{p^2}{2m} + \frac{1}{2}kx^2 + cpx, \tag{4.4.5}$$

where c is a constant. The last term, which is linear in p, makes this Hamiltonian unusual, but we can use a canonical transformation to write it in a more standard form. To see how to do that, let us first rearrange the terms to write the Hamiltonian as

$$H = \frac{(p + cmx)^2}{2m} + \frac{1}{2}(k - c^2m)x^2. \tag{4.4.6}$$

This suggests a transformation

$$P = p + cmx, \quad Q = x. \tag{4.4.7}$$

It is easy to check that they satisfy the conditions (4.4.4),

$$\{Q, Q\} = \{P, P\} = 0, \quad \{Q, P\} = \frac{\partial Q}{\partial x}\frac{\partial P}{\partial p} - \frac{\partial Q}{\partial p}\frac{\partial P}{\partial x} = 1 \times 1 - 0 = 1. \tag{4.4.8}$$

Therefore Q and P and canonical conjugates, and (4.4.7) is a canonical transformation. In terms of Q and P, the Hamiltonian is simply

$$H = \frac{P^2}{2m} + \frac{1}{2}(k - c^2m)Q^2, \tag{4.4.9}$$

which is just a standard harmonic oscillator. This illustrates how one can sometimes use canonical transformations to turn the Hamiltonian into a familiar form, for which the solution is already known. For example, one can always remove a linear term in p by the transformation

$$Q = q, \quad P = p + f(q). \tag{4.4.10}$$

It is important to understand that canonical transformations generally mix coordinates and momenta, and therefore the distinction between them essentially disappears in the Hamiltonian formulation. Because of this, it is interesting to see what a canonical transformation means in the Lagrangian formulation, which uses only coordinates to describe the system. In general, the transformation

changes the Lagrangian, so we denote the new Lagrangian by L'. The Hamiltonian is unchanged, so we have

$$H = p\dot{q} - L = P\dot{Q} - L', \tag{4.4.11}$$

from which we find

$$L' = P\dot{Q} - p\dot{q} + L. \tag{4.4.12}$$

Considering, for example, the transformation (4.4.10), this is

$$L' = L + f(q)\dot{q}. \tag{4.4.13}$$

The extra term is just the total time derivative of the integral $F(q) = \int^q f(q')dq'$,

$$L' = L + \frac{dF}{dt}. \tag{4.4.14}$$

Such a total derivative changes the action by a constant, and therefore does not affect the dynamics.

Chapter 5

Electromagnetic Potentials

5.1 Vector Potential

The dynamics of the electromagnetic field is described by Maxwell's equations,

$$\begin{aligned}
\boldsymbol{\nabla}\cdot\mathbf{E} &= \frac{\rho}{\epsilon_0}, \quad \text{(Gauss's law)} \\
\boldsymbol{\nabla}\times\mathbf{E} &= -\frac{\partial\mathbf{B}}{\partial t}, \quad \text{(Faraday's law)} \\
\boldsymbol{\nabla}\cdot\mathbf{B} &= 0, \quad \text{(magnetic Gauss's law)} \\
\boldsymbol{\nabla}\times\mathbf{B} &= \mu_0\mathbf{J} + \mu_0\epsilon_0\frac{\partial\mathbf{E}}{\partial t}. \quad \text{(Ampère's law)}
\end{aligned} \tag{5.1.1}$$

You have learned in Electricity&Magnetism that in electrostatics, the electric field can be described by the electric potential ϕ as

$$\mathbf{E} = -\boldsymbol{\nabla}\phi\,. \tag{5.1.2}$$

Written in terms of the potential, Gauss's law $\boldsymbol{\nabla}\cdot\mathbf{E} = \rho/\epsilon_0$ becomes the Poisson's equation

$$\nabla^2\phi = -\frac{\rho}{\epsilon_0}. \tag{5.1.3}$$

In a medium with $\epsilon_r \neq 1$ these may be modified slightly but we shall stick to the simpler forms for this discussion.

On the other hand, Eq. (5.1.2) is clearly not sufficient in time-dependent problems, which can be seen by considering Faraday's law,

$$\boldsymbol{\nabla}\times\mathbf{E} = -\frac{\partial\mathbf{B}}{\partial t}\,. \tag{5.1.4}$$

Using Eq. (5.1.2), we can write the curl of the electric field as

$$\boldsymbol{\nabla}\times\mathbf{E} = -\boldsymbol{\nabla}\times\boldsymbol{\nabla}\phi\,, \tag{5.1.5}$$

but this is vanishes because the curl of a gradient is identically zero. Therefore Eqs. (5.1.2) and (5.1.5) are incompatible.

To describe time-dependent situations, we introduce a *vector potential* $\mathbf{A}$, which is related to the electric and magnetic fields by

$$\begin{aligned}
\mathbf{B} &= \boldsymbol{\nabla}\times\mathbf{A}, \\
\mathbf{E} &= -\frac{\partial\mathbf{A}}{\partial t} - \boldsymbol{\nabla}\phi.
\end{aligned} \tag{5.1.6}$$

Let us see how Maxwell's equations (5.1.1) appear in terms of ϕ and $\mathbf{A}$. First, magnetic Gauss's law is trivally satisfied

$$\boldsymbol{\nabla}\cdot\mathbf{B}=\boldsymbol{\nabla}\cdot(\boldsymbol{\nabla}\times\mathbf{A})=0, \tag{5.1.7}$$

because the divergence of a curl is identically zero.[1]

Faraday's law is also automatically satisfied,

$$\boldsymbol{\nabla}\times\mathbf{E}=\boldsymbol{\nabla}\times\left(-\frac{\partial\mathbf{A}}{\partial t}-\boldsymbol{\nabla}\phi\right)=-\frac{\partial\,\boldsymbol{\nabla}\times\mathbf{A}}{\partial t}-\boldsymbol{\nabla}\times\boldsymbol{\nabla}\boldsymbol{\phi}=-\frac{\partial\mathbf{B}}{\partial t}. \tag{5.1.8}$$

Gauss's law becomes

$$\boldsymbol{\nabla}\cdot\mathbf{E}=-\frac{\partial(\boldsymbol{\nabla}\cdot\mathbf{A})}{\partial t}-\nabla^2\phi=\frac{\rho}{\epsilon_0}. \tag{5.1.9}$$

This is a non-trivial equation that the potentials $\mathbf{A}$ and ϕ have to satisfy. It is essentially Poisson's equation (5.1.3) with an additional term.

Finally, Ampère's law becomes

$$\boldsymbol{\nabla}\times\mathbf{B}=\boldsymbol{\nabla}\times(\boldsymbol{\nabla}\times\mathbf{A})=\mu_0\mathbf{J}+\mu_0\epsilon_0\frac{\partial\mathbf{E}}{\partial t}=\mu_0\mathbf{J}-\mu_0\epsilon_0\boldsymbol{\nabla}\frac{\partial\phi}{\partial t}-\mu_0\epsilon_0\frac{\partial^2\mathbf{A}}{\partial t^2}. \tag{5.1.10}$$

Using $\mu_0\epsilon_0=1/c^2$, and rearranging the terms, we obtain

$$\frac{1}{c^2}\frac{\partial^2\mathbf{A}}{\partial t^2}=-\boldsymbol{\nabla}\times(\boldsymbol{\nabla}\times\mathbf{A})-\frac{1}{c^2}\boldsymbol{\nabla}\frac{\partial\phi}{\partial t}+\mu_0\mathbf{J}=\nabla^2\mathbf{A}-\boldsymbol{\nabla}\left(\boldsymbol{\nabla}\cdot\mathbf{A}+\frac{1}{c^2}\frac{\partial\phi}{\partial t}\right)+\mu_0\mathbf{J}. \tag{5.1.11}$$

This is basically the equation of motion for the vector potential $\mathbf{A}$.

Eq. (5.1.11) has the form of a wave equation with some additional terms, so we can try to look for plane wave solutions. As an Ansatz, let us consider a plane wave polarised in the x direction and travelling in the z direction at the speed of light

$$\begin{aligned}\mathbf{A}&=A_0e^{ik(z-ct)}\hat{\mathbf{x}},\\ \phi&=0.\end{aligned} \tag{5.1.12}$$

Substituting this to Eq. (5.1.11), we find

$$\frac{1}{c^2}\frac{\partial^2\mathbf{A}}{\partial t^2}-\nabla^2\mathbf{A}-\boldsymbol{\nabla}\left(\boldsymbol{\nabla}\cdot\mathbf{A}-\frac{1}{c^2}\frac{\partial\phi}{\partial t}\right)-\mu_0\mathbf{J}=\frac{1}{c^2}\frac{\partial^2\mathbf{A}}{\partial t^2}-\frac{\partial^2\mathbf{A}}{\partial z^2}=-k^2\mathbf{A}+k^2\mathbf{A}=0, \tag{5.1.13}$$

so this satisfies Ampère's law. Using Eq. (5.1.6), we find the magnetic and electric fields

$$\begin{aligned}\mathbf{B}&=\boldsymbol{\nabla}\times\mathbf{A}=\begin{vmatrix}\hat{\mathbf{x}}&\hat{\mathbf{y}}&\hat{\mathbf{z}}\\ \frac{\partial}{\partial x}&\frac{\partial}{\partial y}&\frac{\partial}{\partial z}\\ A_x&0&0\end{vmatrix}=\frac{\partial A_x}{\partial z}\hat{\mathbf{y}}=ikA_0e^{ik(z-ct)}\hat{\mathbf{y}},\\ \mathbf{E}&=-\frac{\partial\mathbf{A}}{\partial t}-\boldsymbol{\nabla}\phi=ikcA_0e^{ik(z-ct)}\hat{\mathbf{x}}.\end{aligned} \tag{5.1.14}$$

This is simply an electromagnetic wave travelling in the z direction.

[1] In fact, it is still possible to write down a vector potential that describes magnetic charge (see Problem Sheet 7 or Contemporary Physics 53 (2012) 195).

5.2 Gauge Invariance

The scalar and vector potential are not physically observable fields, only the electric and magnetic fields are. You know already that adding a constant to the scalar potential ϕ doesn't change the resulting electric field. More generally, we can ask whether we have more freedom to modify ϕ and $\mathbf{A}$ without changing the electric and magnetic fields.

To do this, consider adding functions of space and time to $\mathbf{A}$ and ϕ,

$$\begin{aligned} \mathbf{A} &\rightarrow \mathbf{A} + \boldsymbol{\alpha}(\mathbf{x}, t), \\ \phi &\rightarrow \phi + f(\mathbf{x}, t). \end{aligned} \tag{5.2.1}$$

This would change the electric and magnetic fields by

$$\begin{aligned} \mathbf{B} &\rightarrow \boldsymbol{\nabla} \times (\mathbf{A} + \boldsymbol{\alpha}) = \boldsymbol{\nabla} \times \mathbf{A} + \boldsymbol{\nabla} \times \boldsymbol{\alpha} = \mathbf{B} + \boldsymbol{\nabla} \times \boldsymbol{\alpha}, \\ \mathbf{E} &\rightarrow -\frac{\partial(\mathbf{A} + \boldsymbol{\alpha})}{\partial t} - \boldsymbol{\nabla}(\phi + f) = \mathbf{E} - \frac{\partial \boldsymbol{\alpha}}{\partial t} - \boldsymbol{\nabla} f. \end{aligned} \tag{5.2.2}$$

Therefore, $\mathbf{B}$ and $\mathbf{E}$ remain unchanged if

$$\boldsymbol{\nabla} \times \boldsymbol{\alpha} = 0, \quad \text{and} \quad \frac{\partial \boldsymbol{\alpha}}{\partial t} + \boldsymbol{\nabla} f = 0. \tag{5.2.3}$$

The Helmholtz theorem states that any vector field whose curl vanishes can be written as a gradient of a scalar, so we can write

$$\boldsymbol{\alpha} = \boldsymbol{\nabla} \lambda. \tag{5.2.4}$$

The second condition in Eq. (5.2.3) then becomes

$$\boldsymbol{\nabla} f = -\frac{\partial \boldsymbol{\alpha}}{\partial t} = -\boldsymbol{\nabla} \frac{\partial \lambda}{\partial t}. \tag{5.2.5}$$

This means that the physical fields $\mathbf{B}$ and $\mathbf{E}$ are invariant under *gauge transformations*

$$\begin{aligned} \mathbf{A} &\rightarrow \mathbf{A} + \boldsymbol{\nabla} \lambda, \\ \phi &\rightarrow \phi - \frac{\partial \lambda}{\partial t}, \end{aligned} \tag{5.2.6}$$

where $\lambda(\mathbf{x}, t)$ is an arbitrary scalar function. This symmetry, which is known as *gauge invariance* plays a very important role in particle physics, where an analogous gauge invariance determines the properties of elementary particle interactions almost completely.

Note that while $\mathbf{E}$ and $\mathbf{B}$ are invariant under gauge transformations, Eqs. (5.1.9) and (5.1.11) are not. This means that we can use a gauge transformation to make those equations simpler and easier to solve. This is known as fixing the gauge. For example, the divergence $\boldsymbol{\nabla} \cdot \mathbf{A}$ is not gauge invariant but transforms as

$$\boldsymbol{\nabla} \cdot \mathbf{A} \rightarrow \boldsymbol{\nabla} \cdot (\mathbf{A} + \boldsymbol{\nabla} \lambda) = \boldsymbol{\nabla} \cdot \mathbf{A} + \nabla^2 \lambda. \tag{5.2.7}$$

Because we can always find a solution to $\nabla^2 \lambda = g$ for an arbitrary function $g(\mathbf{x}, t)$, we can use a gauge transformation to fix $\boldsymbol{\nabla} \cdot \mathbf{A}$ to any value we like.

One popular way to fix the gauge is the *Coulomb gauge*, in which $\boldsymbol{\nabla} \cdot \mathbf{A} = 0$. In this gauge the non-trivial Maxwell equations (5.1.9) and (5.1.11) become

$$\begin{aligned} \nabla^2 \phi &= -\frac{\rho}{\epsilon_0}, \\ \frac{1}{c^2} \frac{\partial^2 \mathbf{A}}{\partial t^2} &= \nabla^2 \mathbf{A} - \frac{1}{c^2} \nabla \frac{\partial \phi}{\partial t} + \mu_0 \mathbf{J}. \end{aligned} \tag{5.2.8}$$

The main benefit of this gauge is that the equations are simpler: The first equation is simply the familiar Poisson equation, and the second equation is a wave equation with a source term. However, the drawback is that the Poisson equation appears to violate causality because a change is the charge distribution affects the scalar potential immediately at all distances. This is not a serious problem because ϕ is not observable, and the observable fields $\mathbf{E}$ and $\mathbf{B}$ still behave causally.

From the point of view of relativity, a better choice is the *Lorenz gauge*[2] defined by

$$\boldsymbol{\nabla} \cdot \mathbf{A} + \frac{1}{c^2}\frac{\partial \phi}{\partial t} = 0. \tag{5.2.9}$$

In this gauge, the equations of motion are

$$\begin{aligned} \frac{1}{c^2}\frac{\partial^2 \phi}{\partial t^2} - \nabla^2 \phi &= \frac{\rho}{\epsilon_0}, \\ \frac{1}{c^2}\frac{\partial^2 \mathbf{A}}{\partial t^2} - \nabla^2 \mathbf{A} &= \mu_0 \mathbf{J}. \end{aligned} \tag{5.2.10}$$

Now both ϕ and $\mathbf{A}$ satisfy wave equations, and therefore changes in the charge distribution propagate at the speed of light, satisfying causality.

A third gauge choice which is often useful is the *Weyl gauge*, which is also known as the *temporal gauge*. It is defined as $\phi = 0$, which means that the only degree of freedom is $\mathbf{A}$. In this gauge, the Maxwell equations become

$$\begin{aligned} \frac{\partial}{\partial t}\left(\boldsymbol{\nabla} \cdot \mathbf{A}\right) &= -\frac{\rho}{\epsilon_0}, \\ \frac{1}{c^2}\frac{\partial^2 \mathbf{A}}{\partial t^2} &= \nabla^2 \mathbf{A} - \boldsymbol{\nabla}\left(\boldsymbol{\nabla} \cdot \mathbf{A}\right) + \mu_0 \mathbf{J}. \end{aligned} \tag{5.2.11}$$

5.3 Particle in an electromagnetic Field

Let us see how to incorporate the electromagnetic field into the Lagrangian and Hamiltonian methods. For the static electric field, this is straightforward, because one can desribe it by a standard potential term

$$V(\mathbf{r}) = q\phi(\mathbf{r}). \tag{5.3.1}$$

The Lagrangian would therefore be

$$L = T - V = \frac{1}{2}m\dot{\mathbf{r}}^2 - q\phi(\mathbf{r}). \tag{5.3.2}$$

However, this is clearly not invariant under gauge transformations (5.2.6). A gauge transformation $\lambda(t, \mathbf{r})$ would change the Lagrangian into

$$L \to L' = \frac{1}{2}m\dot{\mathbf{r}}^2 - q\phi(\mathbf{r}) + q\frac{\partial \lambda}{\partial t} = L + q\frac{\partial \lambda}{\partial t}. \tag{5.3.3}$$

Adding such an extra term to the Lagrangian will change the physics, unless the extra term is a total time derivative, in which case it would only change the action by a constant. Our term $q\frac{\partial \lambda}{\partial t}$ is not a total time derivative, but could we modify the Lagrangian in such a way that it turns into

$$\frac{d\lambda}{dt} = \frac{\partial \lambda}{\partial t} + \frac{dx}{dt}\frac{\partial \lambda}{\partial x} + \frac{dy}{dt}\frac{\partial \lambda}{\partial y} + \frac{dz}{dt}\frac{\partial \lambda}{\partial z} = \frac{\partial \lambda}{\partial t} + \dot{\mathbf{r}} \cdot \boldsymbol{\nabla}\lambda? \tag{5.3.4}$$

[2]Even though the Lorenz gauge is invariant under Lorentz transformations, they are spelled differently. The Lorenz gauge (with no "t") is named after Ludvig Lorenz, whereas Lorentz transformations (with "t") are named after Hendrik Lorentz.

Comparing with the form of the gauge transformation (5.2.6), we note that the vector potential changes by precisely $\boldsymbol{\nabla}\lambda$, which suggests the Lagrangian

$$L = \frac{1}{2}m\dot{\mathbf{r}}^2 + q\dot{\mathbf{r}}\cdot\mathbf{A} - q\phi. \tag{5.3.5}$$

Indeed, this Lagrangian changes under the gauge transformation (5.2.6) as

$$L \to L' = \frac{1}{2}m\dot{\mathbf{r}}^2 + q\dot{\mathbf{r}}\cdot\mathbf{A} + q\dot{\mathbf{r}}\cdot\boldsymbol{\nabla}\lambda - q\phi(\mathbf{r}) + q\frac{\partial\lambda}{\partial t} = L + q\frac{d\lambda}{dt}. \tag{5.3.6}$$

Now the Lagrangian (5.3.5) changes only by a total derivative, which does not affect the physics.

We can now derive the Euler-Lagrangian equations from our Lagrangian (5.3.5). We do this for the x coordinate, but it is trivial to generalise the result to the y and z coordinates. The Euler-Lagrange equation is

$$\begin{aligned}
\frac{d}{dt}\frac{\partial L}{\partial \dot{x}} - \frac{\partial L}{\partial x} &= \frac{d}{dt}\left(m\dot{x} + qA_x\right) - q\dot{\mathbf{r}}\cdot\frac{\partial\mathbf{A}}{\partial x} + q\frac{\partial\phi}{\partial x} \\
&= m\ddot{x} + q\frac{\partial A_x}{\partial t} + q\left(\dot{x}\frac{\partial A_x}{\partial x} + \dot{y}\frac{\partial A_x}{\partial y} + \dot{z}\frac{\partial A_x}{\partial z}\right) \\
&\quad -q\left(\dot{x}\frac{\partial A_x}{\partial x} + \dot{y}\frac{\partial A_y}{\partial x} + \dot{z}\frac{\partial A_z}{\partial x}\right) + q\frac{\partial\phi}{\partial x} \\
&= m\ddot{x} + q\left[\dot{y}\left(\frac{\partial A_x}{\partial y} - \frac{\partial A_y}{\partial x}\right) + \dot{z}\left(\frac{\partial A_x}{\partial z} - \frac{\partial A_z}{\partial x}\right) + q\frac{\partial\phi}{\partial x} + \frac{\partial A_x}{\partial t}\right] \\
&= m\ddot{x} - q\left(\dot{y}B_z - \dot{z}B_y\right) - qE_x = 0.
\end{aligned} \tag{5.3.7}$$

Combining this with the y and z components and writing in the vector notation we have

$$m\ddot{\mathbf{r}} = q\left(\mathbf{E} + \mathbf{v}\times\mathbf{B}\right), \tag{5.3.8}$$

which is just the usual Lorentz force equation. Note that we were able to derive this by assuming only the static Coulomb force and gauge invariance. This gives a flavour of how gauge invariance can play a key role in determining the properties of elementary particle interactions.

From the Lagrangian (5.3.5) we can derive the canonical momentum

$$p_x = \frac{\partial L}{\partial \dot{x}} = m\dot{x} + qA_x, \tag{5.3.9}$$

or in vector form $\mathbf{p} = m\dot{\mathbf{r}} + q\mathbf{A}$, and the Hamiltonian

$$H = \mathbf{p}\cdot\dot{\mathbf{r}} - L = \frac{(\mathbf{p} - q\mathbf{A})^2}{2m} + q\phi. \tag{5.3.10}$$

Eqs. (5.3.5) and (5.3.10) show that one has to formulate electrodynamics in terms of the scalar and vector potentials in order to use Lagrangian or Hamiltonian mechanics. In contrast, the Newtonian formulation only deals with the physical electric and magnetic fields.

5.4 Retarded Potentials

Let us now adopt the Lorenz gauge (5.2.9), and ask what the potentials ϕ and $\mathbf{A}$ in the presence of a time-dependent charge distribution. We focus on the scalar potential, but because the equation of the vector potential has the same form, the results can be translated directly to it.

Instead of actually solving Eqs. (5.2.10), I will present the solution and then show that it satisfies Eqs. (5.2.10). Let us start from the simplest case of a static point charge q at $\mathbf{r}'$, for which the scalar potential is

$$\phi(\mathbf{r}) = \frac{q}{4\pi\epsilon_0|\mathbf{r}-\mathbf{r}'|}. \tag{5.4.1}$$

It is easy to generalise this to a continuous but static charge distribution $\rho(\mathbf{r})$,

$$\phi(\mathbf{r}) = \frac{1}{4\pi\epsilon_0}\int d^3\mathbf{r}'\frac{\rho(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}. \tag{5.4.2}$$

Because electromagnetic waves travel at the speed of light c, one can guess that if the charge density changes with time, the potential at point $\mathbf{r}$ and time t will depend on the charge at point $\mathbf{r}'$ at the earlier time known as the *retarded time*

$$t' = t - \frac{1}{c}|\mathbf{r}-\mathbf{r}'|. \tag{5.4.3}$$

This suggests that the potential should be given by

$$\phi(\mathbf{r},t) = \frac{1}{4\pi\epsilon_0}\int d^3\mathbf{r}'\frac{\rho(\mathbf{r}',t-|\mathbf{r}-\mathbf{r}'|/c)}{|\mathbf{r}-\mathbf{r}'|}\,. \tag{5.4.4}$$

This is known as the *retarded potential*.

To check that Eq. (5.4.4) is valid, we have to show that it satisfies Eq. (5.2.10). To simplify the notation we write $\mathbf{R} = \mathbf{r}-\mathbf{r}'$ and $R = |\mathbf{r}-\mathbf{r}'|$, so that the retarded time is $t' = t - R/c$. Now, we have the Laplacian with respect to the coordinate $\mathbf{r}$,

$$\boldsymbol{\nabla}^2\phi = \frac{1}{4\pi\epsilon_0}\int d^3\mathbf{r}'\boldsymbol{\nabla}^2\frac{\rho}{R} = \frac{1}{4\pi\epsilon_0}\int d^3\mathbf{r}'\left[\frac{1}{R}\boldsymbol{\nabla}^2\rho + 2\left(\boldsymbol{\nabla}\rho\right)\cdot\left(\boldsymbol{\nabla}\frac{1}{R}\right) + \rho\boldsymbol{\nabla}^2\frac{1}{R}\right]. \tag{5.4.5}$$

The charge density $\rho(t',\mathbf{r}')$ depends on $\mathbf{r}$ only through the retarded time t' given by Eq. (5.4.3), so

$$\boldsymbol{\nabla}\rho = \dot{\rho}\boldsymbol{\nabla}t' = -\frac{1}{c}\dot{\rho}\boldsymbol{\nabla}R. \tag{5.4.6}$$

Taking the divergence of this, we obtain

$$\boldsymbol{\nabla}^2\rho = -\frac{1}{c}\left[\boldsymbol{\nabla}\dot{\rho}\cdot\boldsymbol{\nabla}R + \dot{\rho}\boldsymbol{\nabla}^2R\right] = \frac{1}{c^2}\ddot{\rho}\boldsymbol{\nabla}R\cdot\boldsymbol{\nabla}R - \frac{1}{c}\dot{\rho}\boldsymbol{\nabla}^2R. \tag{5.4.7}$$

Using the identities from vector calculus,

$$\boldsymbol{\nabla}R = \frac{\mathbf{R}}{R}, \quad \boldsymbol{\nabla}\frac{1}{R} = -\frac{\mathbf{R}}{R^3}, \tag{5.4.8}$$

and

$$\boldsymbol{\nabla}^2R = \frac{2}{R}, \quad \boldsymbol{\nabla}^2\frac{1}{R} = -4\pi\delta(\mathbf{R}), \tag{5.4.9}$$

we can write these as

$$\boldsymbol{\nabla}\rho = -\frac{1}{c}\dot{\rho}\frac{\mathbf{R}}{R}, \tag{5.4.10}$$

and

$$\boldsymbol{\nabla}^2\rho = \frac{1}{c^2}\ddot{\rho} - \frac{2}{c}\frac{\dot{\rho}}{R}, \tag{5.4.11}$$

and Eq. (5.4.5) as

$$\begin{aligned}\boldsymbol{\nabla}^2\phi &= \frac{1}{4\pi\epsilon_0}\int d^3\mathbf{r}'\left[\frac{1}{c^2}\frac{\ddot{\rho}}{R}-\frac{2}{c}\frac{\dot{\rho}}{R^2}-\frac{2}{c}\dot{\rho}\frac{\mathbf{R}}{R}\cdot\left(-\frac{\mathbf{R}}{R^3}\right)-4\pi\rho\delta(\mathbf{R})\right]\\ &= \frac{1}{4\pi\epsilon_0}\int d^3\mathbf{r}'\frac{1}{c^2}\frac{\ddot{\rho}}{R}-\frac{\rho}{\epsilon_0}=\frac{1}{c^2}\frac{\partial^2\phi}{\partial t^2}-\frac{\rho}{\epsilon_0}.\end{aligned} \tag{5.4.12}$$

This shows that the retarded potential satisfies Eq. (5.2.10).

Because the equation for $\mathbf{A}$ has the same form, we can also write down the corresponding solution for the vector potential $\mathbf{A}$

$$\mathbf{A}(\mathbf{r},t)=\frac{\mu_0}{4\pi}\int d^3\mathbf{r}'\frac{\mathbf{J}(\mathbf{r}',t-|\mathbf{r}-\mathbf{r}'|/c)}{|\mathbf{r}-\mathbf{r}'|}\,. \tag{5.4.13}$$

The solutions (5.4.4 and 5.4.13) describe electromagnetic waves emitted by a changing charge and current distribution travelling outwards at the speed of light. Note that they are not the most general solutions of Eq. (5.2.10). There is another solution, known as the *advanced potential*, which describes electromagnetic waves travelling from infinity towards the source. This is, however, not useful for most practical applications.

As the first example of retarded potentials, consider an infinite straight wire on the z axis. We switch on a current at time $t=0$, so that the current is given by

$$I(t)=\begin{cases}0 & \text{for } t\le 0,\\ I_0 & \text{for } t>0.\end{cases}$$

Because the current is confined on the z axis, the current density is

$$\mathbf{J}=I(t)\delta(x)\delta(y)\hat{\mathbf{z}}. \tag{5.4.14}$$

There is no electric charge, so $\rho=0$, and therefore $\phi=0$. The vector potential is, however, non-trivial. To simplify the calculation, we calculate it at point $\mathbf{r}=(x,y,0)$. Because the setup is translation invariant in the z direction, the same result applies at all z. Substituing Eq. (5.4.14) into Eq. (5.4.13), we find

$$\begin{aligned}\mathbf{A}(t,\mathbf{r}) &= \frac{\mu_0\hat{\mathbf{z}}}{4\pi}\int d^3\mathbf{r}'\frac{I(t-|\mathbf{r}-\mathbf{r}'|/c)\delta(x')\delta(y')}{|\mathbf{r}-\mathbf{r}'|},\\ &= \frac{\mu_0\hat{\mathbf{z}}}{4\pi}\int dz'\frac{I(t-\sqrt{r^2+z'^2}/c)}{\sqrt{r^2+z'^2}}.\end{aligned} \tag{5.4.15}$$

We now note that the current (5.4.14) vanishes for $z'^2>c^2t^2-r^2$ and is I_0 elsewhere, so we have

$$\begin{aligned}\mathbf{A}(t,\mathbf{r}) &= \frac{\mu_0 I_0\hat{\mathbf{z}}}{4\pi}\int_{-\sqrt{c^2t^2-r^2}}^{\sqrt{c^2t^2-r^2}}dz'\frac{1}{\sqrt{r^2+z'^2}}=\frac{\mu_0 I_0\hat{\mathbf{z}}}{2\pi}\int_0^{\sqrt{c^2t^2-r^2}}dz'\frac{1}{\sqrt{r^2+z'^2}}\\ &= \frac{\mu_0 I_0}{2\pi}\ln\frac{ct+\sqrt{c^2t^2-r^2}}{r}\hat{\mathbf{z}}.\end{aligned} \tag{5.4.16}$$

From this we can calculate the electric and magnetic fields,

$$\begin{aligned}\mathbf{E} &= -\frac{\partial\mathbf{A}}{\partial t}=-\frac{\mu_0 I_0}{2\pi}\frac{c}{\sqrt{c^2t^2-r^2}}\hat{\mathbf{z}},\\ \mathbf{B} &= \boldsymbol{\nabla}\times\mathbf{A}=-\frac{\partial A_z}{\partial r}\hat{\boldsymbol{\phi}}=\frac{\mu_0 I_0}{2\pi r}\frac{ct}{\sqrt{c^2t^2-r^2}}\hat{\boldsymbol{\phi}}.\end{aligned} \tag{5.4.17}$$

At late times these approach

$$\begin{aligned}\mathbf{E} &\to 0\\ \mathbf{B} &\to \frac{\mu_0 I_0}{2\pi r}\hat{\boldsymbol{\phi}},\end{aligned} \tag{5.4.18}$$

which are the familiar results for a time-independent current.

The second example we consider is the *Hertzian dipole*, which is an infinitesimally small oscillating electric dipole. The dipole consists of charge $+q$ at a position $\delta\mathbf{r} = \delta z\hat{\mathbf{z}}$ and an opposite charge $-q$ at $-\delta\mathbf{r}$. It can be parameterised by the dipole moment $\mathbf{p} = 2q\delta z\hat{\mathbf{z}}$. The dipole is oscillating such that $q(t) = q_0\cos(\omega t)$ so that the dipole moment may be written $\mathbf{p}(t) = p_0\cos(\omega t)\hat{\mathbf{z}}$, where $p_0 = 2q_0\delta z$.

The scalar potential ϕ at point $\mathbf{r}$ in spherical coordinates can be obtained from Eq. (5.4.4) as

$$\phi = \frac{1}{4\pi\epsilon_0}\left\{\frac{q_0\cos\omega(t-|\mathbf{r}-\delta\mathbf{r}|/c)}{|\mathbf{r}-\delta\mathbf{r}|} - \frac{q_0\cos\omega(t-|\mathbf{r}+\delta\mathbf{r}|/c)}{|\mathbf{r}+\delta\mathbf{r}|}\right\}. \tag{5.4.19}$$

Because we are assuming that $\delta z = |\delta\mathbf{r}|$ is small, we can simplify this by Taylor expanding it to linear order in δz. First, because $\delta z \ll r$, we expand

$$\begin{aligned}|\mathbf{r}\pm\delta\mathbf{r}| &= \sqrt{(\mathbf{r}\pm\delta\mathbf{r})^2} = \sqrt{\mathbf{r}^2 \pm 2\mathbf{r}\cdot\delta\mathbf{r} + \delta\mathbf{r}^2} = \sqrt{r^2 \pm 2r\delta z\cos\theta + \delta z^2}\\ &= = r \pm \delta z\cos\theta + O(\delta z^2),\end{aligned} \tag{5.4.20}$$

and

$$\frac{1}{|\mathbf{r}\pm\delta\mathbf{r}|} = \frac{1}{r} \mp \frac{\delta z\cos\theta}{r^2} + O(\delta z^2). \tag{5.4.21}$$

These give us

$$\begin{aligned}\phi &= \frac{q_0}{4\pi\epsilon_0}\left\{\frac{\cos\omega\left(t-\frac{r-\delta z\cos\theta}{c}\right) - \cos\omega\left(t-\frac{r+\delta z\cos\theta}{c}\right)}{r}\right.\\ &\qquad \left. +\delta z\cos\theta\frac{\cos\omega\left(t-\frac{r-\delta z\cos\theta}{c}\right) + \cos\omega\left(t-\frac{r+\delta z\cos\theta}{c}\right)}{r^2}\right\}.\end{aligned} \tag{5.4.22}$$

The assumption $\delta z \ll c/\omega$ allows us to Taylor expand the cosines. In the first term, the constant term in the Taylor expansion cancels, and therefore we only obtain a linear term in δz, whereas the second term already contains a factor δz and therefore we only keep the constant term in the Taylor expansion,

$$\begin{aligned}\phi &= \frac{q_0}{4\pi\epsilon_0}\left\{\frac{-2\omega\frac{\delta z\cos\theta}{c}\sin\omega\left(t-\frac{r}{c}\right)}{r} + \delta z\cos\theta\frac{2\cos\omega\left(t-\frac{r}{c}\right)}{r^2}\right\}\\ &= \frac{p_0\cos\theta}{4\pi\epsilon_0}\left\{\frac{1}{r^2}\cos\omega(t-r/c) - \frac{\omega}{rc}\sin\omega(t-r/c)\right\}.\end{aligned} \tag{5.4.23}$$

We are only really interested in what happens far away from the dipole, in the region where $r \gg \lambda$ or $r \gg c/\omega$. In that case we may ignore the term in r^{-2} in Eq. (5.4.23) and write

$$\phi = -\frac{p_0\omega}{4\pi\epsilon_0 c}\left(\frac{\cos\theta}{r}\right)\sin\omega(t-r/c). \tag{5.4.24}$$

When the dipole oscillates, there is a current

$$I(t) = \frac{dq}{dt} = -q_0\omega\sin\omega t \tag{5.4.25}$$

along the line element between the two charges. The current density is

$$\mathbf{J}(t,\mathbf{r}) = \begin{cases} I(t)\delta(x)\delta(y)\hat{\mathbf{z}} & \text{if } -\delta z < z < \delta z, \\ 0 & \text{otherwise.} \end{cases} \tag{5.4.26}$$

Using Eq. (5.4.13) we find the vector potential

$$\begin{aligned} \mathbf{A}(t,\mathbf{r}) &= \frac{\mu_0\hat{\mathbf{z}}}{4\pi}\int_{-\delta z}^{\delta z}\frac{I(t-|\mathbf{r}-z'\hat{\mathbf{z}}|/c)}{|\mathbf{r}-z'\hat{\mathbf{z}}|} \approx -\frac{2\mu_0\delta z\hat{\mathbf{z}}}{4\pi}\omega\frac{\sin\omega(t-r/c)}{r} \\ &= -\frac{\mu_0 p_0}{4\pi}\omega\frac{\sin\omega(t-r/c)}{r}\hat{\mathbf{z}}\,. \end{aligned} \tag{5.4.27}$$

From Eqs. (5.4.24) and (5.4.27), we can calculate the magnetic and electric fields. In spherical coordinates, we find

$$\begin{aligned} \mathbf{B} &= \boldsymbol{\nabla}\times\mathbf{A} = \left(\frac{1}{r}\frac{\partial}{\partial r}(rA_\theta) - \frac{1}{r}\frac{\partial A_r}{\partial\theta}\right)\hat{\boldsymbol{\phi}} \\ &= -\frac{\mu_0 p_0\omega^2}{4\pi c}\left(\frac{\sin\theta}{r}\right)\cos\omega(t-r/c)\hat{\boldsymbol{\phi}} \end{aligned} \tag{5.4.28}$$

$$\begin{aligned} \mathbf{E} &= -\frac{\partial\mathbf{A}}{\partial t} - \boldsymbol{\nabla}\phi \\ &= -\frac{\mu_0 p_0\omega^2}{4\pi}\left(\frac{\sin\theta}{r}\right)\cos\omega(t-r/c)\hat{\boldsymbol{\theta}}\,, \end{aligned} \tag{5.4.29}$$

where we have ignored all terms that fall faster than $1/r$ and used the fact that $\hat{\mathbf{z}} = \cos\theta\hat{\mathbf{r}} - \sin\theta\hat{\boldsymbol{\theta}}$, which implies $A_r = A_z\cos\theta$ and $A_\theta = -A_z\sin\theta$.

Eq. (5.4.29) shows electromagnetic waves travelling radially away from the dipole. To find the power radiated by the dipole, we calculate the Poynting vector $\mathbf{S}$ which given the energy flow through unit area per unit time. We find

$$\begin{aligned} \mathbf{S} &= \frac{1}{\mu_0}\mathbf{E}\times\mathbf{B} \\ &= \frac{\mu_0}{c}\left(\frac{p_0\omega^2}{4\pi}\frac{\sin\theta}{r}\cos\omega(t-r/c)\right)^2\hat{\mathbf{r}}. \end{aligned} \tag{5.4.30}$$

This is oscillating, but we can average over one cycle of oscillation by noting that $\langle\cos^2\omega(t-r/c)\rangle = 1/2$,

$$\langle\mathbf{S}\rangle = \frac{\mu_0 p_0^2\omega^4}{32\pi^2 c}\frac{\sin^2\theta}{r^2}\hat{\mathbf{r}}. \tag{5.4.31}$$

The total power radiated is given by integrating Eq. (5.4.31) over the surface of a sphere

$$\langle P\rangle = \int d\mathbf{a}\cdot\langle\mathbf{S}\rangle = \frac{\mu_0 p_0^2\omega^4}{32\pi^2 c}\int\frac{\sin^2\theta}{r^2}r^2\sin\theta\,\mathrm{d}\theta\,\mathrm{d}\phi = \frac{\mu_0 p_0^2\omega^4}{12\pi c} \tag{5.4.32}$$

which is independent of r as expected from conservation of energy.

5.5 Fields of a Moving Charge

Let us now consider a single moving charge, which follows some trajectory $\mathbf{r}_0(t)$. The charge and current densities are

$$\begin{aligned} \rho(t,\mathbf{r}) &= q\delta^{(3)}(\mathbf{r}-\mathbf{r}_0(t)), \\ \mathbf{J}(t,\mathbf{r}) &= q\mathbf{v}_0(t)\delta^{(3)}(\mathbf{r}-\mathbf{r}_0(t)), \end{aligned} \tag{5.5.1}$$

where $\mathbf{v}_0 = \dot{\mathbf{r}}_0$.

To calculate the scalar potential from Eq. (5.4.4), we use a simple trick: We introduce a delta function that enforces the condition (5.4.3),

$$\begin{aligned} \phi(t,\mathbf{r}) &= \frac{1}{4\pi\epsilon_0}\int dt'\, d^3\mathbf{r}'\, \delta\left(t' - [t - |\mathbf{r}-\mathbf{r}'|/c]\right)\frac{\rho(t',\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} \\ &= \frac{q}{4\pi\epsilon_0}\int dt'\, d^3\mathbf{r}'\, \delta\left(t' - [t - |\mathbf{r}-\mathbf{r}'|/c]\right)\frac{\delta(\mathbf{r}'-\mathbf{r}_0(t'))}{|\mathbf{r}-\mathbf{r}'|} \\ &= \frac{q}{4\pi\epsilon_0}\int dt'\, \delta\left(t' - [t - |\mathbf{r}-\mathbf{r}_0(t')|/c]\right)\frac{1}{|\mathbf{r}-\mathbf{r}_0(t')|}. \end{aligned} \tag{5.5.2}$$

We now have a delta function whose argument is a non-linear function of the integration variable t'. To deal with it, we use the general result that if function $f(x)$ has a zero at $x = x_0$, i.e., $f(x_0) = 0$, then

$$\delta(f(x)) = \frac{\delta(x-x_0)}{|f'(x_0)|}. \tag{5.5.3}$$

In Eq. (5.5.2), the function is

$$f(t') = t' - t + \frac{1}{c}|\mathbf{r}-\mathbf{r}_0(t')|. \tag{5.5.4}$$

Denoting the retarded time by $t_{\rm ret}$, so that $f(t_{\rm ret}) = 0$ or equivalently

$$|\mathbf{r}-\mathbf{r}_0(t_{\rm ret})| = c(t-t_{\rm ret}), \tag{5.5.5}$$

we therefore have

$$\phi(t,\mathbf{r}) = \frac{q}{4\pi\epsilon_0}\int dt'\,\frac{\delta\left(t'-t_{\rm ret}\right)}{|f'(t_{\rm ret}||\mathbf{r}-\mathbf{r}_0(t')|} = \frac{q}{4\pi\epsilon_0}\frac{1}{|f'(t_{\rm ret})||\mathbf{r}-\mathbf{r}_0(t_{\rm ret})|}. \tag{5.5.6}$$

We need to calculate

$$f'(t_{\rm ret}) = \frac{\partial}{\partial t'}\left(t'-t+\frac{1}{c}|\mathbf{r}-\mathbf{r}_0(t')|\right)\Big|_{t'=t_{\rm ret}} = 1 - \frac{(\mathbf{r}-\mathbf{r}_0(t_{\rm ret}))\cdot\mathbf{v}_0(t_{\rm ret})}{c|\mathbf{r}-\mathbf{r}_0(t_{\rm ret})|}. \tag{5.5.7}$$

Denoting

$$\mathbf{R} = \mathbf{r}-\mathbf{r}_0(t_{\rm ret}), \quad \text{and} \quad \mathbf{v}_{\rm ret} = \mathbf{v}_0(t_{\rm ret}), \tag{5.5.8}$$

this is simply

$$f'(t_{\rm ret}) = 1 - \frac{\mathbf{R}\cdot\mathbf{v}_{\rm ret}}{cR}. \tag{5.5.9}$$

Substituting this to Eq. (5.5.6), we obtain the final result

$$\phi\left(\mathbf{r},t\right) = \frac{q}{4\pi\epsilon_0}\frac{1}{R-\mathbf{R}\cdot\mathbf{v}_{\rm ret}/c}. \tag{5.5.10}$$

Note that Eqs. (5.5.5) and (5.5.8) imply that $R = c(t - t_{\rm ret})$. Analogously, one finds the vector potential

$$\mathbf{A}\left(\mathbf{r},t\right) = \frac{\mu_0 q}{4\pi}\frac{\mathbf{v}_0(t_{\rm ret})}{R - \mathbf{R}\cdot\mathbf{v}_{\rm ret}/c}\,. \tag{5.5.11}$$

These forms for the potentials due to a moving charge are known as *Liénard–Wiechert potentials.*

From the potentials (5.5.10) and (5.5.11), we can calculate the electric and magnetic fields. (For details, see 10.3.2 in Griffiths). The result is

$$\mathbf{E}(\mathbf{r},t) = -\frac{\partial\mathbf{A}}{\partial t} - \boldsymbol{\nabla}\phi = \frac{q}{4\pi\epsilon_0}\frac{1}{\left(R - \mathbf{R}\cdot\mathbf{v}_{\rm ret}/c\right)^3}\left\{\frac{\mathbf{R}'}{\gamma_{\rm ret}^2} + \frac{1}{c^2}\mathbf{R}\times\left(\mathbf{R}'\times\dot{\mathbf{v}}_{\rm ret}\right)\right\} \tag{5.5.12}$$

$$\mathbf{B}(\mathbf{r},t) = \boldsymbol{\nabla}\times\mathbf{A} = \frac{\mathbf{R}\times\mathbf{E}}{cR}\,, \tag{5.5.13}$$

where $\mathbf{R}' = \mathbf{R} - \mathbf{v}_{\rm ret}R/c$ is the position where the charge would be at time t if it had continued with the velocity $\mathbf{v}_{\rm ret}$, and $\gamma_{\rm ret} = (1 - v_{\rm ret}^2/c^2)^{-1/2}$ is the *retarded* γ factor.

This is the general solution, valid right up to $v \sim c$, for an accelerating particle. The first term in Eq. (5.5.12) varies as R^{-2} and is independent of acceleration, whereas the second term varies as R^{-1} and depends on acceleration. The first term may be interpreted as a generalised Coulomb field whereas the second term describes electromagentic waves radiated by the moving charge.

As our first example, let us consider a charge moving at constant velocity $\mathbf{v}$, so that $\mathbf{r}_0(t) = \mathbf{v}t$. In that case, Eq. (5.5.12) simplifies to

$$\mathbf{E}(\mathbf{r},t) = \frac{q}{4\pi\epsilon_0}\frac{1}{\left(R - \mathbf{R}\cdot\mathbf{v}_{\rm ret}/c\right)^3}\frac{\mathbf{R}'}{\gamma^2}. \tag{5.5.14}$$

For constant velocity, $\mathbf{R}' = \mathbf{R} - R\mathbf{v}/c = \mathbf{r} - \mathbf{v}t$. To find the denominator, we calculate

$$\begin{aligned}\left(R - \frac{1}{c}\mathbf{R}\cdot\mathbf{v}\right)^2 &= R^2 - \frac{2}{c}R\mathbf{R}\cdot\mathbf{v} + \frac{1}{c^2}(\mathbf{R}\cdot\mathbf{v})^2 = \left(1 - \frac{v^2}{c^2}\right)\mathbf{R}'^2 + \frac{1}{c^2}(\mathbf{R}'\cdot\mathbf{v})^2 \\ &= R'^2\left(1 - \frac{v^2}{c^2} + \frac{v^2}{c^2}\cos^2\theta'\right) = R'^2\left(1 - \frac{v^2}{c^2}\sin^2\theta'\right),\end{aligned} \tag{5.5.15}$$

where θ' is the angle between $\mathbf{R}'$ and $\mathbf{v}$. This shows that

$$R - \frac{1}{c}\mathbf{R}\cdot\mathbf{v} = R'\sqrt{1 - \frac{v^2}{c^2}\sin^2\theta'}. \tag{5.5.16}$$

Substituting this to Eq. (5.5.14), we find the electric field of a charge moving at constant velocity,

$$\mathbf{E} = \frac{q}{4\pi\epsilon_0}\frac{\left(1 - \frac{v^2}{c^2}\right)\mathbf{R}'}{\left(1 - \frac{v^2}{c^2}\sin^2\theta'\right)^{3/2}R'^3}. \tag{5.5.17}$$

Note that the electric field points to the present position $\mathbf{R}'$ of the charge, not to the retarded position $\mathbf{R}$ as one might have expected. This makes sense from the point of view of special relativity, because the field of a moving charge should be just a Lorentz boost of the field of a stationary charge, which naturally points to the charge itself. We can also note that the field strength is reduced in forward and backward directions, where $\sin\theta' \approx 0$, and enhanced in perpendicular directions where $\sin\theta' \approx 1$.

We can also find the magnetic field using Eq. (5.5.13). Noting that $\mathbf{R} = \mathbf{R}' + \mathbf{v}(t - t_{\rm ret})$ and $\mathbf{E} \parallel \mathbf{R}'$, we find

$$\mathbf{B} = \frac{\mathbf{R} \times \mathbf{E}}{cR} = \frac{(t - t_{\rm ret})\mathbf{v} \times \mathbf{E}}{cR} = \frac{\mathbf{v} \times \mathbf{E}}{c^2}. \tag{5.5.18}$$

Considering the moving charge as a Lorentz boost of a stationary charge, it is interesting that it has a non-zero magnetic field. This shows that electric and magnetic fields have to transform to each other under Lorentz boosts.

As another example, consider a charge moving non-relativistically so that $v \ll c$. We can then simplify the expressions in (5.5.12) and (5.5.13) such that $\mathbf{R}' \to \mathbf{R}$ and $R - \mathbf{R} \cdot \mathbf{v}/c \to R$. The electric field becomes

$$\mathbf{E}\left(\mathbf{r}, t\right) = \frac{q}{4\pi\epsilon_0} \frac{1}{R^3} \left(\mathbf{R} + \frac{1}{c^2} \mathbf{R} \times \left(\mathbf{R} \times \dot{\mathbf{v}} \right) \right) . \tag{5.5.19}$$

The first term describes simply the Coulomb field moving with the charge, so if we are interested in the power radiated by the charge, we can ignore it and only consider the second term, which describes radiation. The Poynting vector due to the radiation term is

$$\mathbf{S} = \frac{1}{\mu_0} \mathbf{E}_{\rm rad} \times \mathbf{B} = \frac{1}{\mu_0 cR} \mathbf{E}_{\rm rad} \times \left(\mathbf{R} \times \mathbf{E}_{\rm rad} \right) = \frac{1}{\mu_0 cR} E_{\rm rad}^2 \mathbf{R} = \frac{q^2 \dot{v}^2 \sin^2\theta}{16\pi^2 \epsilon_0 c^3 R^2} \hat{\mathbf{R}}, \tag{5.5.20}$$

where θ is the angle between $\mathbf{R}$ and the acceleration $\dot{\mathbf{v}}$. The total power radiated into a sphere of radius R is

$$P = \int_0^\pi d\theta \, \sin\theta \int_0^{2\pi} d\phi R^2 S_R = \frac{q^2 \dot{v}^2}{6\pi\epsilon_0 c^3} \, , \tag{5.5.21}$$

where the $\dot{v}$ here is evaluated at the retarded time.

Chapter 6

Electrodynamics and Relativity

6.1 Four-Vectors

In the context of special and general relativity particularly, but also in other situations, it is often useful to consider space and time as two aspects of the same quantity rather than as separate. To this end we can write the coordinates of an *event* occurring at position $\mathbf{r} = (x, y, z)$ and time t as at the *four-vector* position (ct, x, y, z) in space-time, where the time component is written as a length ct, the distance light travels in time t, so that it has the same units as the other components. Often this is rewritten in the form (x^0, x^1, x^2, x^3), with superscript indices. The reason for this becomes clear soon. These superscript indices should not be confused with raising x to a power. They are usually denoted by Greek letters, e.g., x^μ, where $\mu \in \{0, 1, 2, 3\}$. Abusing this notation slightly, we often also denote the whole four-vector by x^μ to make it clear that we are referring to a four-vector quantity. If that is obvious, we can also refer to the four-vector by simply x.

Consider now a Lorentz boost in x direction by velocity v. Writing $\gamma = 1/\sqrt{1 - v^2/c^2}$, the time and space coordinates transform as

$$\begin{aligned} t' &= \gamma(t - vx/c^2), \\ x' &= \gamma(x - vt), \\ y' &= y, \\ z' &= z. \end{aligned} \tag{6.1.1}$$

Using the four-vector notation, this can be written as a matrix multiplication

$$\begin{pmatrix} x^0 \\ x^1 \\ x^2 \\ x^3 \end{pmatrix}' = \begin{pmatrix} \gamma & -\beta\gamma & 0 & 0 \\ -\beta\gamma & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x^0 \\ x^1 \\ x^2 \\ x^3 \end{pmatrix}, \tag{6.1.2}$$

where $\beta = v/c$. Denoting the transformation matrix by $\overline{\overline{\mathbf{\Lambda}}}$, we can write the transformation in terms of the four-vector components as

$$x'^\mu = \sum_\nu \Lambda^\mu{}_\nu x^\nu, \tag{6.1.3}$$

where $\Lambda^\mu{}_\nu$ denotes the elements of the matrix $\overline{\overline{\mathbf{\Lambda}}}$. As usual, the first index in $\Lambda^\mu{}_\nu$ refers to the row and the second index to the column. The reason why we write the row index as superscript and the column index as subscript will become clear soon.

A Lorentz transformation can be defined as one that leaves all space-time intervals

$$\ell^2 = c^2t^2 - \mathbf{r}^2 = (x^0)^2 - (x^1)^2 - (x^2)^2 - (x^3)^2 \tag{6.1.4}$$

unchanged. To express this in the four-vector notation, we define a 4×4 matrix known as the *metric tensor*,

$$\overline{\overline{\mathbf{g}}} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}, \tag{6.1.5}$$

whose components we denote by $g_{\mu\nu}$. Note that the metric is symmetric, $g_{\nu\mu} = g_{\mu\nu}$. More specifically, this is known as the *Minkowski metric* to distinguish it from more general metrics that are used in general relativity, and to emphasize that it is often also denoted by $\eta_{\mu\nu}$.

Using the metric tensor, the space-time interval can be written as

$$\ell^2 = \sum_{\mu,\nu} x^\mu g_{\mu\nu} x^\nu. \tag{6.1.6}$$

Note that in Eqs. (6.1.3) and (6.1.6) each Lorentz index appears once as a superscript and once as a subscript. From now on we will follow the *Einstein convention*, in which are Lorentz index that appears once as a superscript and once as a subscript is summed over. We will see later that this is almost always what we want, and in the exceptional cases when we do not want to sum over the index, we state that explicitly.

Using the Einstein convention, the transformation law (6.1.3) becomes

$$x'^\mu = \Lambda^\mu{}_\nu x^\nu, \tag{6.1.7}$$

and the expression for the space-time interval is

$$\ell^2 = x^\mu g_{\mu\nu} x^\nu. \tag{6.1.8}$$

Tensors (i.e. linear relations between a number of four-vectors) can be represented conveniently in this notation. For example, if four-vector x^μ is related to four-vectors y^μ, z^μ and w^μ through a linear relation (i.e., a rank 4 tensor), this can be expressed as

$$x^\mu = M^\mu{}_{\nu\rho\sigma} y^\nu z^\rho w^\sigma. \tag{6.1.9}$$

We can see that in this notation, a rank N tensor appears simply as an object with N Lorentz indices. In contrast, the matrix notation we used to represent the inertia tensor is only suitable for rank 2 tensor.

The component notation is also more flexible than the matrix notation. A product of two matrices (or rank 2 tensors) $\overline{\overline{\mathbf{A}}}$ and $\overline{\overline{\mathbf{B}}}$, with components $A^\mu{}_\nu$ and $B^\mu{}_\nu$, can be written as

$$(\overline{\overline{\mathbf{A}}} \cdot \overline{\overline{\mathbf{B}}})^\mu{}_\nu = A^\mu{}_\rho B^\rho{}_\nu. \tag{6.1.10}$$

In matrix multiplication, the order of the factors matters, $\overline{\overline{\mathbf{A}}} \cdot \overline{\overline{\mathbf{B}}} \neq \overline{\overline{\mathbf{B}}} \cdot \overline{\overline{\mathbf{A}}}$, but in the component notation the symbols represent matrix elements, which are real or complex numbers. Therefore we can change the order of the factors freely, i.e., $A^\mu{}_\rho B^\rho{}_\nu = B^\rho{}_\nu A^\mu{}_\rho$. The labelling of the indices keeps track of how the tensors are multiplied. To actually compute numerical values, it it often convenient to switch to the matrix notation, and it is then important to write the matrices in the right order. Note also that

you can choose freely which Greek letter you use for each summation index, but the same letter can only be used ones in one expression (i.e. once as a superscript and once as a subscript).

Under this transformation, the space-time interval transforms as

$$\ell^2 = x^\mu g_{\mu\nu} x^\nu \to x'^\mu g_{\mu\nu} x'^\nu = \Lambda^\mu{}_\rho x^\rho g_{\mu\nu} \Lambda^\nu{}_\sigma x^\sigma = x^\rho \Lambda^\mu{}_\rho g_{\mu\nu} \Lambda^\nu{}_\sigma x^\sigma = x^\mu \Lambda^\rho{}_\mu g_{\rho\sigma} \Lambda^\sigma{}_\nu x^\nu, \quad (6.1.11)$$

where, in the last step, we used the freedom to change the labelling of the summation indices and swapped $\mu \leftrightarrow \rho$ and $\nu \leftrightarrow \sigma$. In order for the transformation to leave the space-time interval ℓ^2 invariant, Eq. (6.1.11) has to be equal to $x^\mu g_{\mu\nu} x^\nu$, and this requires

$$\Lambda^\rho{}_\mu g_{\rho\sigma} \Lambda^\sigma{}_\nu = g_{\mu\nu}. \quad (6.1.12)$$

We can therefore use Eq. (6.1.12) as the definition of a Lorentz transformation.

Using the metric tensor, we can also define a scalar product of two four-vectors x and y as

$$x \cdot y = x^\mu g_{\mu\nu} y^\nu = x^0 y^0 - x^1 y^1 - x^2 y^2 - x^3 y^3. \quad (6.1.13)$$

This is also invariant under Lorentz transformations,

$$\ell^2 = x^\mu g_{\mu\nu} y^\nu \to x'^\mu g_{\mu\nu} y'^\nu = \Lambda^\mu{}_\rho x^\rho g_{\mu\nu} \Lambda^\nu{}_\sigma y^\sigma = x^\mu \Lambda^\rho{}_\mu g_{\rho\sigma} \Lambda^\sigma{}_\nu y^\nu = x^\mu g_{\mu\nu} y^\nu. \quad (6.1.14)$$

To simplify the notation further, we define a *covariant* vector x_μ by

$$x_\mu = g_{\mu\nu} x^\nu = (x^0, -x^1, -x^2, -x^3), \quad (6.1.15)$$

and indicate it by using a subscript index. We say that we use the metric to *lower the index*. The original position four-vector x^μ with a superscript index is called a *contravariant* vector. For example, the scalar product (6.1.13) is then simply

$$x \cdot y = x^\mu y_\mu. \quad (6.1.16)$$

To *raise the index*, i.e., turn a covariant vector back to a contravariant one, we need the inverse $\overline{\overline{\mathbf{g^{-1}}}}$ of the metric tensor, so that

$$x^\mu = (g^{-1})^{\mu\nu} x_\nu. \quad (6.1.17)$$

(Note that we use superscript indices to be consistent with the Einstein convention.) Eq. (6.1.17) is equivalent to saying that it is the inverse matrix of $g_{\mu\nu}$, defined in the usual way by

$$(g^{-1})^{\mu\nu} g_{\nu\rho} = \delta^\mu_\rho, \quad (6.1.18)$$

where

$$\delta^\mu_\rho = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad (6.1.19)$$

is the 4×4 unit matrix.

For the Minkowski metric (6.1.5) it is easy to find the inverse, and it turns out to be the same matrix as the metric itself,

$$(g^{-1})^{\mu\nu} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} = g_{\mu\nu}, \quad (6.1.20)$$

but this is not the case in general relativity. In any case, using the definition of the inverse metric (6.1.18), we can see that it satisfies

$$g_{\mu\rho}g_{\nu\sigma}(g^{-1})^{\rho\sigma} = g_{\mu\nu}. \tag{6.1.21}$$

This means that when we lower the indices of the inverse metric $(g^{-1})^{\mu\nu}$ in the same way as in Eq. (6.1.15) we obtain the original metric $g_{\mu\nu}$. We can therefore think of the inverse metric $(g^{-1})^{\mu\nu}$ as simply the contravariant counterpart of the covariant metric $g_{\mu\nu}$. In particular, this means that there is no need to indicate the inverse metric by "-1" and we can simply write

$$(g^{-1})^{\mu\nu} = g^{\mu\nu} \tag{6.1.22}$$

without any risk of confusion. Then Eq. (6.1.18) becomes

$$g^{\mu\nu}g_{\nu\rho} = \delta^{\mu}_{\rho}, \tag{6.1.23}$$

and the expression for raising the index (6.1.17) simplifies to

$$x^{\mu} = g^{\mu\nu}x_{\nu}. \tag{6.1.24}$$

We can treat all Lorentz indices in this way, using $g_{\mu\nu}$ to lower a contravariant superscript index to a covariant subscript, and $g^{\mu\nu}$ to raise a covariant subscript index to a contravariant superscript. In particular, if we multiply both sides of Eq. (6.1.12) by $g^{\lambda\mu}$, we find

$$g^{\lambda\mu}\Lambda^{\rho}{}_{\mu}g_{\rho\sigma}\Lambda^{\sigma}{}_{\nu} = g^{\lambda\mu}g_{\mu\nu} = \delta^{\lambda}_{\nu}. \tag{6.1.25}$$

Comparing this with the definition of the *inverse Lorentz transformation* $\overline{\overline{\mathbf{\Lambda^{-1}}}}$, which takes the system back from the boosted to the original frame,

$$(\Lambda^{-1})^{\lambda}{}_{\sigma}\Lambda^{\sigma}{}_{\nu} = \delta^{\lambda}_{\nu}, \tag{6.1.26}$$

we find that

$$(\Lambda^{-1})^{\lambda}{}_{\sigma} = g^{\lambda\mu}\Lambda^{\rho}{}_{\mu}g_{\rho\sigma} \equiv \Lambda_{\sigma}{}^{\lambda}. \tag{6.1.27}$$

We can also derive the transformation law for covariant vectors,

$$x'_{\mu} = g_{\mu\nu}x'^{\nu} = g_{\mu\nu}\Lambda^{\nu}{}_{\rho}x^{\rho} = g_{\mu\nu}\Lambda^{\nu}{}_{\rho}g^{\rho\lambda}x_{\lambda} = \Lambda_{\mu}{}^{\lambda}x_{\lambda}. \tag{6.1.28}$$

Comparing with Eq. (6.1.27) we see that the transformation matrix for the covariant vectors is the inverse of the contravariant transformation matrix.

Besides the position four-vector x^{μ}, there are other quantities that transform in the same way under Lorentz transformations and can therefore be naturally written as four-vectors. These include

- The four-velocity u^{μ}, defined as

$$u^{\mu} = \frac{dx^{\mu}}{d\tau}, \tag{6.1.29}$$

  where τ is the *proper time* is the time measured by the observer moving along the trajector defined by x^{μ}. It is defined as $c^2d\tau^2 = c^2dt^2 - dx^2 - dy^2 - dz^2$, or $dt = \gamma d\tau$. It follows that

$$u^{\mu} = (\gamma c, \gamma v_x, \gamma v_y, \gamma v_z). \tag{6.1.30}$$

- Four-momentum
$$p^\mu = mu^\mu = (E/c, p_x, p_y, p_z). \tag{6.1.31}$$
- Four-current density
$$j^\mu = nqu^\mu = (\gamma nqc, \gamma nqv_x, \gamma nqv_y, \gamma nqv_z) = (\rho c, j_x, j_y, j_z). \tag{6.1.32}$$

These all transform as contravariant vectors, i.e., $u'^\mu = \Lambda^\mu{}_\nu u^\nu$ etc., although of course we can always lower the index with the metric to turn them into the covariant form, $u_\mu = g_{\mu\nu}u^\nu$, when it is more convenient.

For an example of a four-vector that is more natural to think of as a covariant vector, consider a scalar function $f(x)$ of spacetime, and its derivative with respect to the contravariant position vector x^μ. Using the chain rule of derivatives, and the inverse Lorentz transformation $x^\nu = (\Lambda^{-1})^\nu{}_\mu x'^\mu = \Lambda_\mu{}^\nu x'^\mu$, we find
$$\frac{\partial f(x)}{\partial x'^\mu} = \sum_\nu \frac{\partial x^\nu}{\partial x'^\mu}\frac{\partial f(x)}{\partial x^\nu} = \Lambda_\mu{}^\nu \frac{\partial f(x)}{\partial x^\nu}. \tag{6.1.33}$$
Comparing with Eq. (6.1.28), we see that a derivative with respect to a contravariant vector transforms as a covariant vector. Therefore we use the notation
$$\partial_\mu \equiv \frac{\partial}{\partial x^\mu}, \tag{6.1.34}$$
to make this explicit. In this notation, Eq. (6.1.33) becomes
$$\partial'_\mu f = \Lambda_\mu{}^\nu \partial_\nu f. \tag{6.1.35}$$
Similarly, a derivative with respect to a covariant vector transforms as a contravariant vector, and therefore we write
$$\partial^\mu \equiv \frac{\partial}{\partial x_\mu}. \tag{6.1.36}$$

Remember that a tensor is a linear relationship between two or more vectors. In special relativity we expect that the same linear relationship remains valid in all reference frames, in which case it is called a *Lorentz tensor*. Consider, for example, the rank 4 tensor $M^\mu_{\nu\rho\sigma}$ in Eq. (6.1.9). Lorentz boosting the right-hand-side, we find
$$x'^\mu = \Lambda^\mu{}_\lambda x^\lambda = \Lambda^\mu{}_\lambda M^\lambda{}_{\nu\rho\sigma} y^\nu z^\rho w^\sigma = \Lambda^\mu{}_\lambda M^\lambda{}_{\nu\rho\sigma}\Lambda_\alpha{}^\nu y'^\alpha \Lambda_\beta{}^\rho z'^\beta \Lambda_\gamma{}^\sigma w'^\gamma, \tag{6.1.37}$$
where in the last step we used the inverse Lorentz transformation. We want to be able to write this as
$$x'^\mu = M'^\mu{}_{\alpha\beta\gamma} y'^\alpha z'^\beta w'^\gamma, \tag{6.1.38}$$
which means that the boosted tensor has to be
$$M'^\mu{}_{\alpha\beta\gamma} = \Lambda^\mu{}_\lambda \Lambda_\alpha{}^\nu \Lambda_\beta{}^\rho \Lambda_\gamma{}^\sigma M^\lambda{}_{\nu\rho\sigma} \tag{6.1.39}$$
We can see that each superscript index transforms with the contravariant transformation matrix, and each subscript index with the covariant transformation matrix, just like in four-vectors.

Whenever an index is summed over (*contracted*) according to the Einstein convention, the sum is Lorentz invariant, so summed indices can be ignored when doing Lorentz transformations. For example,
$$M'^\mu{}_{\mu\beta\gamma} = \Lambda^\mu{}_\lambda \Lambda_\mu{}^\nu \Lambda_\beta{}^\rho \Lambda_\gamma{}^\sigma M^\lambda{}_{\nu\rho\sigma} = \delta^\nu_\lambda \Lambda_\beta{}^\rho \Lambda_\gamma{}^\sigma M^\lambda{}_{\nu\rho\sigma} = \Lambda_\beta{}^\rho \Lambda_\gamma{}^\sigma M^\mu{}_{\mu\rho\sigma} \tag{6.1.40}$$
where we used the property (6.1.25). This shows why the Einstein convention is so useful in special relativity: Because the laws of nature are supposed to be the same in all inertial frames, pairs of indices should only appear in this Lorentz invariant form.

6.2 Relativistic Electrodynamics

Historically, electrodynamics played a key role in the development of the theory of relativity, and electrodynamics appears much more elegent in a fully relativistic formulation. However, it is not entirely trivial to write electric and magnetic fields in a four-vector form. For example, we saw in Section 5.5 that electric and magnetic fields have to somehow transform to each other under Lorentz transformations, so they cannot be two separate four-vectors.

To derive the relativistic formulation, we start from the expression for the Lorentz force,

$$\frac{d\mathbf{p}}{dt} = \mathbf{F} = q(\mathbf{E} + \mathbf{v} \times \mathbf{B}). \tag{6.2.1}$$

It follows directly that the derivative with respect to the proper time is

$$\frac{d\mathbf{p}}{d\tau} = \gamma \frac{d\mathbf{p}}{dt} = q\left(\frac{u^0}{c}\mathbf{E} + \mathbf{u} \times \mathbf{B}\right), \tag{6.2.2}$$

where $\mathbf{u} = \gamma\mathbf{v}$ is the spatial part of the four-velocity u^μ.

The time derivative of the energy of the particle is

$$\frac{dE}{dt} = \mathbf{F} \cdot \mathbf{v} = q\mathbf{E} \cdot \mathbf{v}, \tag{6.2.3}$$

from which we obtain the derivative with respect to the proper time as

$$\frac{dE}{d\tau} = \gamma \frac{dE}{dt} = q\mathbf{E} \cdot \mathbf{u}. \tag{6.2.4}$$

We can now combine Eqs. (6.2.2) and (6.2.4) into the proper time derivative of the four-momentum $p^\mu = (E/c, \mathbf{p})$,

$$\frac{dp^\mu}{d\tau} = \frac{d}{d\tau}\begin{pmatrix} E/c \\ p_x \\ p_y \\ p_z \end{pmatrix} = q\begin{pmatrix} 0 & E_x/c & E_y/c & E_z/c \\ E_x/c & 0 & B_z & -B_y \\ E_y/c & -B_z & 0 & B_x \\ E_z/c & B_y & -B_x & 0 \end{pmatrix}\begin{pmatrix} u^0 \\ u^1 \\ u^2 \\ u^3 \end{pmatrix}. \tag{6.2.5}$$

The matrix appearing in this expression is called the *Faraday tensor* or the field-strength tensor and denoted by $F^\mu{}_\nu$, so we can write more compactly the relativistic Lorentz force equation as

$$\frac{dp^\mu}{d\tau} = qF^\mu{}_\nu u^\nu. \tag{6.2.6}$$

The Faraday tensor is often written with two contravariant indices as

$$F^{\mu\nu} = F^\mu{}_\rho g^{\rho\nu} = \begin{pmatrix} 0 & -E_x/c & -E_y/c & -E_z/c \\ E_x/c & 0 & -B_z & B_y \\ E_y/c & B_z & 0 & -B_x \\ E_z/c & -B_y & B_x & 0 \end{pmatrix}. \tag{6.2.7}$$

The Lorentz force equation (6.2.6) then becomes

$$\frac{dp^\mu}{d\tau} = qF^{\mu\nu}u_\nu. \tag{6.2.8}$$

Note that this tensor is antisymmetric, $F^{\nu\mu} = -F^{\mu\nu}$.

In order for the right-hand-side of Eq. (6.2.6) to be Lorentz contravariant, $F^{\mu\nu}$ has to transform as a contravariant rank 2 Lorentz tensor,

$$F'^{\mu\nu} = \Lambda^{\mu}{}_{\rho}\Lambda^{\nu}{}_{\sigma}F^{\rho\sigma}. \tag{6.2.9}$$

This tells us how the electric and magnetic fields must transform. For example, considering a boost in z direction,

$$\Lambda^{\mu}{}_{\rho} = \begin{pmatrix} \gamma & 0 & 0 & -\beta\gamma \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -\beta\gamma & 0 & 0 & \gamma \end{pmatrix}, \tag{6.2.10}$$

we find that the electric and magnetic fields transform as

$$\begin{aligned} E'_x &= \gamma\left(E_x - vB_y\right), \\ E'_y &= \gamma\left(E_y + vB_x\right), \\ E'_z &= E_z\,, \end{aligned} \tag{6.2.11}$$

$$\begin{aligned} B'_x &= \gamma\left(B_x + vE_y/c^2\right), \\ B'_y &= \gamma\left(B_y - vE_x/c^2\right), \\ B'_z &= B_z\,. \end{aligned} \tag{6.2.12}$$

More generally, we can write

$$\begin{aligned} \mathbf{E}'_{\parallel} &= \mathbf{E}_{\parallel}, \\ \mathbf{B}'_{\parallel} &= \mathbf{B}_{\parallel}, \\ \mathbf{E}'_{\perp} &= \gamma\left(\mathbf{E}_{\perp} + \mathbf{v}\times\mathbf{B}\right), \\ \mathbf{B}'_{\perp} &= \gamma\left(\mathbf{B}_{\perp} - \mathbf{v}\times\mathbf{E}/c^2\right) = \gamma\left(\mathbf{B}_{\perp} - \mu_0\epsilon_0\mathbf{v}\times\mathbf{E}\right), \end{aligned} \tag{6.2.13}$$

where $\parallel$ refers to the component parallel to the boost velocity, and $\perp$ to the perpendicular components. If $\mathbf{B} = 0$, then

$$\mathbf{B}' = -\frac{\mathbf{v}\times\mathbf{E}'}{c^2}, \tag{6.2.14}$$

in agreement with Eq. (5.5.18) (note the opposite sign of $\mathbf{v}$).

6.3 Maxwell's Equations

Now that we have combined the electric and magnetic fields into one Lorentz tensor $F^{\mu\nu}$, we want to write Maxwell's equations (5.1.1) in terms of it. We start by noting that, according to Eq. (6.2.7), the electric and magnetic fields are given by

$$\begin{aligned} \mathbf{E} &= (cF^{10}, cF^{20}, cF^{30}), \\ \mathbf{B} &= (F^{32}, F^{13}, F^{21}). \end{aligned} \tag{6.3.1}$$

We now write Gauss's law as

$$\begin{aligned} \boldsymbol{\nabla}\cdot\mathbf{E} - \frac{\rho}{\epsilon_0} &= \frac{\partial E_x}{\partial x} + \frac{\partial E_y}{\partial y} + \frac{\partial E_z}{\partial z} - \frac{\rho}{\epsilon_0} = c\left(\frac{\partial F^{10}}{\partial x} + \frac{\partial F^{20}}{\partial y} + \frac{\partial F^{30}}{\partial z} - \frac{\rho}{c\epsilon_0}\right) \\ &= c\left(\partial_1 F^{10} + \partial_2 F^{20} + \partial_3 F^{30} - \frac{\rho}{c\epsilon_0}\right) = c\left(\partial_\mu F^{\mu 0} - \frac{\rho}{c\epsilon_0}\right). \end{aligned} \tag{6.3.2}$$

Therefore we have

$$\partial_\mu F^{\mu 0} = \frac{\rho}{c\epsilon_0} = \mu_0 c\rho. \tag{6.3.3}$$

To deal with Ampère's law,

$$\boldsymbol{\nabla} \times \mathbf{B} - \mu_0 \mathbf{J} - \mu_0 \epsilon_0 \frac{\partial \mathbf{E}}{\partial t} = 0 \tag{6.3.4}$$

we write the x component

$$\begin{aligned} \frac{\partial B_z}{\partial y} - \frac{\partial B_y}{\partial z} - \mu_0 j_x - \mu_0 \epsilon_0 \frac{\partial E_x}{\partial t} &= \frac{\partial F^{21}}{\partial y} - \frac{\partial F^{13}}{\partial z} - \mu_0 j_x - \frac{1}{c}\frac{\partial F^{10}}{\partial t} \\ &= \partial_0 F^{01} + \partial_2 F^{21} + \partial_3 F^{31} - \mu_0 j_x \\ &= \partial_\mu F^{\mu 1} - \mu_0 j_x = 0, \end{aligned} \tag{6.3.5}$$

so we have

$$\partial_\mu F^{\mu 1} = \mu_0 j_x, \tag{6.3.6}$$

and similarly for the other components. We can now write Eqs. (6.3.3) and (6.3.6) as one equation in terms of the four-current $j^\mu = (c\rho, \mathbf{j})$,

$$\partial_\mu F^{\mu\nu} = \mu_0 j^\nu. \tag{6.3.7}$$

The magnetic Gauss's law reads

$$\begin{aligned} \boldsymbol{\nabla} \cdot \mathbf{B} &= \frac{\partial B_x}{\partial x} + \frac{\partial B_y}{\partial y} + \frac{\partial B_z}{\partial z} = \frac{\partial F^{32}}{\partial x} + \frac{\partial F^{13}}{\partial y} + \frac{\partial F^{21}}{\partial z} = \partial_1 F^{32} + \partial_2 F^{13} + \partial_3 F^{21} \\ &= \partial^1 F^{23} + \partial^2 F^{31} + \partial^3 F^{12} = 0. \end{aligned} \tag{6.3.8}$$

For Faraday's law

$$\boldsymbol{\nabla} \times \mathbf{E} + \frac{\partial \mathbf{B}}{\partial t} = 0, \tag{6.3.9}$$

we again take the x component,

$$\begin{aligned} \frac{\partial B_x}{\partial t} + \frac{\partial E_z}{\partial y} - \frac{\partial E_y}{\partial z} &= \frac{\partial F^{32}}{\partial t} + c\frac{\partial F^{30}}{\partial y} - c\frac{\partial F^{20}}{\partial z} = c\left(\partial_0 F^{32} - \partial_2 F^{03} - \partial_3 F^{20}\right) \\ &= c\left(\partial^0 F^{32} + \partial^2 F^{03} + \partial^3 F^{20}\right) = 0 \end{aligned} \tag{6.3.10}$$

By comparing Eqs. (6.3.8) and (6.3.10), we note that we can combine them into one equation

$$\partial^\mu F^{\nu\rho} + \partial^\nu F^{\rho\mu} + \partial^\rho F^{\mu\nu} = 0. \tag{6.3.11}$$

Thus, we have found that in the four-vector notation, the four Maxwell's equations (5.1.1) can be expressed by just two equations (6.3.7) and (6.3.11).

6.4 Four-vector Potential

In Section 5.1, we saw that we can express the electric and magnetic fields using the scalar and vector potentials, and that reduced tha number of non-trivial Maxwell's equations from four to two. We will now see that we can do the same to the Faraday tensor $F^{\mu\nu}$, and that way reduce Maxwell's equations to just one.

Using Eq. (5.1.6), we can write

$$F^{10} = \frac{E_x}{c} = -\frac{1}{c}\left(\frac{\partial A_x}{\partial t} + \frac{\partial \phi}{\partial x}\right) = -\partial_0 A_x - \partial_1\left(\frac{\phi}{c}\right) = \partial^1\left(\frac{\phi}{c}\right) - \partial^0 A_x, \tag{6.4.1}$$

and

$$F^{21} = B_z = \frac{\partial A_y}{\partial x} - \frac{\partial A_x}{\partial y} = \partial_1 A_y - \partial_2 A_x = \partial^2 A_x - \partial^1 A_y. \tag{6.4.2}$$

By defining the *four-vector potential* $A^\mu = (\phi/c, \mathbf{A})$, we can write these two equations as

$$\begin{aligned} F^{10} &= \partial^1 A^0 - \partial^0 A^1 \\ F^{21} &= \partial^2 A^1 - \partial^1 A^2. \end{aligned} \tag{6.4.3}$$

Similar relations apply to other components of $F^{\mu\nu}$, so we can combine them into one equation

$$F^{\mu\nu} = \partial^\mu A^\nu - \partial^\nu A^\mu. \tag{6.4.4}$$

Using this equation, we can write the Faraday tensor in terms of the four-vector potential. Finally, we want to express Maxwell's equations (6.3.7) and (6.3.11) in terms of A^μ. Eq. (6.3.7) becomes

$$\partial_\mu F^{\mu\nu} = \partial_\mu \partial^\mu A^\nu - \partial^\nu \partial_\mu A^\mu = \mu_0 j^\nu. \tag{6.4.5}$$

For Eq. (6.3.11), we find that

$$\partial^\mu F^{\nu\rho} + \partial^\nu F^{\rho\mu} + \partial^\rho F^{\mu\nu} = \partial^\mu\partial^\nu A^\rho - \partial^\mu\partial^\rho A^\nu + \partial^\nu\partial^\rho A^\mu - \partial^\nu\partial^\mu A^\rho + \partial^\rho\partial^\mu A^\nu - \partial^\rho\partial^\nu A^\mu = 0 \tag{6.4.6}$$

identically, because each term appears twice with opposite signs. Therefore, Eq. (6.3.11) is automatically satisfied when the Faraday tensor is expressed in terms of the four-vector potential. The only non-trivial equation is therefore Eq. (6.4.5).

6.5 Lagrangian for Electrodynamics

Finally, let us see how we can describe electrodynamics in the Lagrangian formulation. Because the electromagnetic fields are continuous fields, we need to find a Lagrangian density $\mathcal{L}$ as defined in Section 3.6, and we will express it in terms of the four-vector potential A^μ.

We know that electrodynamics is invariant under both gauge and Lorentz transformations. Therefore the Lagrangian density $\mathcal{L}$ has to be a Lorentz scalar, and to be gauge invariant it can only depend on the four-vector potential through the Faraday tensor $F^{\mu\nu}$. It also makes sense to demand that the Lagrangian density should contain only first time derivatives and they should appear only in quadratic form. This corresponds to a "natural" system as defined in Section 3.5.1, and it ensures that the Euler-Lagrange equations have the familiar form. The expression that satisfies these requirements is $F^{\mu\nu}F_{\mu\nu}$, and therefore we are led to consider a Lagrangian density of the form

$$\mathcal{L} = aF^{\mu\nu}F_{\mu\nu}, \tag{6.5.1}$$

where a is some constant. Actually, the numerical value of a does not matter because it will drop out the Euler-Lagrange equation, but we will see later that the sign should be negative. It is conventional to choose $a = -1/4$, so that we have

$$\mathcal{L} = -\frac{1}{4} F^{\mu\nu} F_{\mu\nu}. \tag{6.5.2}$$

In terms of the four-vector potential, this becomes

$$\mathcal{L} = -\frac{1}{4} (\partial^\mu A^\nu - \partial^\nu A^\mu)(\partial_\mu A_\nu - \partial_\nu A_\mu) = -\frac{1}{2} \partial^\mu A^\nu (\partial_\mu A_\nu - \partial_\nu A_\mu). \tag{6.5.3}$$

In order to derive the Euler-Lagrange equation, we first write Eq. (3.6.12) in a four-vector form,

$$\partial_\mu \frac{\partial \mathcal{L}}{\partial(\partial_\mu y)} - \frac{\partial \mathcal{L}}{\partial y} = 0, \tag{6.5.4}$$

and generalise it to the current case by replacing y with A_ν. Because the Lagrangian $\mathcal{L}$ depends only on its derivatives, we find the equation

$$\partial_\mu \frac{\partial \mathcal{L}}{\partial(\partial_\mu A_\nu)} = 0. \tag{6.5.5}$$

To avoid any confusion about the derivative in this expression, it is best to lower all the indices in Eq. (6.5.3) and write it in the form

$$\mathcal{L} = -\frac{1}{2} g^{\kappa\rho} g^{\lambda\sigma} \partial_\kappa A_\lambda (\partial_\rho A_\sigma - \partial_\sigma A_\rho). \tag{6.5.6}$$

Then the derivative is easy to take by noting that it is non-zero only if the Lorentz indices match, that is,

$$\frac{\partial (\partial_\kappa A_\lambda)}{\partial (\partial_\mu A_\nu)} = \delta^\mu_\kappa \delta^\nu_\lambda. \tag{6.5.7}$$

We find

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial(\partial_\mu A_\nu)} &= -\frac{1}{2} g^{\kappa\rho} g^{\lambda\sigma} \left[\delta^\mu_\kappa \delta^\nu_\lambda (\partial_\rho A_\sigma - \partial_\sigma A_\rho) + \partial_\kappa A_\lambda \left(\delta^\mu_\rho \delta^\nu_\sigma - \delta^\mu_\sigma \delta^\nu_\rho\right)\right] \\ &= -\frac{1}{2} \left[g^{\mu\rho} g^{\nu\sigma} (\partial_\rho A_\sigma - \partial_\sigma A_\rho) + \partial_\kappa A_\lambda \left(g^{\mu\kappa} g^{\nu\lambda} - g^{\mu\lambda} g^{\nu\kappa}\right)\right] \\ &= -\frac{1}{2} \left[\partial^\mu A^\nu - \partial^\nu A^\mu + \partial^\mu A^\nu - \partial^\nu A^\mu\right] = -F^{\mu\nu}, \end{aligned} \tag{6.5.8}$$

and therefore the Euler-Lagrange equation (6.5.5) is

$$\partial_\mu \frac{\partial \mathcal{L}}{\partial(\partial_\mu A_\nu)} = -\partial_\mu F^{\mu\nu} = 0. \tag{6.5.9}$$

Which is exactly the Maxwell equation (6.3.7) in vacuum, i.e., with $j^\mu = 0$. Because the other Maxwell equation (6.3.11) is satisfied identically when using the four-vector potential A^μ, we have shown that the laws of electrodynamics in vacuum are correctly described by the Lagrangian (6.5.3) which we obtained by assuming essentially only gauge and Lorentz invariance. This demonstrates how powerful symmetry considerations can be in physics, and in fact the properties of the other fundamental interactions (strong and weak nuclear force, and gravity) are also determined by their corresponding gauge invariances.