# Foundations of Mathematical Physics

Paul P. Cook*

*Department of Mathematics, King's College London*
*The Strand, London WC2R 2LS, UK*

*email: paul.cook@kcl.ac.uk

# Contents


**1 Classical Mechanics** **5**

1.1 Lagrangian Mechanics . . . . . 5

1.1.1 Conserved Quantities . . . . . 10

1.2 Hamiltonian Mechanics . . . . . 10

1.2.1 Hamilton's equations. . . . . 11

1.2.2 Poisson Brackets . . . . . 11

1.2.3 Duality and the Harmonic Oscillator . . . . . 12

1.3 Noether's Theorem . . . . . 13

1.3.1 A Sideways Glance at Noether's Theorem. . . . . 17

1.3.2 Noether's theorem in the Hamiltonian formulation. . . . . 18

**2 Special Relativity and Component Notation** **23**

2.1 The Special Theory of Relativity . . . . . 23

2.1.1 The Lorentz Group and the Minkowski Inner Product. . . . . 26

2.2 Component Notation. . . . . 31

2.2.1 Matrices and Matrix Multiplication. . . . . 32

2.2.2 Common Four-Vectors . . . . . 34

2.2.3 Maxwell's Equations. . . . . 36

2.2.4 Electromagnetic Duality . . . . . 40

2.2.5 Field Theory Equations of Motion . . . . . 40

**3 Quantum Mechanics** **43**

3.1 Canonical Quantisation . . . . . 44

3.1.1 The Hilbert Space and Observables. . . . . 45

3.1.2 Eigenvectors and Eigenvalues . . . . . 47

3.1.3 A Countable Basis. . . . . 48

3.1.4 A Continuous Basis. . . . . 51

3.2 The Schrödinger Equation. . . . . 53

3.2.1 The Heisenberg and Schrödinger Pictures. . . . . 54

**4 Group Theory** **59**

4.1 The Basics . . . . . 60

4.2 Common Groups . . . . . 61

4.2.1 The Symmetric Group $S_n$ . . . . . 61

4.2.2 Back to Basics . . . . . 65

4.3 Group Homomorphisms . . . . . 67

4.3.1 The First Isomomorphism Theorem . . . . . 70
4.4 Some Representation Theory . . . . . 71
4.4.1 Schur's Lemma . . . . . 72
4.4.2 The Direct Sum and Tensor Product . . . . . 75
4.5 Lie Groups . . . . . 79
4.5.1 Infinitesimal Generators and the Invariance of Physical Law . . . 81
4.5.2 Lie Algebras . . . . . 84
4.5.3 Examples of interest in theoretical physics . . . . . 87

**5 Special topics in Mathematical Physics** **95**
5.1 Famous Differential Equations in Physics . . . . . 95
5.2 Classical Orthogonal Polynomials . . . . . 98
5.2.1 Recurrence Relations . . . . . 101
5.2.2 The Hermite Polynomials . . . . . 102
5.3 Green's Functions . . . . . 105
5.3.1 The Dirac delta function and spherical coordinates. . . . . 106
5.3.2 The Green's function for the Laplacian operator . . . . . 110
5.4 Some Complex Calculus . . . . . 112
5.4.1 Differentiation of Complex Functions . . . . . 112
5.4.2 Integration of Complex Functions . . . . . 114
5.4.3 Laurent Series . . . . . 117

# Chapter 1

# Classical Mechanics

## 1.1 Lagrangian Mechanics

Newton's second law of motion states that for a body of constant mass $m$ acted on by a force $\mathbf{F}$

$$\mathbf{F} = \frac{d}{dt}(\mathbf{p}) = m\ddot{\mathbf{x}} \tag{1.1}$$

where $\mathbf{p}$ is the linear momentum ($\mathbf{p} \equiv m\dot{\mathbf{x}}$), $x$ is the position of the body and $\dot{\mathbf{x}} \equiv \frac{d\mathbf{x}}{dt}$. Hence if $\mathbf{F} = 0$ then the linear momentum is conserved $\dot{\mathbf{p}} = 0$.

$\mathbf{F}$ is called a *conservative force* if the energy of the system is constant for motion under the force. This is equivalent to two statements:

(i) The work done under the force is path-independent, and

(ii) The force may derived from a scalar field.

The work done by a mass $m$ subject to a force $\mathbf{F}$ moving on a path from $\mathbf{x}(t_1)$ to $\mathbf{x}(t_2)$ is

$$\begin{aligned}
\Delta W &= \int_{\mathbf{x}(t_1)}^{\mathbf{x}(t_2)} \mathbf{F} \cdot d\mathbf{x} \\
&= \int_{t_1}^{t_2} \mathbf{F} \cdot \dot{\mathbf{x}} dt \\
&= \int_{t_1}^{t_2} m\ddot{\mathbf{x}} \cdot \dot{\mathbf{x}} dt \\
&= \int_{t_1}^{t_2} m\frac{d}{dt}(\frac{1}{2}\dot{\mathbf{x}}^2)\, dt \\
&= \frac{1}{2}m\dot{x}^2(t_2) - \frac{1}{2}m\dot{x}^2(t_1) \\
&\equiv \Delta T
\end{aligned} \tag{1.2}$$

where $T \equiv \frac{1}{2}m\dot{x}^2$ is the kinetic energy.

In general the work done depends on the precise path taken from $\mathbf{x}(t_1)$ to $\mathbf{x}(t_2)$. It would seem common-sense that to push a supermarket trolley from $\mathbf{x}(t_1)$ to $\mathbf{x}(t_2)$ requires an amount of work that is path-dependent - a path may be short or long, it might traverse a hill or go around it - and one might expect the amount of work to vary for each path. For many theoretical examples including these where work has to

be done against and by the force of gravity the work function is path-independent. An example of a path-dependent work function is the work done against friction[1]

Whenever $\Delta W$ is path-independent the force $\mathbf{F}$ is called conservative and such a force can always be derived from a scalar field $V$, called the potential, as

$$\mathbf{F} = -\boldsymbol{\nabla} V. \tag{1.3}$$

When $\mathbf{F}$ is conservative the work function $\Delta W$ depends only on the values of $V$ at the endpoints of the path:

$$\begin{aligned} \Delta W &= \int_{t_1}^{t_2} -\boldsymbol{\nabla} V \cdot \dot{\mathbf{x}}\, dt \\ &= \int_{t_1}^{t_2} -(\frac{\partial V}{\partial x}\frac{dx}{dt} + \frac{\partial V}{\partial y}\frac{dy}{dt} + \frac{\partial V}{\partial z}\frac{dz}{dt})\, dt \\ &= \int_{t_1}^{t_2} -(\frac{dV}{dt})\, dt \\ &= -(V(t_2) - V(t_1)). \end{aligned} \tag{1.4}$$

In terms of kinetic energy we had $\Delta W = T(t_2) - T(t_1)$ hence,

$$\begin{aligned} & T(t_2) - T(t_1) = V(t_1) - V(t_2) \\ \Rightarrow \quad & (T+V)(t_1) = (T+V)(t_2). \end{aligned} \tag{1.5}$$

Hence a conservative force conserves energy $E \equiv T + V$ over time.

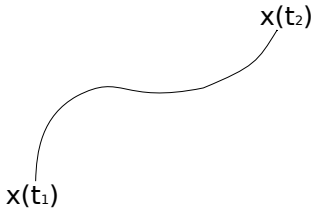
In terms of the potential $V$, Newton's second law of motion (for a constant mass) becomes:

$$-\frac{\partial V}{\partial x^i} \equiv -\partial_i V = m\ddot{x}^i \tag{1.6}$$

where $x^i$ are the components of the vector $x$ (i.e. $i \in \{1,2,3\}$) and we have introduced the notation $\partial_i$ for $\frac{\partial}{\partial x^i}$. This law of motion may be derived from a variational principle on the functional[2]

$$S = \int_{t_1}^{t_2} dt\mathcal{L} \tag{1.7}$$

called the action, where $\mathcal{L}$ is the Lagrangian. To each path the action assigns a number using the Lagrangian.




(1.8)

You may recall from optics the principle of least time which is used to discover which path a photon travels in moving from A to B. The path a photon takes when it is

[1] You might consider the work done moving around a closed loop. For a conservative force the work is zero (split the closed loop into two journeys from A to B and from B to A, as the work done by a conservative force depends only on A and B we have $W_{AB} = T_A - T_B = -W_{BA}$, hence total work around the loop equals $W_{AB} + W_{BA} = 0$). For work against a friction force there is positive contribution around every leg of the journey to the work which does not vanish when summed.

[2] A functional takes a vector as its argument and returns a scalar. The action is a function of the vectors $\mathbf{x}, \dot{\mathbf{x}}$ as well as the scalar time, $t$ and returns a real-valued number.

diffracted as it moves between two media is dictated by this principle. The situation for diffraction is analagous to the physicist on the beach who observes a drowning swimmer out at sea. The physicist knows that she can travel faster on the sand than she can swim so her optimal route will travel not in a straight line towards the swimmer but along a line which minimises the journey to the swimmer. This line will be bent in the middle and composed of two straight lines which change direction at the boundary between the sand and the sea. How does she work out which path she should follow to get to the swimmer in optimal time? Well she first derives a function which for each path to the swimmer computes the time the path takes to travel. Then she considers the infinitude of all possible paths to the swimmer and reads off from her function the time each path will take. The path that takes the shortest time will extremise her function (as will the longest time, if it exists), and she can find the quickest path to take in this way. Of course the swimmer may not thank her for taking so long. In a similar manner the action assigns a number to each motion a system may make, and the dynamical motion is determined when the action is extremised. The action contains the Lagrangian which is defined by

$$\begin{aligned}\mathcal{L}(\mathbf{x},\dot{\mathbf{x}};t) &\equiv T - V \\ &= \sum_{i=1}^{n} \frac{1}{2} m_i \dot{\mathbf{x}}_i^2 - \sum_{i=1}^{n} V_i \end{aligned} \tag{1.9}$$

for a system of $n$ particles of masses $m_i$ with position vectors $\mathbf{x}_i$ and velocities $\dot{\mathbf{x}}_\mathbf{i}$. Note that here we are not referring to the $i$'th component of a vector but rather the properties of the $i$'th particle. The equations of motion are found by extremising the action $S$. For simplicity of notation we will consider only a one-particle system (i.e. $n = 1$),

$$\begin{aligned}\delta S &= \int_{t_1}^{t_2} dt\, \delta\mathcal{L} \\ &= \int_{t_1}^{t_2} dt\, \delta(\frac{1}{2} m\dot{\mathbf{x}}^2 - V(\mathbf{x})) \\ &= \int_{t_1}^{t_2} dt\, [m\dot{\mathbf{x}}\delta\dot{\mathbf{x}} - \delta V((\mathbf{x}))] \\ &= \int_{t_1}^{t_2} dt\, [m\dot{\mathbf{x}}\frac{d}{dt}(\delta\mathbf{x}) - \partial_i V \delta x^i] \\ &= \int_{t_1}^{t_2} dt\, [-\frac{d}{dt}(m\dot{x_i}) - \partial_i V]\delta x^i + [\delta x^i m\dot{x_i}]_{t_1}^{t_2}\end{aligned} \tag{1.10}$$

where we have used integration by parts in the final line. Under the variation the action is expected to change at all orders:

$$S(\mathbf{x}+\delta\mathbf{x}) = S(\mathbf{x}) + \frac{\partial S}{\partial \mathbf{x}}\delta\mathbf{x} + \mathcal{O}((\delta\mathbf{x})^2) \equiv S + \delta S + \mathcal{O}((\delta\mathbf{x})^2) \tag{1.11}$$

When the first order variation of $S$ vanishes ($\frac{\partial S}{\partial \mathbf{x}} = 0$) the action is extremised. Each path from $\mathbf{x}(t_1)$ to $\mathbf{x}(t_2)$ gives a different value of the action, and the extremisation of the action occurs only for certain paths between the fixed points. From above we see that when $\delta S = 0$ (and noting that the endpoints of the path are fixed hence

$\delta\mathbf{x}(t_1) = \delta\mathbf{x}(t_2) = 0$) then

$$\begin{aligned}\delta S &= \int_{t_1}^{t_2} dt\,[-\frac{d}{dt}(m\dot{x}_i) - \partial_i V]\delta x^i \\ &= 0 \end{aligned} \tag{1.12}$$

for all $\delta x^i$. Which is satisfied only when Newton's law of motion is satisfied for the path with components $x_i$, (i.e. when $-\partial_i V = \frac{d}{dt}(m\dot{x}_i)$). This is no coincidence as Lagrange's equations may be derived from the Newton's second law.

More generally a generic dynamical system may be described by $n$ generalised coordinates $q_i$ and $n$ generalised velocities $\dot{q}_i$ where $i = 1, 2, 3, \ldots n$ and $n$ is the number of independent degrees of freedom of the system. The choice of generalised coordinates is where the art of dynamics resides. Imagine a system of $N$ particles moving in a three dimensional space $V$. There are $2 \times 3N$ Cartesian coordinates and velocities which describe this system. Now suppose further that the particles are all constrained to move on the surface of a sphere of radius $R$. One could make the change of coordinates to spherical coordinates but for each particle the radial coordinate would be redundant (since it is fixed to equal the sphere's readius $R$) and the new coordinates would be awash with trigonemtric functions. As the surface of the sphere is two-dimensional only two coordinates on the surface of the sphere are needed to identify a unique position. One reasonable choice is the angular variables $\theta$ and $\phi$ defined relative to the $x$-axis and the $z$-axis for example. These are independent coordinates and are an example of generalised coordinates. To summarise the example, each particle has three Cartesian coordinates which must satisfy one constraint: the equation $x^2 + y^2 + z^2 = R^2$, hence there are only two generalised coordinates per particle which may be chosen as $(\theta, \phi)$.

The Lagrangian function is defined via Cartesian coordinates, but constraint equations allow one to rewrite the Largrangian in terms of $q_i$ and $\dot{q}_i$, i.e. $\mathcal{L} = \mathcal{L}(q_i, \dot{q}_i; t)$. The equations of motion for the system are the (Euler-)Lagrange equations:

$$\frac{d}{dt}\left(\frac{\partial \mathcal{L}}{\partial \dot{q}_i}\right) - \frac{\partial \mathcal{L}}{\partial q_i} = 0 \tag{1.13}$$

**Problem 1.1.1.** *Derive the Lagrange equations for an abstract Lagrangian $\mathcal{L}(q_i, \dot{q}_i)$ by extremising the action $S$.*

**Example 1: The free particle.**

For a single free particle in $\mathbb{R}^3$ we have:

$$\mathcal{L} = T - V \tag{1.14}$$

$$= \frac{1}{2}m(\dot{x}^2 + \dot{y}^2 + \dot{z}^2) - V \tag{1.15}$$

The generalised coordinates may be picked to be any $n$ quantities which completely paramaterise the resulting path of the particle, in this case Cartesian coordinates suffice (i.e. let $q_1 \equiv x, q_2 \equiv y, q_3 \equiv z$). The particle is not subject to a force, hence $V = 0$ and hence the Lagrange equations (1.13) give

$$\frac{d}{dt}(m\dot{q}_i) = 0 \tag{1.16}$$

i.e. that linear momentum is conerved.

**Example 2: The linear harmonic oscillator.**

The system has one coordinate, $q$, and the potential is $V(q) = \frac{1}{2}kq^2$ where $k > 0$ (n.b. $\Rightarrow F = -kq$). The Lagrangian is

$$\mathcal{L} = \frac{1}{2}m\dot{q}^2 - \frac{1}{2}kq^2 \tag{1.17}$$

and the equation of motion (1.13) gives

$$\begin{aligned} \frac{d}{dt}(m\dot{q}) + kq = 0 \\ \Rightarrow \ddot{q} = -\frac{k}{m}q \end{aligned} \tag{1.18}$$

Hence we find

$$q(t) = A\cos(\omega t) + B\sin(\omega t) \tag{1.19}$$

where $\omega \equiv \sqrt{\frac{k}{m}}$ is the frequency of oscillation and $A$ and $B$ are real constants.

**Example 3: Circular motion.**

Consider a bead of mass $m$ constrained to move under gravity on a frictionless, circular, immobile, rigid hoop of radius $R$ such that the hoop lies in a vertical plane.

The Lagrangian formulation offers a neat way to ignore the forces of constraint (which keep the bead attached to the hoop) via the use of generalised coordinates. If the hoop rests in the $xz$-plane and is centred at $z = R$ then the Cartesian coordinates (in terms of a suitable chosen generalised coordinate $q \equiv \theta$) of the bead are:

$$\begin{aligned} x = R\cos\theta \quad &\Rightarrow \quad \dot{x} = -R\sin\theta\dot{\theta} \\ y = 0 \quad &\Rightarrow \quad \dot{y} = 0 \\ z = R + R\sin\theta \quad &\Rightarrow \quad \dot{z} = R\cos\theta\dot{\theta} \end{aligned} \tag{1.20}$$

These encode the statement that the bead is constrained to move on the hoop but without needing to consider any of the forces acting to keep the bead on the hoop. The Lagrangian is

$$\mathcal{L} = \frac{1}{2}m(\dot{x}^2 + \dot{y}^2 + \dot{z}^2) - V \tag{1.21}$$

$$= \frac{1}{2}m(R^2\dot{\theta}^2) - mg(R\sin\theta + R) \tag{1.22}$$

where we have used the gravitational potential $V = mgz (\Rightarrow -\partial_z V = -mg \equiv F_G)$. The equations of motion (1.13) are

$$\begin{aligned} \frac{d}{dt}(mR^2\dot{\theta}) - mgR\cos\theta = 0 \\ \Rightarrow mR^2\ddot{\theta} = mgR\cos\theta \\ \therefore \ddot{\theta} = \left(\frac{g}{R}\right)\cos\theta \\ = \left(\frac{g}{R}\right)(1 - \frac{\theta^2}{2} + \mathcal{O}(\theta^4)) \end{aligned} \tag{1.23}$$

For $\theta << 1$ we have $\ddot{\theta} \approx \frac{g}{R} \Rightarrow \theta \approx \frac{1}{2}(\frac{g}{R})t^2 + At + B$ where $A$ and $B$ are real constants. Obviously the assumption used for this approximation fails after a short time!

### 1.1.1 Conserved Quantities

For every ignorable coordinate in the Lagrangian there is an associated conserved quantity. That is if $\mathcal{L}(q_i, \dot{q}_i; t)$ satisfies $\frac{\partial \mathcal{L}}{\partial q_i} = 0$ then, as a consequence of 1.13,

$$\frac{d}{dt}\left(\frac{\partial \mathcal{L}}{\partial \dot{q}_i}\right) = 0 \tag{1.24}$$

and $\frac{\partial \mathcal{L}}{\partial \dot{q}_i}$ is conserved. This quantity is called the generalised momentum $p_i$ associated to the generalised coordinate $q_i$:

$$p_i \equiv \frac{\partial \mathcal{L}}{\partial \dot{q}_i}. \tag{1.25}$$

For example, consider free circular motion (set V=0 in the last example), where we have:

$$\mathcal{L} = \frac{1}{2} m R^2 \dot{\theta}^2. \tag{1.26}$$

We observe that $\theta$ is an ignorable coordinate as $\frac{\partial \mathcal{L}}{\partial \theta} = 0$ and hence $p_\theta = mR^2\dot{\theta}$ is conserved. This is the conservation of angular momentum, as $|\mathbf{r} \wedge \mathbf{p}| = p_\theta$, as you may confirm.

## 1.2 Hamiltonian Mechanics

Hamiltonians also encode the dynamics of a physical system. There is an invertible map from a Lagrangian to a Hamiltonian so no information is lost. The map is the Legendre transform and is used to define the Hamiltonian $\mathcal{H}$:

$$\mathcal{H}(q_i, p_i; t) = \sum_i \dot{q}_i p_i - \mathcal{L} \tag{1.27}$$

**Example.**

Let $\mathcal{L} = \sum_i \frac{1}{2}\dot{q}_i^2 - V(q)$ then $p_i = \frac{\partial \mathcal{L}}{\partial \dot{q}_i} = m\dot{q}_i$ so that

$$\begin{aligned} \mathcal{H} &= \sum_i \dot{q}_i (m\dot{q}_i) - \sum_i \frac{1}{2} m q_i^2 + V(q) \\ &= \sum_i \frac{1}{2} m q_i^2 + V(q) \\ &= \sum_i \frac{p_i^2}{2m} + V(q). \end{aligned} \tag{1.28}$$

**N.B. The Hamiltonian is a function of $q_i$ and $p_i$ and not $q_i$ and $\dot{q}_i$.**
The Hamiltonian is closely related to the energy of the system. While the dynamics of the Lagrangian system are described by a single point $(q)$ in an $n$-dimensional vector space called 'configuration space', the equivalent structure for Hamiltonian dynamics is the $2n$-dimensional 'phase space' where a single point is described by the vector $(q, p)$. This is a little more than cosmetics as the equations of motion describing the two systems differ. The Lagrangian has $n$ second order differential equations describing the motion, while the Hamiltonian system has $2n$ first order equations of motion. In both cases $2n$ boundary conditions are required to completely solve the equations of motion.

### 1.2.1 Hamilton's equations.

As $\mathcal{H} \equiv \mathcal{H}(q_i, p_i; t)$ then,

$$d\mathcal{H} = \sum_i \left( \frac{\partial \mathcal{H}}{\partial q_i} dq_i + \frac{\partial \mathcal{H}}{\partial p_i} dp_i + \frac{\partial \mathcal{H}}{\partial t} dt \right). \tag{1.29}$$

While as $\mathcal{H} = \sum_i \dot{q}_i p_i - \mathcal{L}$ we also have

$$\begin{aligned} d\mathcal{H} &= \sum_i \left( d\dot{q}_i p_i + \dot{q}_i dp_i - \frac{\partial \mathcal{L}}{\partial q_i} dq_i - \frac{\partial \mathcal{L}}{\partial \dot{q}_i} d\dot{q}_i - \frac{\partial \mathcal{L}}{\partial t} dt \right) \\ &= \sum_i \left( \dot{q}_i dp_i - \frac{\partial \mathcal{L}}{\partial q_i} dq_i - \frac{\partial \mathcal{L}}{\partial t} dt \right) \end{aligned} \tag{1.30}$$

where we have used the definition of the conjugate momentum $p_i = \frac{\partial \mathcal{L}}{\partial \dot{q}_i}$ to eliminate two terms in the final line. By comparing the coefficients of $dq_i$, $d\dot{q}_i$ and $dt$ in the two expressions for $d\mathcal{H}$ we find

$$\dot{q}_i = \frac{\partial \mathcal{H}}{\partial p_i}, \qquad \dot{p}_i = -\frac{\partial \mathcal{H}}{\partial q_i}, \qquad \frac{\partial \mathcal{H}}{\partial t} = -\frac{\partial \mathcal{L}}{\partial t} \tag{1.31}$$

where we have used Lagrange's equation 1.13 to observe that $\dot{p}_i = \frac{\partial \mathcal{L}}{\partial q_i}$. The first two of the above equations are usually referred to as Hamilton's equations of motion. Notice that these are $2n$ first order differential equations compared to Lagrange's equations which are $n$ second-order differential equations.

**Example.**

If

$$\mathcal{H} = \frac{p^2}{2m} + V(q) \tag{1.32}$$

then

$$\dot{q} = \frac{\partial \mathcal{H}}{\partial p} = \frac{p}{m} \qquad \text{and} \qquad \dot{p} = -\frac{\partial \mathcal{H}}{\partial q} = -\frac{\partial V}{\partial q}. \tag{1.33}$$

In other words we find, for this simple system, $p = m\dot{q}$ (the definition of linear momentum if $q$ is a Cartesian coordinate) and $F = -\frac{\partial V}{\partial q} = \dot{p}$ (Newton's second law).

### 1.2.2 Poisson Brackets

The Hamiltonian formulation of mechanics, while equivalent to the Lagrangian fomrulation, makes manifest a symmetry of the dynamical system. Notice that if we interchange $q_i$ and $p_i$ in Hamilton's equations, the two equations are interchanged up to a minus sign. This kind of skew-symmetry indicates that Hamiltonian dynamical systems possess a symplectic structure and the phase space is related to a symplectic manifold $Sp(2n)$ (see the group theory chapter for the definition of the symplectic group). There is, consequently, a useful skew-symmetric structure that exists on the phase space. It is called the Poisson bracket and is defined by

$$\{f, g\} \equiv \sum_i \left( \frac{\partial f}{\partial q_i} \frac{\partial g}{\partial p_i} - \frac{\partial g}{\partial q_i} \frac{\partial f}{\partial p_i} \right) \tag{1.34}$$

where $f = f(q_i, p_i)$ and $g = g(q_i, p_i)$ are abitrary functions on phase space.

One can write the equations of motion using the Poisson bracket as

$$\dot{q} = \{q_i, \mathcal{H}\} = \frac{\partial \mathcal{H}}{\partial p_i} \qquad \text{and} \qquad \dot{p} = \{p_i, \mathcal{H}\} = -\frac{\partial \mathcal{H}}{\partial q_i}. \tag{1.35}$$

Being curious pattern-spotters we may wonder whether it is generally the case that $\dot{f} \stackrel{?}{=} \{f, \mathcal{H}\}$ for an arbitrary function $f(q_i, p_i)$ on phase space. It is indeed the case as

$$\begin{aligned} \{f, \mathcal{H}\} &= \sum_i \left( \frac{\partial f}{\partial q_i}\frac{\partial \mathcal{H}}{\partial p_i} - \frac{\partial \mathcal{H}}{\partial q_i}\frac{\partial f}{\partial p_i} \right) \\ &= \sum_i \left( \frac{\partial f}{\partial q_i}\frac{dq_i}{dt} + \frac{dp_i}{dt}\frac{\partial f}{\partial p_i} \right) \\ &= \frac{df}{dt} \end{aligned} \tag{1.36}$$

if $f = f(q_i, p_i)$.

The set of Poisson brackets acting on simply $q_i$ and $p_j$ are known as the fundamental or canonical Poisson brackets. They have a simple form:

$$\begin{aligned} \{q_i, p_j\} &= \delta_{ij} \\ \{q_i, q_j\} &= 0 \\ \{p_i, p_j\} &= 0 \end{aligned} \tag{1.37}$$

which one may confirm by direct computation.

### 1.2.3 Duality and the Harmonic Oscillator

In string theory there are a number of surprising transformations called T-duality which leave the theory unchanged but give a new interpretation to the setting. By T-duality one observes that the theory is unchanged whether the fundamental distance is $R$ or $\frac{1}{R}$. This is a most unusual statement[3] which you will learn more about elsewhere. The prototype for duality transformations in a physical theory is the electromagnetic duality which we will look at briefly after we have discussed special relativity and tensor notation. The most simple duality transformation is exhibited in the harmonic oscillator. We have seen that the Lagrangian and Hamiltonian of the harmonic oscillator are

$$\begin{aligned} \mathcal{L} &= \frac{1}{2}m\dot{q}^2 - \frac{1}{2}kq^2 \\ \mathcal{H} &= \frac{p^2}{2m} + \frac{kq^2}{2}. \end{aligned} \tag{1.38}$$

The Hamilton equations are

$$\begin{aligned} \dot{q} = \frac{p}{m} \qquad \text{and} \qquad \dot{p} &= -kq \\ \Rightarrow \ddot{q} &= -\left(\frac{k}{m}\right)q \end{aligned} \tag{1.39}$$

and these have the solution

$$q = A\cos(\omega t) + B\sin(\omega t) \qquad \text{where} \qquad \omega = \sqrt{\frac{k}{m}}. \tag{1.40}$$

[3] If we were able to make such a transformation of the world we observe we would expect it to appear very different - if we survived

The solution is unchanged under the transformation

$$(m, k) \to (\frac{1}{k}, \frac{1}{m}) \tag{1.41}$$

as $\omega \to \sqrt{(\frac{1}{m})k} = \omega$. The transformation which we call a duality leaves the solution of the equations of motion unchanged. However the Lagrangian is transformed as

$$\mathcal{L} \to \mathcal{L}' = \frac{\dot{q}^2}{2k} - \frac{q^2}{2m} \tag{1.42}$$

and looks rather different. The Hamiltonian is transformed as

$$\mathcal{H} \to \mathcal{H}' = \frac{kp^2}{2} + \frac{q^2}{2m} \tag{1.43}$$

which up to a canonical transformation is identical to the original Hamiltonian $\mathcal{H}$. The precise canonical transformation is

$$\begin{aligned} q &\to q' = p \\ p &\to p' = -q \end{aligned} \tag{1.44}$$

which takes $\mathcal{H}' \to \mathcal{H}$. The transformation above is canonical as the Poisson brackets are preserved: $\{q', p'\} = \{p, -q\} = 1$. The Hamiltonian with dual parameters is canonically equivalent to the original Hamiltonian. Investigation of dualities can be rewarding, for example it is surprising to realise that the harmonic oscillator with large mass $m$ and large spring constant $k$ is equivalent to the same system with small mass $\frac{1}{k}$ and small spring constant $\frac{1}{m}$.

## 1.3 Noether's Theorem

**Theorem 1.3.1. (Noether)** *To every continuous symmetry of an action there is an associated conserved quantity.*

Let us denote the action by $S_R[q]$ where

$$S_R[q] \equiv \int_R dt \mathcal{L}(q, \dot{q}) \qquad \text{where} \qquad R = [t_1, t_2]. \tag{1.45}$$

There are two types of symmetry that we would like to consider,

(i.) Spatial: $S_R[q'] = S_R[q]$ and

(ii.) Space-time: $S_{R'}[q'] = S_R[q]$.

These two types foreshadow the symmetries that appear in field theory where an *internal* symmetry such as an $SO(n)$ scalar symmetry rotates the Lagrangian into itself, other types of symmetry of the action are called *external*. The spatial symmetries above are a symmetry of the Lagrangian alone and would be the prototype of an internal symmetry. We will consider Noether's theorem for a spatial symmetry first and find the associated conserved quantity (also called the conserved charge).

Let

$$q_i \to q_i' = q_i + \epsilon \chi_i(q) \equiv q_i + \delta q_i \tag{1.46}$$

then

$$\begin{aligned}\delta S_R &= S_R[q+\delta q] - S_R[q] + \mathcal{O}(\delta q)^2 \\ &= \int_R dt\mathcal{L}(q_i+\delta q_i, \dot{q}_i+\delta\dot{q}_i) - \int_R dt\mathcal{L}(q_i,\dot{q}_i) + \mathcal{O}(\delta q)^2\end{aligned} \tag{1.47}$$

Now,

$$\mathcal{L}(q_i+\delta q_i,\dot{q}_i+\delta\dot{q}_i) = \mathcal{L}(q_i,\dot{q}_i) + \sum_i\left(\delta q_i\frac{\partial\mathcal{L}}{\partial q_i} + \delta\dot{q}_i\frac{\partial\mathcal{L}}{\partial\dot{q}_i}\right) + \mathcal{O}(\delta q)^2 \tag{1.48}$$

so that

$$\begin{aligned}\delta S_R &= \int_R dt\sum_i\left(\delta q_i\frac{\partial\mathcal{L}}{\partial q_i} + \delta\dot{q}_i\frac{\partial\mathcal{L}}{\partial\dot{q}_i} + \mathcal{O}(\delta q_i)^2\right) \\ &= \int_R dt\sum_i\left(\delta q_i\left[\frac{\partial\mathcal{L}}{\partial q_i} - \frac{d}{dt}\left(\frac{\partial\mathcal{L}}{\partial\dot{q}_i}\right)\right]\right) + \int_R\sum_i\left(\frac{d}{dt}\left(\delta q_i\frac{\partial\mathcal{L}}{\partial\dot{q}_i}\right) + \mathcal{O}(\delta q_i)^2\right) \\ &= \int_R\left(\frac{d}{dt}\left(\sum_i\delta q_i\frac{\partial\mathcal{L}}{\partial\dot{q}_i}\right) + \sum_i\mathcal{O}(\delta q_i)^2\right)\end{aligned} \tag{1.49}$$

Where we have used Lagrange's equation to arrive at the final line. If the transformation $q_i \to q_i'$ is a symmetry then by definition $\delta S_R = 0$ up to terms of $\mathcal{O}(\delta q_i^2)$ and so

$$\frac{d}{dt}\left(\sum_i\delta q_i\frac{\partial\mathcal{L}}{\partial\dot{q}_i}\right) = 0 \tag{1.50}$$

Rewriting in terms of the generators $\chi_i$ we have that

$$\mathcal{Q} \equiv \sum_i\chi_i\frac{\partial\mathcal{L}}{\partial\dot{q}_i} \tag{1.51}$$

is a conserved quantity.

The space-time symmetries are chracterised by an additional time translation such that

$$\begin{aligned}q_i \to q_i' &= q_i + \epsilon\chi_i(q) \\ t \to t' &= t + \epsilon\xi(t).\end{aligned} \tag{1.52}$$

If the action is invariant under these transformations we have

$$S_{R'}[q_i'] = S_R[q_i] \tag{1.53}$$

where $R'$ is the image of the interval $R$ under the time translation. Consequently,

$$\begin{aligned}0 &= S_{R'}[q_i'] - S_R[q_i] \\ &= (S_{R'}[q_i'] - S_R[q_i']) + (S_R[q_i'] - S_R[q_i]) \\ &= (S_{R'}[q_i'] - S_R[q_i']) + \int_R dt\left(\frac{d}{dt}\left(\sum_i\epsilon\chi_i\frac{\partial\mathcal{L}}{\partial\dot{q}_i}\right)\right)\end{aligned} \tag{1.54}$$

where we have noted that the difference $(S_R[q_i'] - S_R[q_i])$ corresponds to an internal symmetry of the type we have computed above, hence we may immediately replace it with the conserved charge associated to the spatial symmetry. Now as

$$dt' = dt + \epsilon d\xi = dt + \epsilon\frac{d\xi}{dt}dt = dt(1+\epsilon\frac{d\xi}{dt}) \tag{1.55}$$

$$
\begin{aligned}
(S_{R'}[q_i'] - S_R[q_i']) &= \int_{R'} dt' \mathcal{L}(q'(t')) - \int_R dt \mathcal{L}(q'(t)) \qquad (1.56)\\
&= \int_R dt(1+\epsilon\frac{d\xi}{dt})(\mathcal{L}(q'(t) + \delta t \frac{d\mathcal{L}(q'(t))}{dt} + \mathcal{O}((\delta t)^2)) - \int_R dt \mathcal{L}(q'(t))\\
&= \int_R dt(\mathcal{L}(q'(t) + \delta t \frac{d\mathcal{L}(q'(t))}{dt} + \epsilon\frac{d\xi}{dt}\mathcal{L}(q'(t)) + \mathcal{O}((\delta t)^2))) - \int_R dt \mathcal{L}(q'(t))\\
&= \int_R dt(\mathcal{L}(q'(t))\epsilon\frac{d\xi}{dt} + \epsilon\xi\frac{d\mathcal{L}(q'(t))}{dt} + \mathcal{O}(\epsilon^2))\\
&= \int_R dt(\mathcal{L}(q(t))\epsilon\frac{d\xi}{dt} + \delta t\frac{d\mathcal{L}(q(t))}{dt} + \mathcal{O}(\epsilon^2))
\end{aligned}
$$

where $\delta t = \epsilon\xi$ and in the final line we have observed that $\epsilon\xi\frac{d}{dt}(\mathcal{L}(q'(t)) = \epsilon\xi\frac{d}{dt}(\mathcal{L}(q(t)) + \mathcal{O}(\epsilon^2)$. Recalling that $\delta t \equiv \epsilon\xi$ we have

$$
\begin{aligned}
(S_{R'}[q_i'] - S_R[q_i']) &= \int_R dt(\mathcal{L}\epsilon\frac{d\xi}{dt} + \epsilon\xi\frac{d\mathcal{L}}{dt} + \mathcal{O}(\epsilon^2)) \qquad (1.57)\\
&= \int_R dt(\frac{d}{dt}\Big(\mathcal{L}\epsilon\xi\Big) + \mathcal{O}(\epsilon^2))
\end{aligned}
$$

Including the spatial symmetry we have in all,

$$
\begin{aligned}
0 &= S_{R'}[q_i'] - S_R[q_i] \qquad (1.58)\\
&= \int_R dt\Big(\frac{d}{dt}\Big(\mathcal{L}\epsilon\xi + \sum_i \epsilon\chi_i\frac{\partial\mathcal{L}}{\partial\dot{q}_i}\Big) + \mathcal{O}(\epsilon^2)\Big)
\end{aligned}
$$

Neglecting the higher order terms in $\epsilon$ we deduce that

$$
Q \equiv \mathcal{L}\xi + \sum_i \chi_i\frac{\partial\mathcal{L}}{\partial\dot{q}_i} \qquad (1.59)
$$

is a conserved quantity under the space-time transformation symmetry of the action.

**Example 1**

Suppose that the spatial translation given by

$$
q_i \to q_i' = q_i + \epsilon a_i \qquad (1.60)
$$

where $a_i$ is a constant shift in the i'th generalised coordinate is a symmetry of the action. Then we see that the conserved charge is

$$
Q = \sum_i a_i\frac{\partial\mathcal{L}}{\partial\dot{q}_i} = \sum_i a_i p_i \qquad (1.61)
$$

where $p_i$ are the generalised momenta. The conserved quantity is a linear sum of the generalised momenta which are all independently conserved.

**Example 2**

Suppose that the temporal translation is a symmetry of the action. Let the translation be

$$
t \to t' = t + \epsilon b \qquad (1.62)
$$

where $b$ is a constant. Let us isolate the temporal shift from its associated spatial shift, i.e. as $q(t) \to q(t') = q(t + \epsilon b)$, by simultaneously transforming the coordinates as

$$q \to q' = q(t - \epsilon b) = q(t) - \epsilon b \dot{q}(t) + \mathcal{O}(\epsilon^2). \tag{1.63}$$

Then we see that the conserved charge is

$$Q = b\mathcal{L} - b\sum_i \dot{q}_i \frac{\partial \mathcal{L}}{\partial \dot{q}_i} = -b(\sum_i \dot{q}_i p_i - \mathcal{L}) = -bE \tag{1.64}$$

where $E \equiv \sum_i \dot{q}_i p_i - \mathcal{L}$ is the energy function, the precursor to the Hamiltonian. The conserved quantity proportional to the energy. If $b = -1$ the energy is the conserved quantity.

**Example 3**

Consider the Lagrangian for the simple harmonic oscillator in two dimensions:

$$\mathcal{L} = \frac{m}{2}(\dot{x}^2 + \dot{y}^2) - \frac{k}{2}(x^2 + y^2). \tag{1.65}$$

Let's make a change of coordinates in order to make manifest the rotation symmetry of the system, let $z = x + iy$, so that

$$\mathcal{L} = \frac{m}{2}\dot{z}\dot{\bar{z}} - \frac{k}{2}z\bar{z}. \tag{1.66}$$

The rotation in the complex plane is a symmetry of the action. Let

$$z \to z' = e^{i\omega}z = z + i\omega z + \mathcal{O}(\omega^2) \tag{1.67}$$

so that,

$$\begin{aligned} z\bar{z} \to z'\bar{z'} &= e^{i\omega}ze^{-i\omega}\bar{z} = z\bar{z} \\ \dot{z}\dot{\bar{z}} \to \dot{z'}\dot{\bar{z'}} &= e^{i\omega}\dot{z}e^{-i\omega}\dot{\bar{z}} = \dot{z}\dot{\bar{z}} \end{aligned} \tag{1.68}$$

and evidently $\mathcal{L} \to \mathcal{L}$. The infinitesimal transformations of $z$ and $\bar{z}$ are given by

$$\delta z = i\omega z \qquad \text{and} \qquad \delta\bar{z} = -i\omega\bar{z} \tag{1.69}$$

and hence the conserved charge is

$$Q = iz\frac{\partial \mathcal{L}}{\partial \dot{z}} - i\bar{z}\frac{\partial \mathcal{L}}{\partial \dot{\bar{z}}} = \frac{im}{2}(z\dot{\bar{z}} - \bar{z}\dot{z}). \tag{1.70}$$

One can use the equations of motion to check that this quantity is conserved. The equations of motion give:

$$\begin{aligned} \frac{m}{2}\ddot{\bar{z}} &= -\frac{k}{2}\bar{z} \\ \frac{m}{2}\ddot{z} &= -\frac{k}{2}z \end{aligned} \tag{1.71}$$

and hence,

$$\begin{aligned} \frac{dQ}{dt} &= \frac{im}{2}(\dot{z}\dot{\bar{z}} + z\ddot{\bar{z}} - \dot{\bar{z}}\dot{z} - \bar{z}\ddot{z}) \\ &= \frac{im}{2}(z\ddot{\bar{z}} - \bar{z}\ddot{z}) \\ &= -\frac{ik}{2}(z\bar{z} - \bar{z}z) \\ &= 0 \end{aligned} \tag{1.72}$$

### 1.3.1 A Sideways Glance at Noether's Theorem.

It is worth our while thinking again about Noether's theorem. First let us recall that the Lagrangian for a dynamical system is only defined up to a total time derivative $\frac{dJ(q)}{dt}$, as if one transforms the Lagrangian in the following manner

$$\mathcal{L} \to \mathcal{L}' = \mathcal{L} + \frac{dJ(q)}{dt} \tag{1.73}$$

then the equations of motion remain unchanged

$$\begin{aligned}
\frac{d}{dt}\left(\frac{\partial \mathcal{L}'}{\partial \dot{q}_i}\right) - \frac{\partial \mathcal{L}'}{\partial q_i} &= \frac{d}{dt}\left(\frac{\partial}{\partial \dot{q}_i}\left(\mathcal{L} + \frac{dJ(q)}{dt}\right)\right) - \frac{\partial}{\partial q_i}\left(\mathcal{L} + \frac{dJ(q)}{dt}\right) \\
&= \underbrace{\frac{d}{dt}\left(\frac{\partial \mathcal{L}}{\partial \dot{q}_i}\right) - \frac{\partial \mathcal{L}}{\partial q_i}}_{=0} + \frac{d}{dt}\left(\frac{\partial}{\partial \dot{q}_i}\left(\frac{dJ(q)}{dt}\right)\right) - \frac{\partial}{\partial q_i}\left(\frac{dJ(q)}{dt}\right) \\
&= \frac{d}{dt}\left(\frac{\partial}{\partial \dot{q}_i}\left(\frac{\partial J(q)}{\partial q_i}\dot{q}_i\right)\right) - \frac{\partial}{\partial q_i}\left(\frac{dJ(q)}{dt}\right) \\
&= \frac{d}{dt}\left(\frac{\partial J(q)}{\partial q_i}\right) - \frac{\partial}{\partial q_i}\left(\frac{dJ(q)}{dt}\right) \\
&= 0.
\end{aligned} \tag{1.74}$$

It is quicker to make this same observation directly using the action:

$$S'_R = \int_R dt\left(\mathcal{L} + \frac{dJ}{dt}\right) = \int_R dt\mathcal{L} + \underbrace{J(t_2) - J(t_1)}_{\text{constant}} = S_R + \text{constant} \tag{1.75}$$

hence $\delta A'_R = \delta A_R$ under the transformation $\mathcal{L} \to \mathcal{L}'$. Returning to Noether's theorem let us suppose that we can find a transformation of the coordinates and velocities such that $\mathcal{L}' = \mathcal{L} + \frac{dJ}{dt} \equiv \mathcal{L} + \delta\mathcal{L}$. In other words we are considering transformations that leave the equations of motion invariant but which might[4] result in a change in the Lagrangian. Then we would have at first order that

$$\begin{aligned}
\delta\mathcal{L} &= \frac{\partial \mathcal{L}}{\partial q_i}\delta q_i + \frac{\partial \mathcal{L}}{\partial \dot{q}_i}\delta \dot{q}_i \\
&= \frac{\partial \mathcal{L}}{\partial q_i}\delta q_i + \frac{d}{dt}\left(\frac{\partial \mathcal{L}}{\partial \dot{q}_i}\delta q_i\right) - \frac{d}{dt}\left(\frac{\partial \mathcal{L}}{\partial \dot{q}_i}\right)\delta q_i \\
&= \frac{d}{dt}\left(\frac{\partial \mathcal{L}}{\partial \dot{q}_i}\delta q_i\right) \\
&= \frac{dJ}{dt}.
\end{aligned} \tag{1.76}$$

Hence we find there exists a conserved quantity as

$$\frac{d}{dt}\left(\frac{\partial \mathcal{L}}{\partial \dot{q}_i}\delta q_i - J\right) = 0. \tag{1.77}$$

If we follow the standard definition that a symmetry implies that the first order change in the Lagrangian vanishes $\delta\mathcal{L}$ then by construction $J$ is a constant - the Noether charge is just the part identified earlier as coming from a spatial symmetry. If we allow our

[4] We err on the side of caution as we note that if $J$ is constant the total time derivative change in the Lagrangian is zero.

definition to be loosened to: "a symmetry of a dynamical system leaves the equations of motion invariant" then we allow the possibility of a less-trivial $J$ and a manifest first order change in $\mathcal{L}$. This is sufficient to derive the conservation of energy from a time translation of the Lagrangian alone, without recourse to the action. A dynamical time translation transforms

$$\begin{aligned} q_i &\to q_i' = q_i + \epsilon \dot{q}_i \\ \dot{q}_i &\to \dot{q}_i' = \dot{q}_i + \epsilon \ddot{q}_i \end{aligned} \tag{1.78}$$

to first order in $\epsilon$ where $\epsilon = \delta t$. Hence

$$\mathcal{L}' = \mathcal{L} + \frac{\partial \mathcal{L}}{\partial q_i}\epsilon \dot{q}_i + \frac{\partial \mathcal{L}}{\partial \dot{q}_i}\epsilon \ddot{q}_i + \mathcal{O}(\epsilon^2) = \mathcal{L} + \epsilon \frac{d\mathcal{L}}{dt} + \mathcal{O}(\epsilon^2) \tag{1.79}$$

and the first order change in the Lagrangian does not vanish. Yet there is an associated conserved charge as the change is a total time derivative so we have $J = \mathcal{L}$ and the conserved charge is

$$\mathcal{Q} = \sum_i \dot{q}_i \frac{\partial \mathcal{L}}{\dot{q}_i} - \mathcal{L} = H \tag{1.80}$$

which is the energy function, or Hamiltonian.

### 1.3.2 Noether's theorem in the Hamiltonian formulation.

Canonical transformations ($q_i \to q_i'$, $p_i \to p_i'$) are those transformations which preserve the form of the equations of motion written in the transformed variables, i.e. under a canonical transformation the equations of motion are transformed into

$$\dot{q}_i' = \frac{\partial \mathcal{H}(q_i'.p_i')}{\partial p_i'} \qquad \text{and} \qquad \dot{p}_i' = -\frac{\partial \mathcal{H}(q_i'.p_i')}{\partial q_i'}. \tag{1.81}$$

A necessary and sufficient condition for a transformation to be canonical is that the fundamental Poisson brackets are preserved under the transformation, i.e.

$$\{q_i', p_j'\} = \delta_{ij}, \quad \{q_i', q_j'\} = 0 \quad \text{and} \qquad \{p_i', p_j'\} = 0. \tag{1.82}$$

Consequently a canonical transformation may be generated by an arbitrary function $f(q_i, p_i)$ on phase space via

$$\begin{aligned} q_i &\to q_i' = q_i + \alpha\{q_i, f\} \equiv q_i + \delta q_i \\ p_i &\to p_i' = p_i + \alpha\{p_i, f\} \equiv p_i + \delta p_i \end{aligned} \tag{1.83}$$

and if $\alpha \ll 1$ then the transformation is an infinitesimal canonical transformation. It is easy to check that this preserves the fundamental Poisson brackets up to terms of order $\mathcal{O}(\alpha^2)$, e.g.

$$\begin{aligned} \{q_i', p_j'\} &= \{q_i + \alpha\{q_i, f\}, p_j + \alpha\{p_j, f\}\} \\ &= \{q_i, p_j\} + \alpha(\{\{q_i, f\}, p_j\} + \{q_i, \alpha\{p_j, f\}\}) + \mathcal{O}(\alpha^2) \\ &= \{q_i, p_j\} + \alpha(\{\frac{\partial f}{\partial p_i}, p_j\} + \{q_i, -\frac{\partial f}{\partial q_j}\}) + \mathcal{O}(\alpha^2) \\ &= \delta_{ij} + \alpha(\frac{\partial^2 f}{\partial q_j \partial p_i} - \frac{\partial^2 f}{\partial p_i \partial q_j}) + \mathcal{O}(\alpha^2) \\ &= \delta_{ij} + \mathcal{O}(\alpha^2). \end{aligned} \tag{1.84}$$

If an infinitesimal canonical transformation is a symmetry of the Hamiltonian then $\delta\mathcal{H} = 0$ under the transformation. Now,

$$\begin{aligned}\delta\mathcal{H} &= \sum_i \left(\frac{\partial\mathcal{H}}{\partial q_i}\delta q_i + \frac{\partial\mathcal{H}}{\partial p_i}\delta p_i\right) \\ &= \alpha\sum_i \left(\frac{\partial\mathcal{H}}{\partial q_i}\frac{\partial f}{\partial p_i} - \frac{\partial\mathcal{H}}{\partial p_i}\frac{\partial f}{\partial q_i}\right) \\ &= \alpha\{\mathcal{H}, f\} \\ &= -\alpha\frac{df}{dt}\end{aligned} \tag{1.85}$$

where we have assumed that $f$ is an explicit function of the phase space variables and not time, i.e. $\frac{\partial f}{\partial t} = 0$. Hence if the transformation is a symmetry $\delta\mathcal{H} = 0$ then $f(q_i, p_i)$ is a conserved quantity.

**Problem 1.3.1.** *The Lagrangian of non-relativistic particle with mass $m$ and charge $e$ propagating in a manifold $\mathcal{M}$ with metric $ds^2 = g_{\mu\nu}dx^\mu dx^\nu$ and coupled to a magnetic gauge potential $A_\mu(x)$ is*

$$\mathcal{L} = \frac{m}{2}g_{\mu\nu}(x)\dot{x}^\mu\dot{x}^\nu + eA_\mu(x)\dot{x}^\mu$$

*where $x^\mu$ are the coordinates of the manifold, $\mu = 1, ..., dim(\mathcal{M})$ and $\dot{x}^\mu$ is the time derivative of the coordinate $x^\mu$.*

*(i.) Give the equations of motion in the Lagrangian formalism. In particular express the equations of motion in terms of the Levi-Civita connection.*

*(ii.) Find the canonical momentum and the Hamiltonian of the theory.*

*(iii.) Noether's charge for general (space)time symmetries is*

$$Q = \xi\mathcal{L} + X^\mu(x)\frac{\partial\mathcal{L}}{\partial\dot{x}^\mu}$$

*Specialise this to time translations and use the invariance of the action under time translations to calculate the energy $E$ of the charged particle. Is $E$ expressed in terms of the velocities dependent on $A$? Verify that $E$ is conserved subject to the equations of motion.*

**Problem 1.3.2.** *The Lagrangian for a two-dimensional harmonic oscillator is*

$$\mathcal{L} = \frac{m}{2}(\dot{x}^2 + \dot{y}^2) - \frac{k}{2}(x^2 + y^2)$$

*where $x$ and $y$ are Cartesian coordinates, $\dot{x}$ and $\dot{y}$ are their time-derivatives, $m$ is the mass of the oscillator and $k$ is a constant.*

*(a.) Rewrite the Lagrangian in terms of the complex coordinate $z = x + iy$, its complex conjugate $\bar{z}$ and their time-derivatives.*

*(b.) Show that*

$$z \to z' = e^{i\omega}z = z + i\omega z + \mathcal{O}(\omega^2)$$

*is a symmetry of the Lagrangian.*

*(c.) Prove Noether's theorem for an action $A_R[q_i]$ where*

$$A_R[q_i] \equiv \int_R dt\mathcal{L}(q_i, \dot{q}_i) \qquad \text{with} \quad R = [t_1, t_2]$$

*which is invariant under the transformation: $q_i \to q_i' = q_i + \epsilon X_i(q_i)$ by showing that $X_i \frac{\partial \mathcal{L}}{\partial \dot{q}_i}$ is a conserved quantity for the transformation. (N.B. the Einstein summation convention is used for the repeated $i$ index.)*

*(d.) Consider the infinitesimal version of the transformation given in part (b.) so that $\delta z = i\omega z$. Find the conserved quantity $Q$ associated to this transformation and use the equations of motion to prove directly that its time-derivative $\frac{dQ}{dt}$ is zero.*

**Problem 1.3.3.** *The Lagrangian of a relativistic particle with mass $m$ and charge $e$ and coupled to an electromagnetic field is*

$$\mathcal{L} = -\frac{mc^2}{\gamma} - e\phi(\mathbf{x}, t) + \sum_i eA_i(\mathbf{x}, t)\dot{x}^i$$

*where $x^i$ are the coordinates of the particle with $i = 1, 2, 3$, $\gamma = (1 - \frac{\dot{x}^2}{c^2})^{-\frac{1}{2}}$, $\dot{x}^i$ is the time derivative of the coordinate $x^i$, $\phi(\mathbf{x}, t)$ is the electric scalar potential and $\mathbf{A}(\mathbf{x}, t)$ is the magnetic vector potential.*

*(a.) Show that the equations of motion may be written in vector form as*

$$\frac{d}{dt}\left(m\gamma\dot{\mathbf{x}}\right) = -e\frac{\partial \mathbf{A}}{\partial t} - e\nabla\phi + \dot{\mathbf{x}} \wedge \nabla \wedge \mathbf{A}.$$

*(b.) Find the Hamiltonian of the system.*

*(c.) Show that the rest energy of the system (i.e. when $\mathbf{p} = \mathbf{0}$) is*

$$mc^2 + \frac{1}{2}\frac{e^2}{m}\mathbf{A}^2 + e\phi + \mathcal{O}(\frac{1}{c^2}).$$

**Problem 1.3.4.** *A conformal rescaling of the metric acts as $g'_{\mu\nu} = e^{\alpha} g_{\mu\nu}$ where $\alpha \in \mathbb{R}$.*

*(a.) Consider the change of coordinates given by $x^\mu \to x'^\mu = x^\mu + \epsilon^\mu(x)$ and which transforms the metric as*

$$g_{\mu\nu} = \frac{\partial x'^\rho}{\partial x^\mu}\frac{\partial x'^\sigma}{\partial x^\nu} g'_{\rho\sigma}.$$

*Use the above transformation of the metric to show that if $x^\mu \to x'^\mu = x^\mu + \epsilon^\mu(x)$ generates a conformal transformation of the metric then*

$$g_{\mu\nu} = (g_{\mu\nu} + \partial_\mu\epsilon_\nu + \partial_\nu\epsilon_\mu)e^{\alpha} + \mathcal{O}(\epsilon^2).$$

*(b.) Hence show that*

$$(e^{-\alpha} - 1)D = 2\partial_\mu\epsilon^\mu$$

*where $D$ is the dimension of spacetime.*

*(c.) Use the expression in part (c.) to eliminate $e^\alpha$ in the final expression of part (a.) to show that*

$$\frac{2}{D}\partial_\kappa \epsilon^\kappa g_{\mu\nu} = \partial_\mu \epsilon_\nu + \partial_\nu \epsilon_\mu.$$

*Act on this expression with $\partial^\mu \partial^\nu$ to obtain:*

$$\left(\frac{2}{D} - 2\right)\partial^\kappa \partial_\kappa (\partial_\lambda \epsilon^\lambda) = 0$$

*allowng the observation that $\epsilon^\mu$ is at most a quadratic function in $x$ for a conformal transformation.*

*Consider the action for a massless, real scalar field $\phi$ with a quartic potential in Minkowksi space-time:*

$$S = \int d^4\mathbf{x}\mathcal{L} = \int d^4\mathbf{x}\left(\frac{1}{2}\partial_\mu\phi\partial^\mu\phi - \lambda\phi^4\right)$$

*where $\lambda \in \mathbb{R}$ is a constant. Under a conformal transformation the field transforms as $\phi \to \phi' \equiv \phi + \kappa x^\mu \partial_\mu \phi + \kappa\phi$ where $\kappa$ is the infinitesimal parameter for the transformation.*

*(d.) Show that the variatation of the Lagrangian under the conformal transformation is given by (upto order $\kappa^2$):*

$$\mathcal{L} \to \mathcal{L} + \kappa\partial_\mu(x^\mu\mathcal{L}).$$

*(e.) Hence show that there is an associated conserved quantity*

$$j^\mu \equiv \partial_\mu\phi(x^\nu\partial_\nu\phi + \phi) - x^\mu\mathcal{L}.$$

*(f.) Find the equation of motion for $\phi$ and use this to show explicitly that $\partial_\mu j^\mu = 0$.*

# Chapter 2

# Special Relativity and Component Notation

In 1905 Einstein published four papers which each changed the world. In the first he established that energy occurs in discrete quanta, which since the work of Max Planck had been thought to be a property of the energy transfer mechanism rather than energy itself - this work really opened the door for the development of quantum mechanics. In his second paper Einstein used an analysis of brownian motion to establish the physical existence of atoms. In his third and fourth papers he set out the special theory of relativity and derived the most famous equation in physics, if not mathematics, relating energy to rest mass $E = mc^2$. Hence 1905 is often referred to as Einstein's annus mirabilis.

At the time Einstein had been refused a number of academic positions and was working in the patent office in Bern. He was living with his wife and two young children while he was writing these historic papers. Not only was he insightful but perhaps, more importantly, he was dedicated and industrious. He must also have been pretty tired too. In 1921 Einstein was awarded the Nobel prize for his work on the photoelectric effect (the work in the first of his four papers that year) but special relativity was overlooked (partly because it was very difficult to verify its predictions accurately at the time). If there is any message to be taken from the decision of the Nobel committee it is probably that you should keep your own counsel with regard to the quality of your work.

In this chapter we will give a brief description of the special theory of relativity - a more complete description of the theory will require group theory and will be covered again the group theory chapter. One consequence of relativity is that time and space are put on equal footing and we will need to develop the notation we have used for classical mechanics in which time was a special variable. Consequently we will spend some time developing our notation and will also consider the component notation for tensors. Sometimes a good notation is as good as a new idea.

## 2.1 The Special Theory of Relativity

The theory was constructed on two simple postulates:

(1.) the laws of physics are independent of the inertial reference frame of the observer,

and

(2.) the speed of light is a constant for all observers.

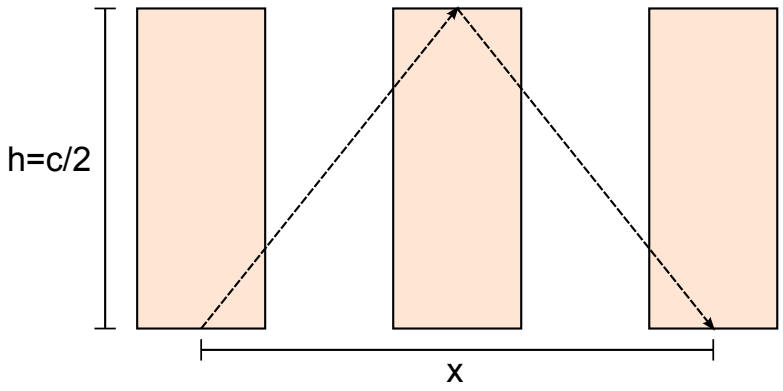
Surprisingly these simple postulates necessitated that coordinate and time transformations between two different frames $F$ and $F'$ moving at relative speed $v$ in the $x$-direction were no longer the Gallilean transformation but rather the Lorentz transformations:

$$\begin{aligned} t' &= \gamma(t - \frac{xv}{c^2}) \\ x' &= \gamma(x - vt) \\ y' &= y \\ z' &= z \end{aligned} \tag{2.1}$$

where

$$\gamma \equiv \left(\sqrt{1 - \frac{v^2}{c^2}}\right)^{-1}. \tag{2.2}$$

Let us consider two thought experiments to motivate these transformations, the first will demonstrate time dilation and the second the shortening of length. Consider a clock formed of two perfect mirrors separated vertically such that a photon bouncing between the mirrors takes one second to travel from the bottom mirror to the top mirror and back again. It is consequently a very tall clock, it has height $h = \frac{c}{2}$ metres where $c$ is the speed of light (hence $h \approx \frac{299792458}{2} = 149,896,229$ metres in a vacuum!). Let us set the clock in motion with a speed $v$ in the $+x$-direction and consider two observers: one in the rest frame of the clock $F'$ and a second in a frame $F$ and a second observer in frame $F'$ which moves at speed $v$ along the x-axis. Suppose at time $t = 0$ the two clocks are at the origin of frame $F$ (i.e. the origin of both frames $F$ and $F'$ coincide at $t = 0$). As the observer at the origin of frame $F'$ moves off at speed $v$ the observer in frame $F$ observes the "ticking" of the relatively moving photon clock slow down. Schematically we indicate a view of the moving clock as seen from frame $F'$ below:




The photon in the moving clock now is seen to move along the hypotenuse of a right-angled triangle as the clock moves horizontally. What are the dimensions of this triangle as seen from frame $F'$? The height is the same as the clock at $\frac{c}{2}$. As viewed from the frame $F'$ where the clock appears to be moving $t'$ seconds are observed to pass, in which time the clock's base has moved a distance $vt'$. Now using the Pythagorean formula and the first postulate of special relativity (that the speed of light is a constant) we find that

the photon travels a distance $d = ct'$ where

$$ct' = 2(\sqrt{\frac{c^2}{4} + \frac{v^2 t'^2}{4}}) = \sqrt{c^2 + v^2 t'^2}. \tag{2.3}$$

Rearranging we find that, after one second has passed as measured in the rest frame of the clock, that $t'$ seconds have passed as viewed from the frame $F'$ in which the clock is moving and

$$1 = \left(\sqrt{1 - \frac{v^2}{c^2}}\right) t' = \frac{1}{\gamma} t'. \tag{2.4}$$

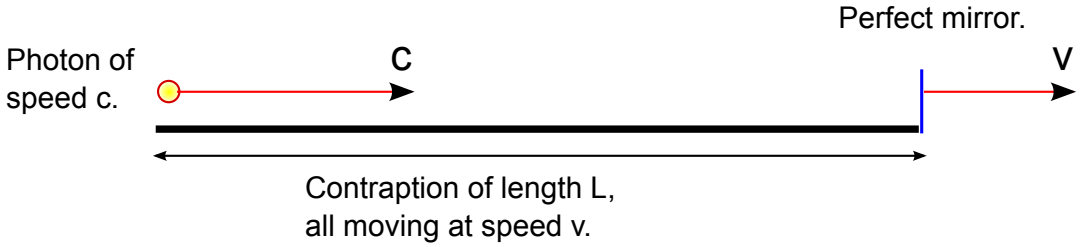
We deduce that after $t$ oscillations of the moving photon clock

$$ct' = \sqrt{c^2 t^2 + v^2 t'^2} \quad \Rightarrow \quad t' = \gamma t. \tag{2.5}$$

As $\gamma \geq 1$ the time measured on a moving clock has slowed. This derivation of time dilation is only a toy model as we assumed we could instantaneously know when the photon on the moving clock had completed its oscillation. In practise the observer would sit at the origin of frame $F'$ and record measurements from there, information would take time to be transported back to their frame's origin and a second property of special relativity would need to be considered, that of length contraction.

Let us consider a second toy model that will indicate length contraction as a consequence of the postulates of special relativity.

Suppose we construct a contraption, consisting of a straight rigid rod with a perfect mirror attached to one end (as drawn below), whose rest length is $L$. We will aim to measure its length using a photon, whose arrival and departure time we will suppose we can measure accurately. The experiment will involve the photon traversing the length of the rod, being reflected by the perfect mirror and returning to its starting point. When conducted at rest the photon returns to its starting point in time $t = \frac{2L}{c}$. Now we will change frames so that in $F'$ the contraption is seen to be moving with speed $v$ in the positive $x$ direction (left-to-right horizontally across the page as drawn below) and repeat the experiment.




Now we know that on the first leg of the journey the photon will take a longer time to reach the mirror, as the mirror is travelling away from the photon. However on the return leg the photon's starting point at the other end of our contraption is moving towards the photon. So we may wonder if the total journey time for the photon has changed overall. We compute the time taken for each of the two legs $t_1$ and $t_2$, and note that viewed from the stationary frame the contraption has length $L$:

$$ct'_1 = L + vt'_1 \Rightarrow t'_1 = \frac{L}{c - v} \tag{2.6}$$

$$ct'_2 = L - vt'_2 \Rightarrow t'_2 = \frac{L}{c + v} \tag{2.7}$$

So the total time taken for the photon to traverse twice the contraption length when it is moving at speed $v$ is

$$\frac{c}{2}(t_1' + t_2') = \frac{c}{2}\left(\frac{L}{c-v} + \frac{L}{c+v}\right) = \frac{Lc^2}{c^2 - v^2} = L\gamma^2 \tag{2.8}$$

Meanwhile using the Lorentz transformations for time between frames we have

$$L_0 = \frac{c}{2}(t_1 + t_2) = \frac{1}{\gamma}\frac{c}{2}(t_1' + t_2') = L\gamma. \tag{2.9}$$

As $\gamma \geq 1$ then $L \geq L_0$. Length appears to have contracted in the moving frame. Denoting $L_0 = x_2' - x_1'$ and $L = x_2 - x_1$ implies that $x' = \gamma x$ (for a measurement where $dt = 0$).

Let us complete this thought experiment by bringing together time dilation and length contraction to find the Lorentz transformations given in equation (2.1). Consider an event occuring in the stationary event at spacetime point $(t, x)$[1] The event is the arrival taken of a photon having started at the origin at $t = 0$, i.e. $x = ct$. Observing the same motion of a photon in the moving frame we deduce (as for the first leg in the thought experiment used to derive length contraction):

$$x' + vt' = ct' \quad \Rightarrow \quad x' = (c - v)t' \tag{2.10}$$

Using the time dilation $t' = \gamma t$ gives

$$x' = \gamma(ct - vt) = \gamma(x - vt) \tag{2.11}$$

(as $x = ct$). As the speed of light is unchanged in either frame we have $\frac{x}{t} = \frac{x'}{t'}$, and using equation (2.11) we have

$$t' = x'\frac{t}{x} = \gamma(x - vt)\frac{t}{x} = \gamma(t - \frac{vt^2}{x}) = \gamma(t - \frac{vx}{c^2}) \tag{2.12}$$

where we have used $t = \frac{x}{c}$ which is valid for photon motion. Thus we have arrived at the Lorentz transformations of equation (2.1).

These simple thought experiments changed the world and demonstrate the possibility for thought alone to outstrip intuition and experiment.

### 2.1.1 The Lorentz Group and the Minkowski Inner Product.

As we will see in the chapter on group theory, the Lorentz transformations form a group denoted $O(1,3)$. The subgroup of proper Lorentz transformations has determinant one and is denoted $SO(1,3)$. When the Lorentz transformations are combined with the translations in space and time the new larger group formed is called the Poincare group. It is the relativistic analogue of the Gallilean group which map between inertial frames in Newtonian mechanics[2]. The Lorentz group $O(1,3)$ is defined by

$$O(1,3) \equiv \{\Lambda \in GL(4,\mathbb{R}) | \Lambda^T \eta \Lambda = \eta; \eta \equiv diag(1,-1,-1,-1)\}$$

[1] We suppress the $y$ and $z$ cooridinates as they are unchanged for a Lorentz transformation in the $x$-direction only.

[2] The Gallilean group consists of 10 transformations: 3 space rotations, 3 space translations, 3 Gallilean velocity boosts $v \to v + u$ and one time tranlsation.

$GL(4,\mathbb{R})$ is the set of invertible four-by-four matrices whose entries are elements of $\mathbb{R}$, $\Lambda^T$ is the transpose of the matrix $\Lambda$ and $\eta$, the Minkowski metric, is a four-by-four matrix whose diagonal elements are non-zero and given in full matrix notation by

$$\eta \equiv \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}. \tag{2.13}$$

It is not yet obvious that either the Lorentz transformations do form a group nor that the definition of $O(1,3)$ encodes the Lorentz transformations as given in section 2.1. We will wait until we encounter the definition of a group before checking the first assertion. The group $SO(1,3)$ itself is the rotation group in a Minkowski space the numbers $(1,3)$ indicate the signature of the spacetime and corresponds to a spacetime with one timelike coordinate and three spatial coordinates or $\mathbb{R}^{1,3}$. Rather more mathematically the matrix $\eta$ defines the signature of the Minkowski metric[3] which is preserved by the Lorentz transformations. It is the insightful observation that the Lorentz transformations leave invariant the Minkowski inner product between two four vectors that will give the first hint that Lorentz transformations are related to the definition of $O(1,3)$. The equivalent statement in Euclidean space $\mathbb{R}^3$ is that rotations leave distances unchanged. The inner product on $\mathbb{R}^{1,3}$ is defined between any two four-vectors

$$\mathbf{v} = \begin{pmatrix} v^0 \\ v^1 \\ v^2 \\ v^3 \end{pmatrix} \qquad \text{and} \qquad \mathbf{w} = \begin{pmatrix} w^0 \\ w^1 \\ w^2 \\ w^3 \end{pmatrix} \tag{2.14}$$

in $\mathbb{R}^{1,3}$ by

$$< \mathbf{v}, \mathbf{w} > \equiv \mathbf{v}^T \eta \mathbf{w} \tag{2.15}$$

$$= (v^0, v^1, v^2, v^3) \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} \begin{pmatrix} w^0 \\ w^1 \\ w^2 \\ w^3 \end{pmatrix} \tag{2.16}$$

$$= v^0 w^0 - v^1 w^1 - v^2 w^2 - v^3 w^3. \tag{2.17}$$

Now we can see clearly that the Minkowski inner product $< \mathbf{v}, \mathbf{w} >$ is not positive for all vectors $\mathbf{v}$ and $\mathbf{w}$.

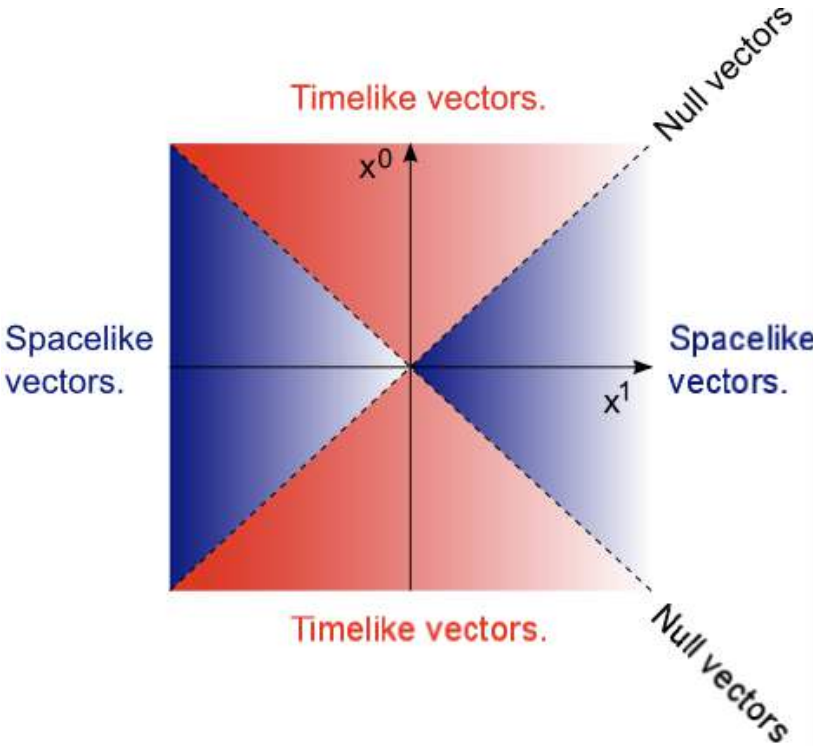
**Problem 2.1.1.** *Show that under the Lorentz transformation* $\mathbf{x}^2 \equiv x^\mu x^\nu \eta_{\mu\nu}$ *is invariant, where* $x^0 = ct$, $x^1 = x$, $x^2 = y$ *and* $x^3 = z$.

[3]We commence the abuse of our familiar mathematical definitions here as the Minkowski metric is not positive definite as is implied by the definition of a metric, similarly the Minkowski inner product is also not positive definite but the constructions of both Minkowski inner product and Minkowski metric are close enough to the standard definitons that the misnomers have remained, and the lack of vocabulary will not confuse our work. Properly Minkowski space is a pseudo-Riemannian manifold in contrast to Euclidean space equipped with the standard metric which is a Riemannian manifold.

It is worthwhile keeping the comparison with $\mathbb{R}^3$ in mind. The equivalent group would be $SO(3)$ and its elements are the rotations in three-dimensional space, the inner product on the space is defined using the identity matrix $I$ whose diagonal entries are all one and whose off-diagonal entries are zero. The Euclidean inner product on $\mathbb{R}^3$ between two vectors $\mathbf{x}$ and $\mathbf{y}$ is $\mathbf{x}^t I\mathbf{y} \equiv x^1y^1 + x^2y^2 + x^3y^3$. The vector length squared $\mathbf{x}^2 = \mathbf{x}^T I\mathbf{x} \equiv \mathbf{x}\cdot\mathbf{x}$ is positive definite when $\mathbf{x} \neq \mathbf{0}$. The rotation of a vector leaves invariant the length of any vector in the space, or in other words leaves the inner product invariant. In the comparison with Lorentz transformations in Minkowski space the crucial difference is that the metric is no longer positive definite and hence four-vectors fall into one of three classes:

$$< \mathbf{v}, \mathbf{v} > \begin{cases} > 0 & \mathbf{v} \text{ is called timelike} \\ = 0 & \mathbf{v} \text{ is called lightlike or null} \\ < 0 & \mathbf{v} \text{ is called spacelike} \end{cases} . \tag{2.18}$$

Consider the subspace of $\mathbb{R}^{1,3}$ consisting of the $x^0$ and the $x^1$ axes. Vectors in this two-dimensional sub-space are labelled by points which lie in one of, or at the meeting points of, the four sectors indicated below:




Let

$$\mathbf{v} = \begin{pmatrix} v^0 \\ v^1 \\ 0 \\ 0 \end{pmatrix} \tag{2.19}$$

be an arbitrary vector in $\mathbb{R}^{1,3}$ also lying entirely within $\mathbb{R}^{1,1}$ due to the zeroes in the the third and fourth compoenents. So

$$< \mathbf{v}, \mathbf{v} >= (v^0)^2 - (v^1)^2 \tag{2.20}$$

and hence if

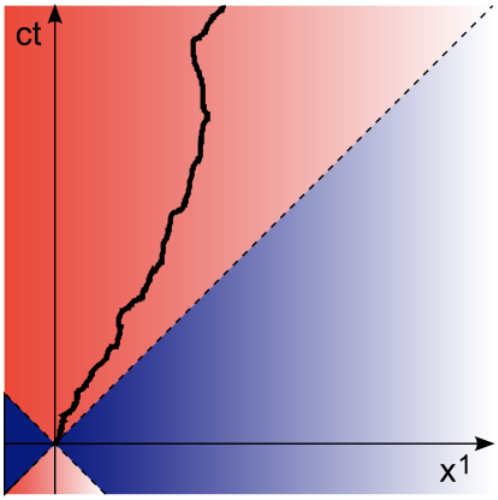
$$\begin{aligned} v^0 > v^1 \quad & \mathbf{v} \text{ is timelike.} \\ v^0 = v^1 \quad & \mathbf{v} \text{ is lightlike or null.} \\ v^0 < v^1 \quad & \mathbf{v} \text{ is spacelike.} \end{aligned} \tag{2.21}$$

In relativity Minkowski space, $\mathbb{R}^{1,3}$ equipped with the Minkowski metric $\eta$, is used to model spacetime. Spacetime, which we have taken for granted so far, has a local basis of coordinates which we are associated with time $t$ and the Cartesian coordinates $(x, y, z)$ by

$$x^0 = ct, \quad x^1 = x, \quad x^2 = y \quad \text{and} \quad x^3 = z \tag{2.22}$$

where $(x^0, x^1, x^2, x^3)$ are the components of a four-vector $\mathbf{x}$, $c$ is the speed of light - a useful constant that ensures that the dimenional units of $x^0$ are metres, the same as $x^1$, $x^2$ and $x^3$.

If we plot the graph of a one-dimensional (here $x^1$) motion of a particle against $x^0 = ct$ the resulting curve is called the worldline of the particle. We measure the position $x^1$ of the particle at a sequence of times and plot we might find a graph that looks like:




What is the gradient of the worldline?

$$\text{Gradient} = \frac{\Delta(ct)}{\Delta(x^1)} = \frac{c}{v^1} \tag{2.23}$$

where $v^1$ is the speed of the particle in the $x^1$ direction. Hence if the particle moves at the speed of light, $c$, then the gradient of the worldline is 1. In this case, when $x^1 = v^1 t = ct$ (and recalling the particle is only moving in the $x^1$ direction) then

$$\mathbf{x}^2 = (x^0)^2 - (x^1)^2 = (ct)^2 - (x^1)^2 = 0 \tag{2.24}$$

so $\mathbf{x}$ is a lightlike or null vector. If the gradient of the worldline is greater than one then $v^1 < c$ and $\mathbf{x}$ is timelike, otherwise if the gradient is less than one then $v^1 > c$ and $\mathbf{x}$ is a spacelike vector. One of the consequences of the special theory of relativity is that objects cannot cross the lightspeed barrier and objects with non-zero rest-mass cannot be accelerated to the speed of light.

**Problem 2.1.2.** *Compute the transformation of the space-time coordinates given by two consecutive Lorentz boosts along the $x$-axis, the first with speed $v$ and the second with speed $u$.*

**Problem 2.1.3.** *Compare your answer to problem 2.1.2 to the single Lorentz transformation given by $\Lambda(u \oplus v)$ where $\oplus$ denotes the relativistic addition of velocities. Hence show that*

$$u \oplus v = \frac{u+v}{1+\frac{uv}{c^2}}.$$

The spacetime at each point is split into four pieces. In the sketch above the set of null vectors form the boundaries of the light-cone for the origin. Given any arbitrary point in spcaetime $\mathbf{p}$ the set of vectors $\mathbf{x} - \mathbf{p}$ are all either timelike, spacelike or null. In the diagram above this would correspond to shifting the origin to the point $\mathbf{p}$, with spacetime again split into four pieces and their boundaries. The points which are connected to $\mathbf{p}$ by a timelike vector lie in the future or past lightcone of $\mathbf{p}$, those connected by a null vector lie on the surface lightcone of $\mathbf{p}$ and those connected by a spacelike vector to $\mathbf{p}$ are outside the lightcone. As nothing may cross the lightspeed barrier any point in spacetime can only exchange information with other points in spacetime which lie within or on its past or future lightcone.

In the two-dimensional spacetime that we have sketched it would be proper to refer to the forward or past light-triangle. The extension to four-dimensional spacetime is not easy to visualise. First consider extending the picture to a three-dimensional spacetime: add a second spatial axis $x^2$, as no spatial direction is singled out (there is a symmetry in the two spatial coordinates) the light-triangle of two-dimensions extends by rotating the the light-triangle around the temporal axis into the $x^2$ direction[4]. Rotating the light-triangle through three-dimensions gives the light-cone. The full picture for four-dimensional spacetime (being four-dimensional) is not possible to visualise and we refer still to the light-cone. However it is useful to be cautious when considering a drawing of a light cone and understand which dimensions (and how many) it really represents, e.g. a light-cone in four dimensions could be indicated by drawing a cone in three-dimensions with the implicit understanding that each point in the cone represents a two-dimensional space the drawing of which has been suppressed.

In all dimensions the lightcone is the cone at a point $\mathbf{p}$ is traced out by all the lightlike vectors connected to $\mathbf{p}$. No spacelike separated points can exchange a signal since the message would have to travel at a speed exceeding that of light.

We finish this section by making an observation that will make the connection between the definition of $O(1,3)$ and the Lorentz transformatons explicit. But which will be most usefully digested a second time after having read through the group theory chapter. Consider again the Lorentz boost transformation shown in equation (2.1). By making the substitution $\gamma = \cosh\xi$ the transformations are re-written in a way that looks a little like a rotation, it is in fact a hyperolic rotation. We note that $\cosh^2\xi - \sinh^2\xi = 1 = \gamma^2 - \sinh^2\xi$, i.e. $\sinh^2\xi = \gamma^2 - 1$, therefore we have the

[4]By taking a slice of the three dimensional graph through $ct$ and perpendicular to the $(x^1, x^2)$ plane the two-dimensional light-triangle structure reappear.

useful relation

$$\tanh\xi = \frac{1}{\gamma}(\gamma^2-1)^{\frac{1}{2}} = (1-\frac{1}{\gamma^2})^{\frac{1}{2}} = (1-(1-\frac{v^2}{c^2}))^{\frac{1}{2}} = \frac{v}{c}. \tag{2.25}$$

Hence we can rewrite the Lorentz boost in (2.1) as

$$ct' = c\cosh\xi\left(t - \frac{x}{c}\tanh\xi\right) = ct\cosh\xi - x\sinh\xi \tag{2.26}$$

$$x' = \cosh\xi\left(x - ct\tanh\xi\right) = x\cosh\xi - ct\sinh\xi \tag{2.27}$$

$$y' = y \tag{2.28}$$

$$z' = z \tag{2.29}$$

or in matrix form as

$$\mathbf{x}' \equiv \begin{pmatrix} ct' \\ x' \\ y' \\ z' \end{pmatrix} = \begin{pmatrix} \cosh\xi & -\sinh\xi & 0 & 0 \\ -\sinh\xi & \cosh\xi & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} ct \\ x \\ y \\ z \end{pmatrix} = \Lambda(\xi)\mathbf{x} \tag{2.30}$$

where $\Lambda$ is the four-by-four matrix indicated above and is a group element of $SO(1,3)$. The Lorentz boost is a hyberbolic rotation of $x$ into $ct$ and vice-versa.

**Problem 2.1.4.** *Show that* $\Lambda(\xi) \in SO(1,3)$.

## 2.2 Component Notation.

We have introduced the concept of the position four-vector implicitly as the extension of the usual three-vector in Cartesian coordinates to include a temporal coordinate. The position four vector is a particular four-vector $\mathbf{x}$ which specifies a unique position in space-time:

$$\mathbf{x} = \begin{pmatrix} ct \\ x \\ y \\ z \end{pmatrix}. \tag{2.31}$$

The components of the postion four-vector are denoted $x^\mu$ where $\mu \in \{0,1,2,3\}$ such that

$$x^0 = ct, \quad x^1 = x, \quad x^2 = y \quad \text{and} \quad x^3 = z. \tag{2.32}$$

It is frequently more useful to work with the components of the vector $x^\mu$ rather than the abstract vector $\mathbf{x}$ or the column vector in full. Consequently we will now develop a formalism for denoting vectors, their transposes, matrices, matrix multiplication and matrix action on vectors all in terms of component notation.

The notation $x^\mu$ with a single raised index we have defined to mean the entries in a single-column vector, hence the raised index denotes a row number (the components of a vector are labelled by their row). We have already met the Minkowski inner product which may be used to find the length-squared of a four-vector: it maps a pair of vectors

to a single scalar. Now a scalar object needs no index notation it is specified by a single number, i.e.

$$< \mathbf{x}, \mathbf{x} >= \mathbf{x}^2 = (x^0)^2 - (x^1)^2 - (x^2)^2 - (x^3)^2. \tag{2.33}$$

On the right-hand-side we see the distribution of the components of the vector. Our aim is to develop a notation that is useful, intuitive and carries some meaning within it. A good notation will improve our computation. We propose to develop a notation so that

$$\mathbf{x}^2 = x_\mu x^\mu \tag{2.34}$$

where $x_\mu$ is a row vector, although not always the simple transpose of $\mathbf{x}$. To do this we will develop matrix multiplication and the Einstein summation convention in the component notation.

### 2.2.1 Matrices and Matrix Multiplication.

Let us think gently about index notation and develop our component notation. Let $A$ be an invertible four-by-four matrix with real entries (i.e. $A \in GL(4, \mathbb{R})$). The matrix may multiply the four-vector $\mathbf{x}$ to give a new four-vector $\mathbf{x}'$. This means that in component notation matrix multiplication takes the component $x^\mu$ to $x'^\mu$, i.e. $\mathbf{x}' = A\mathbf{x}$. In terms of components we write the matrix entry for the $\mu$'th row and $\nu$'th column by $A^\mu{}_\nu$ and matrix multiplication is written as

$$x'^\mu = \sum_\nu A^\mu{}_\nu x^\nu. \tag{2.35}$$

This notation for matrix multiplication is consistent with our notation for a column vector $x^\mu$ and row vector $x_\nu$: raised indices indicate a row number while lowered indices indicate a column number. Hence the summation above is a sum of a product of entries in a row of the matrix and column of the vector - as the summation index $\nu$ is a column label (the matrix row $\mu$ stays constant in the sum). The special feature we have developped here is to distinguish the meaning of a raised and lowered index, otherwise teh expressions above are very familiar.

In more involved computations it becomes onerous to write out multiple summation symbols. So we adopt in most cases the *Einstein summation convention*, so called because it was notably adopted by Einstein in a 1916 paper on general relativity. As can be seen above the summation occurs over a pair of repeated indices, so it is not necessary to use the summation sign. Instead the Einstein summation convention assumes that there is an implicit summation over any pair of repeated indices in an expression. Hence the matrix multiplication written above becomes

$$x'^\mu = A^\mu{}_\nu x^\nu \tag{2.36}$$

when the Einstein summation convention is assumed. In four dimensions this means explcitly

$$x'^\mu = A^\mu{}_\nu x^\nu = A^\mu{}_0 x^0 + A^\mu{}_1 x^1 + A^\mu{}_2 x^2 + A^\mu{}_3 x^3. \tag{2.37}$$

The summed over indices no longer play any role on the right hand side and the index structure matches on either side of the expression: on both sides there is one free

raised $\mu$ index indiciating that we have the components of a vector on both sides of the equality. The repeated pair of indices which will be summed over and missing from the final expression are called 'dummy-indices'. It does not matter which symbol is used to denote a pair of indices to be summed over as they will vanish in the final expression, that is

$$x'^{\mu} = A^{\mu}{}_{\nu}x^{\nu} = A^{\mu}{}_{\sigma}x^{\sigma} = A^{\mu}{}_{\tau}x^{\tau} = A^{\mu}{}_{0}x^{0} + A^{\mu}{}_{1}x^{1} + A^{\mu}{}_{2}x^{2} + A^{\mu}{}_{3}x^{3}. \tag{2.38}$$

The index notation we have adopted is useful as free indices are matched on either side as are the positions of the indices.

So far so good, now we will run into an oddity in our conventions: the Minkowski metric does not have the index structure of a matrix in our conventions, even thought we wrote $\eta$ as a matrix previously! Recall that we aimed to be able to write $\mathbf{x}^2 = x_\mu x^\mu$. Now we understand the meaning of the right-hand-side, applying the Einstein summation convention we have

$$x_\mu x^\mu = x_0x^0 + x_1x^1 + x_2x^2 + x_3x^3 \tag{2.39}$$

but we have seen already that the Minkowski inner product is

$$< \mathbf{x}, \mathbf{x} >= (x^0)^2 - (x^1)^2 - (x^2)^2 - (x^3)^2 \tag{2.40}$$

so we gather that $x_0 = x^0$, $x_1 = -x^1$, $x_2 = -x^2$ and $x_3 = -x^3$ and as we hinted $x_\mu$ is not simply the components of the transpose of $\mathbf{x}$. It is the Minkwoski metric on Minkowski space that we may use to lower indices on vectors:

$$x_\mu \equiv \eta_{\mu\nu}x^\nu. \tag{2.41}$$

This is the analogue of vector transpose in Euclidean space (where the natural inner product is the identity matrix $\delta_{ij}$ and the transpose does not change the sign of the components as $x_i = \delta_{ij}x^j$. Now we note the flaw in our notation, as $\eta$ can lower indices then we could form an object $A_{\mu\nu} = \eta_{\mu\kappa}A^{\kappa}{}_{\nu}$ which is obviously related to a matrix $A^{\kappa}{}_{\nu}$. So we write $\eta$ as a matrix

$$\eta = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} \tag{2.42}$$

we are forced to defy our own conventions and understand $\eta_{\mu\nu}$ to mean the entry in the $\mu$'th row and $\nu$'th column of the matrix above.

Now we can write the Minkowski inner product in component notation:

$$\eta_{\mu\nu}x^\mu x^\nu = x^\mu x_\mu = x_\nu x^\nu = (x^0)^2 - (x^1)^2 - (x^2)^2 - (x^3)^2 =< \mathbf{x}, \mathbf{x} > . \tag{2.43}$$

The transpose has generalised to the raising and lowering of indices using the Minkowski metric $(x^\mu)^T = \eta_{\mu\nu}x^\nu = x_\mu$. To raise indices we use the inverse Minkowski metric denoted $\eta^{\mu\nu}$ and defined by

$$\eta_{\mu\nu}\eta^{\nu\rho} = \delta^{\mu}_{\rho} \tag{2.44}$$

which is the component form of $\eta\eta^{-1} = \mathrm{I}$. From the matrix form of $\eta$ we note that $\eta^{-1} = \eta$. We can raise indices with the inverse Minkowski metric: $x^\mu = \eta^{\mu\nu}x_\nu$.

***Exercise*** Show that the matrix multiplication $\Lambda^T \eta \Lambda = \eta$ used to define the matrices $\Lambda \in O(1,3)$ in component notation may be written as $\Lambda^\mu{}_\rho \eta_{\mu\nu} \Lambda^\nu{}_\sigma = \eta_{\rho\sigma}$.

***Solution***

$$\begin{aligned}
(\Lambda^T)^\mu{}_\rho \eta_{\mu\nu} \Lambda^\nu{}_\sigma &= \Lambda_\kappa{}^\tau \eta^{\mu\kappa} \eta_{\tau\rho} \eta_{\mu\nu} \Lambda^\nu{}_\sigma \\
&= \Lambda_\kappa{}^\tau \eta_{\tau\rho} \delta^\kappa_\nu \Lambda^\nu{}_\sigma \\
&= \Lambda_\kappa{}^\tau \eta_{\tau\rho} \Lambda^\kappa{}_\sigma \\
&= \Lambda_{\kappa\rho} \Lambda^\kappa{}_\sigma \\
&= \Lambda^\lambda{}_\rho \eta_{\lambda\kappa} \Lambda^\kappa{}_\sigma \\
&= \Lambda^\mu{}_\rho \eta_{\mu\nu} \Lambda^\nu{}_\sigma \\
&= \eta_{\rho\sigma}
\end{aligned}$$

where we have used the Minkowski metric to take the matrix transpose.

Since the components of vectors and matrices are numbers the order of terms in products is irrelevant in component notation e.g.

$$\eta_{\mu\nu} x^\nu = x^\nu \eta_{\mu\nu}$$

or

$$A^\mu{}_\nu x_\mu = (\mathbf{x}^T) A = x_\mu A^\mu{}_\nu.$$

We are also free to raise and lower simultaneously pairs of dummy indices:

$$x_\mu x^\mu = x^\nu \eta_{\mu\nu} x^\mu = x^\nu x_\nu = x^\mu x_\mu.$$

So we have many ways to write the same expression, but the key point for us are the things that do not vary: the objects involved in the expression ($\mathbf{x}$ and $A$ below) and the free indices (although the dummy indices may be redistributed):

$$\begin{aligned}
x^T A &= x_\mu A^\mu{}_\nu \\
&= x^\mu A_{\mu\nu} \\
&= A_{\mu\nu} x^\mu \\
&= A^{\rho\sigma} \eta_{\mu\rho} \eta_{\sigma\nu} x^\mu \\
&= A^{\rho\sigma} \eta_{\sigma\nu} x_\rho \\
&= A^\rho{}_\nu x_\rho
\end{aligned}$$

### 2.2.2 Common Four-Vectors

We have seen that the Minkwoski inner product gives a Lorentz-invariant quantity for any pair of four-vectors. We can make use of this Lorentz invariance to construct new but familiar four-vectors. Consider two events, one occurring at the 4-vector $\mathbf{x}$ and another at $\mathbf{y}$ where

$$\mathbf{x} = \begin{pmatrix} ct_1 \\ x_1 \\ y_1 \\ z_1 \end{pmatrix} \quad \text{and} \quad \mathbf{y} = \begin{pmatrix} ct_2 \\ x_2 \\ y_2 \\ z_2 \end{pmatrix}. \tag{2.45}$$

In Newtonian physics the difference in the time $\Delta t \equiv |t_2 - t_1|$ the two events occurred at and the distance in space between the locations of the two events $\Delta r \equiv \sqrt{\sum_{i=1}^{3} |x^i - y^i|^2}$ are both invariants of the Gallilean transformations. As we have seen, under the Lorentz transformations a new single invariant emerges: $|\mathbf{x} - \mathbf{y}|^2 =\equiv c^2\tau_{xy}$ where $\tau_{xy}$ is called the *proper time* between two events $\mathbf{x}$ and $\mathbf{y}$, i.e.

$$c^2\tau_{xy}^2 = c^2(t_2 - t_1)^2 - (x_2 - x_1)^2 - (y_2 - y_1)^2 - (z_2 - z_1)^2. \tag{2.46}$$

Every point $\mathbf{x}$ in space-time has a proper-time associated to it by

$$c^2\tau_x^2 = c^2t_1^2 - x_1^2 - y_1^2 - z_1^2 = x^\mu x_\mu \tag{2.47}$$

We have already shown in problem 2.1.1 that this is invariant under the under the Lorentz transformations and one can show that $\tau_{xy}$ is also invariant as $c^2\tau_{xy}^2 =< \mathbf{x} - \mathbf{y}, \mathbf{x} - \mathbf{y} >= (x-y)^\mu(x-y)_\mu$. Now as $< \mathbf{x}-\mathbf{y}, \mathbf{x}-\mathbf{y} >= \mathbf{x}^2 - 2 < \mathbf{x}, \mathbf{y} > +\mathbf{y}^2$ is invariant then we can conlude that $< \mathbf{x}, \mathbf{y} >$ is also an invariant as $\mathbf{x}^2$ and $\mathbf{y}^2$ are also invariant under the Lorentz transformations.

**Problem 2.2.1.** *Show explicitly that* $< \mathbf{x}, \mathbf{y} >= x^\mu y_\mu$ *is invariant under the Lorentz group.*

These quantities are all called Lorentz-invariant quantities. You will notice that they do not have any free indices for the Lorentz group to act on.

All four-vectors transform in the same way as the position four-vector $\mathbf{x}$ under a Lorentz transformation (just as 3D vectors all transform in the same way under $SO(3)$ rotations). We can find other physically relevant four-vectors by combining the position four-vector $\mathbf{x}$ with Lorentz invariant quantities. For example the Lorentz four-velocity $\mathbf{u}$ is defined using the proper time, which is Lorentz invariant, rather than time which is not:

$$\mathbf{u} = \frac{d\mathbf{x}}{d\tau} = \frac{d\mathbf{x}}{dt}\frac{dt}{d\tau} = \frac{dt}{d\tau}\begin{pmatrix} c \\ u^1 \\ u^2 \\ u^3 \end{pmatrix} \tag{2.48}$$

where $\begin{pmatrix} u^1 \\ u^2 \\ u^3 \end{pmatrix}$ is the usual Newtonian velocity vector in $\mathbb{R}^3$. Let us compute $\frac{dt}{d\tau}$, starting from

$$\tau = \frac{1}{c}\sqrt{c^2t^2 - x^2 - y^2 - z^2} \tag{2.49}$$

then

$$\begin{aligned} \frac{d\tau}{dt} &= \frac{1}{2c^2\tau}(2c^2t - 2xu_1 - 2yu_2 - 2zu_3) \\ &= \frac{(t - \frac{xu_1}{c^2} - \frac{yu_2}{c^2} - \frac{zu_3}{c^2})}{\tau} \\ &= \frac{t(1 - \frac{u^2}{c^2})}{\tau} \\ &= \frac{\gamma}{\gamma^2} \\ &= \gamma^{-1} \end{aligned} \tag{2.50}$$

where $u^2 = (u^1)^2 + (u^2)^2 + (u^3)^2$ and $\frac{1}{\gamma} = \sqrt{1 - \frac{u^2}{c^2}}$. Hence the four velocity is given by

$$\mathbf{u} = \gamma \begin{pmatrix} c \\ u^1 \\ u^2 \\ u^3 \end{pmatrix}. \tag{2.51}$$

We can check that $\mathbf{u}^2$ is invariant:

$$\mathbf{u}^2 = u^\mu u_\mu = \gamma^2(c^2 - u^2) = c^2\gamma^2(1 - \frac{u^2}{c^2}) = c^2 \tag{2.52}$$

The four-momentum is defined as $\mathbf{p} = m\mathbf{u}$ where $m$ is the rest-mass. The spatial part of the four-momentum is the usual Newtonian momentum $p_N$ multiplied by $\gamma$, while the zeroth component is proportional to energy:

$$p^0 = \frac{E}{c} = \gamma mc. \tag{2.53}$$

The invariant quantity associated to $\mathbf{p}$ is

$$p^\mu p_\mu = (\frac{E}{c})^2 - \gamma^2 p_N^2 = m^2c^2 \tag{2.54}$$

Rearranging gives

$$E = (m^2c^4 + \gamma^2 p_N^2 c^2)^{\frac{1}{2}} \tag{2.55}$$

which is the relativistic version of $E = \frac{1}{2}mu^2$ and you could expand the above expression to find the usual kinetic energy term together with other less familiar terms. For a particle at rest we have $\gamma = 1$ and $p_N = 0$ hence we find a particle's rest energy $E_0$ is

$$E_0 = mc^2. \tag{2.56}$$

### 2.2.3 Maxwell's Equations.

The first clue that there was a democracy between time and space came with the discovery of Maxwell's equations. James Clerk Maxwell's work that led to his equations began in his 1861 paper 'On lines of physical force' which was written while he was at King's College London (1860-1865). The equations include an invariant speed of propagation for electromagnetic waves $c$, the speed of light, which is one of the two assumptions in Einstein's special theory of relativity. Consequently they have an elegant formulation when written in terms of Lorentz tensors.

Maxwell's theory of electromagnetism (in the absence of charge density and current) can be derived from an action $S$ given by

$$S = \int d^4\mathbf{x}\, \mathcal{L} \tag{2.57}$$

where

$$\mathcal{L} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} \tag{2.58}$$

with

$$F_{\mu\nu} \equiv \partial_\mu A_\nu - \partial_\nu A_\mu \tag{2.59}$$

and $\mu, \nu \in \{0, 1, 2, 3\}$.

**Problem 2.2.2.** *Show that the transformation*

$$A_\mu \to A_\mu + \partial_\mu \Lambda \tag{2.60}$$

*where $\Lambda$ is an arbitrary function of $x^\mu$ leaves the Lagrangian invariant. N.B. $\Lambda$ is* **not** *a Lorentz transformation just an arbitrary function of $x^\mu$.*

The fact that one may arbitrarily shift the potential $A_\mu$ in this way without changing $\mathcal{L}$ is an example of a gauge symmetry. These symmetries are a pivotal part of the standard model of particle physics and this "$U(1)$" gauge symmetry of electromagnetism is the prototypical example of gauge symmetry.

We would like to use the action above to find the equations of motion but we are immediately at a loss if we attempt to write Lagrange's equations. The problem is we have put space and time on an equal footing in relativity, and in the above action, while in Lagrangian mechanics the temporal derivative plays a special role and is distinguished from the spatial derivative. Lagrange's equations are not covariant. We will return to this problem and address how to upgrade Lagrange's equations to space-time. Here we will vary the fields $A_\mu$ in the action directly and read off the equation of motion. To simplify the expressions we begin by writing the variation of the Lagrangian:

$$\delta_A \mathcal{L} = -\frac{1}{4}\delta_A(F_{\mu\nu})F^{\mu\nu} - \frac{1}{4}F_{\mu\nu}\delta_A(F^{\mu\nu}) \tag{2.61}$$

$$= -\frac{1}{2}\delta_A(F_{\mu\nu})F^{\mu\nu} \tag{2.62}$$

Now under a variation of $A_\mu$ the field strength $F_{\mu\nu}$ transforms as

$$F_{\mu\nu} \to \partial_\mu(A_\nu + \delta A_\nu) - \partial_\nu(A_\mu + \delta A_\mu) \equiv F_{\mu\nu} + \delta_A(F_{\mu\nu}) \tag{2.63}$$

so we read off

$$\delta_A(F_{\mu\nu}) = \partial_\mu(\delta A_\nu) - \partial_\nu(\delta A_\mu). \tag{2.64}$$

So from the variation of the Lagrangian we have:

$$\delta_A \mathcal{L} = -\frac{1}{4}\delta_A(F_{\mu\nu})F^{\mu\nu} - \frac{1}{4}F_{\mu\nu}\delta_A(F^{\mu\nu}) \tag{2.65}$$

$$= -\frac{1}{2}\Big(\partial_\mu(\delta A_\nu) - \partial_\nu(\delta A_\mu)\Big)F^{\mu\nu} \tag{2.66}$$

$$= -\partial_\mu(\delta A_\nu)F^{\mu\nu} \tag{2.67}$$

where we have used the antisymmetry of $F^{\mu\nu} = -F^{\nu\mu}$ and a relabelling of the dummy indices in the second term of the second line to arrive at the final expression. To take the derivative off of $A_\mu$ we use the same technique as when one integrates by parts (although here there is no integral, but when we put the Lagrangian variation back into the action there will be) namely we rewrite the expression using the observation that

$$\partial_\mu(\delta A_\nu F^{\mu\nu}) = \partial_\mu(\delta A_\nu)F^{\mu\nu} + \delta A_\nu \partial_\mu(F^{\mu\nu}) \tag{2.68}$$

to give

$$\delta_A \mathcal{L} = -\partial_\mu(\delta A_\nu F^{\mu\nu}) + \delta A_\nu \partial_\mu(F^{\mu\nu}). \tag{2.69}$$

Returning to the action we have

$$\delta_A S = \int d^4\mathbf{x}\Big( - \partial_\mu(\delta A_\nu F^{\mu\nu}) + \delta A_\nu \partial_\mu(F^{\mu\nu})\Big). \tag{2.70}$$

The first term we can integrate diretl - it is called a boundary term as it is a total derivative - but it vanishes as the term $\delta A_\nu$ vanishes at the fixed points of the path (in field space) we are varying leaving us with

$$0 = \delta_A S = \int d^4\mathbf{x}\delta A_\nu \partial_\mu(F^{\mu\nu}). \tag{2.71}$$

Hence the field equation is

$$\partial_\mu F^{\mu\nu} = 0. \tag{2.72}$$

This is a space-time equation. If we split it up into spatial and temporal components we can reconstruct Maxwell's equations in their familiar form. To do this we introduce the electric $\mathbf{E}$ and magnetic $\mathbf{B}$ fields in terms of components of the field strength:

$$F^{0i} = E^i \quad \text{and} \quad F^{ij} = \epsilon^{ijk}B^k \tag{2.73}$$

where $E^i$ and $B^i$ are the components of $\mathbf{E}$ and $\mathbf{B}$ respectively, $i, j, k \in \{1, 2, 3\}$ and $\epsilon^{ijk}$ is the Levi-Civita symbol normalised such that $\epsilon^{123} = 1$. We will meet the Levi-Civita symbol when we study tensor representations in group theory, at this point it is sufficient to know that it has six components which take the values:

$$\begin{aligned} &\epsilon^{123} = 1, \quad \epsilon^{231} = 1, \quad \epsilon^{312} = 1 \\ &\epsilon^{213} = -1, \quad \epsilon^{132} = -1, \quad \epsilon^{321} = -1 \end{aligned} \tag{2.74}$$

note that swapping of any neighbouring indices changes the sign of the Levi-Civita symbol - the Levi-Civita symbol is an 'antisymmetric' tensor. We will split the equation of motion in equation (2.72) into its temporal part $\nu = 0$ and its spatial part $\nu = i$ where $i \in \{1, 2, 3\}$. Taking $\nu = 0$ we have

$$\partial_0 F^{00} + \partial_i F^{i0} = -\partial_i E^i = 0 \tag{2.75}$$

that is

$$\nabla \cdot \mathbf{E} = 0 \tag{2.76}$$

From the spatial equations ($\nu = i$) we have

$$\partial_0 F^{0i} + \partial_j F^{ji} = \partial_0 E^i + \partial_j(\epsilon^{jik}B^k) = \frac{1}{c}\partial_t E^i - \epsilon^{ijk}\partial_j(B^k) = 0 \tag{2.77}$$

i.e.

$$\nabla \times \mathbf{B} = \frac{1}{c}\frac{\partial \mathbf{E}}{\partial t}. \tag{2.78}$$

That is all we obtain from the equation of motion, so we seem to be two equations short! However there is an identity that is valid on the field strength simply due to its definition. Formerly $F_{\mu\nu}$ is an 'exact form' as it is the 'exterior derivative' of the 'one-form' $A_\mu$[5]. Exact forms vanish when their exterior derivative, which is the antisymmetrised partial derivative, is taken.

[5]Differential forms are a subset of the tensors whose indices are antisymmetric. They are introduced and studied in depth in the Manifolds course.

**Problem 2.2.3.** *Show that*

$$3\partial_{[\mu}F_{\nu\rho]} \equiv \partial_\mu F_{\nu\rho} + \partial_\nu F_{\rho\mu} + \partial_\rho F_{\mu\nu} = 0 \tag{2.79}$$

The identity $\partial_{[\mu}F_{\nu\rho]} = 0$ is called the Bianchi identity for the field strength and is a consequence of its antisymmetric construction. However it is non-trivial and it is from the Bianchi identity for $F_{\mu\nu}$ that the remaining two Maxwell equations emerge.

Let us consider all the non-trivial spatial and temporal components of $\partial_{[\mu}F_{\nu\rho]} = 0$. We note that we cannot have more than one temporal index before the identity trivialises, e.g. let $\mu = \nu = 0$ and $\rho = i$ then we have

$$\partial_0 F_{0i} + \partial_0 F_{i0} + \partial_i F_{00} = \partial_0 F_{0i} - \partial_0 F_{0i} = 0 \tag{2.80}$$

from which we learn nothing. When we take $\mu = 0$, $\nu = i$ and $\rho = j$ we have

$$\partial_0 F_{ij} + \partial_i F_{j0} + \partial_j F_{0i} = 0 \tag{2.81}$$

We must use the Minkowski metric to find the components $F_{\mu\nu}$ of the field strength in terms of $\mathbf{E}$ and $\mathbf{B}$:

$$F_{ij} = \eta_{i\mu}\eta_{j\nu}F^{\mu\nu} = \eta_{ik}\eta_{jl}F^{kl} = F^{ij} = \epsilon^{ijk}B^k \tag{2.82}$$

$$F_{0i} = \eta_{0\mu}\eta_{i\nu}F^{\mu\nu} = \eta_{ik}F^{0k} = -F^{0i} = -E^i. \tag{2.83}$$

Substituting these expressions into equation (2.81) gives

$$\partial_0(\epsilon^{ijk}B^k) + \partial_i E^j - \partial_j E^i = 0. \tag{2.84}$$

To reformulate this in a more familiar way we can make use of an identity on the Levi-Civita symbol:

$$\epsilon_{ijm}\epsilon^{ijk} = 2\delta^k_m. \tag{2.85}$$

**Problem 2.2.4.** *Prove that* $\epsilon_{ijm}\epsilon^{ijk} = 2\delta^k_m$.

Contracting $\epsilon_{ijm}$ with equation (2.84) gives

$$\begin{aligned}\epsilon_{ijm}\partial_0(\epsilon_{ijk}B^k) + \epsilon_{ijm}\partial_i E^j - \epsilon_{ijm}\partial_j E^i &= 2\partial_0(B^m) + \epsilon_{ijm}\partial_i E^j - \epsilon_{ijm}\partial_j E^i \\ &= 2\partial_0(B^m) + 2\epsilon_{ijm}\partial_i E^j = 0\end{aligned} \tag{2.86}$$

which we recognise as

$$\nabla \times \mathbf{E} = -\frac{1}{c}\frac{\partial \mathbf{B}}{\partial t}. \tag{2.87}$$

The final Maxwell equation comes from setting $\mu = i$, $\nu = j$ and $\rho = k$ in equation (2.79):

$$\partial_i F_{jk} + \partial_j F_{ki} + \partial_k F_{ij} = \partial_i(\epsilon^{jkl}B^l) + \partial_j(\epsilon^{kil}B^l) + \partial_k(\epsilon^{ijl}B^l) = 0 \tag{2.88}$$

Contracting this with $\epsilon_{ijk}$ gives

$$\begin{aligned}\epsilon_{ijk}\Big(\partial_i(\epsilon^{jkl}B^l) + \partial_j(\epsilon^{kil}B^l) + \partial_k(\epsilon^{ijl}B^l)\Big) &= \partial_i(2\delta^l_i B^l) + \partial_j(2\delta^l_j B^l) + \partial_k(2\delta^l_k B^l) \\ &= 6\partial_i B^i \\ &= 0\end{aligned} \tag{2.89}$$

That is,

$$\nabla \cdot \mathbf{B} = 0. \tag{2.90}$$

### 2.2.4 Electromagnetic Duality

The action for electromagnetism can be rewritten in terms of $\mathbf{E}$ and $\mathbf{B}$ where it has a very simple form. Now

$$\begin{aligned} F_{\mu\nu}F^{\mu\nu} &= F_{0\nu}F^{0\nu} + F_{i\nu}F^{i\nu} && (2.91)\\ &= F_{00}F^{00} + F_{0i}F^{0i} + F_{i0}F^{i0} + F_{ij}F^{ij} && (2.92)\\ &= -2E^iE^i + \epsilon^{ijk}B^k\epsilon^{ijl}B^l && (2.93)\\ &= -2E^iE^i + 2B^iB^i && (2.94)\\ &= -2\mathbf{E}^2 + 2\mathbf{B}^2. && (2.95) \end{aligned}$$

Hence,

$$\mathcal{L} = \frac{1}{2}(\mathbf{E}^2 - \mathbf{B}^2) \tag{2.96}$$

Some symmetry is apparent in the form of the Lagrangian and the equations of motion. We notice (after some reflection) that if we interchange $\mathbf{E} \to -\mathbf{B}$ and $\mathbf{B} \to \mathbf{E}$ that while the Lagrangian changes sign, the equations of motion are unaltered. This is electromagnetic duality: an ability to swap electric fields for magnetic fields while preserving Maxwell's equations[6].

**Problem 2.2.5.** *Show that under the electromagnetic duality transformations* $(\mathbf{E}, \mathbf{B}) \to (-\mathbf{B}, \mathbf{E})$ *Maxwell's equations in a vacuum are invariant while* $\mathcal{L} \to -\mathcal{L}$.

**Problem 2.2.6.** *With the addition of electric charge and currents the Lagrangian becomes*

$$\mathcal{L} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + A_\mu J^\mu \tag{2.97}$$

*where* $J^\mu$ *are the components of the current four-vector:* $J^0 = c\rho$, $J^i = j^i$ *where* $\rho$ *is the charge density and* $\mathbf{j}$ *is the current density. Show that electromagnetic duality is no longer a symmetry of the modified Maxwell equations.*

Curiously electromagnetic duality is much more apparent in the associated Hamiltonian which takes the form

$$\mathcal{H} = \frac{1}{2}(\mathbf{E}^2 + \mathbf{B}^2) \tag{2.98}$$

which is itself invariant under $(\mathbf{E}, \mathbf{B}) \to (-\mathbf{B}, \mathbf{E})$.

### 2.2.5 Field Theory Equations of Motion

The covariant action for electromagnetism is a functional on fields $A_\mu$ and as such we are not able to apply the Lagrange equations. Of course it was no problem to vary the fields and read off the equation of motion directly, but for curiousity's sake we write out the field theory upgrade of Lagrange's equations. It will be helpful to first consider the equations of motion for a specific field theory action for a scalar field $\phi$ from which we will deduce the field theory Lagrange equations:

$$S = \int d^4\mathbf{x}\left(\frac{1}{2}\partial_\mu\phi\,\partial^\mu\phi - V(\phi)\right) \tag{2.99}$$

[6]The eagle-eyed reader will notice that the electromagnetic duality transformation exchanges equations of motion for Bianhci identities.

where $\phi = \phi(\mathbf{x})$ is a scalar field in space-time, and $V(\phi)$ is an arbitrary potential term. The first term is the kinetic term for the field, but as it is Lorentz invariant it includes spatial derivatives of $\phi$ as well as the velocity $\dot{\phi}$. We may extremise this simple action with respect to a change in the field: $\phi \to \phi + \delta\phi$ giving

$$\delta S = \int d^4\mathbf{x}\left(\partial_\mu \delta\phi\, \partial^\mu \phi - \frac{\partial V}{\partial \phi}\delta\phi\right) = \int d^4\mathbf{x}\, \delta\phi\left( - \partial_\mu \partial^\mu \phi - \frac{\partial V}{\partial \phi}\right) = 0. \tag{2.100}$$

The equation of motion is

$$\partial_\mu \partial^\mu \phi + \frac{\partial V}{\partial \phi} = 0. \tag{2.101}$$

Compared with the usual Lagrange equations we can compare the generalised coordinate $q$ with the field $\phi$, and the velocity $\dot{q}$ with $\partial_\mu \phi$. This comparison aids our appreciation of the field theory equation of motion, which at first glance can appear unsettlingly alien. In terms of the original Lagrangian $\mathcal{L}$ the equation of motion is

$$\partial_\mu\left(\frac{\partial \mathcal{L}}{\partial \partial_\mu \phi}\right) - \frac{\partial \mathcal{L}}{\partial \phi} = 0. \tag{2.102}$$

By replacing the scalar field above with arbitrary tensor fields we can find the equation of motion for more general field theory actions including the action for Maxwell's electromagnetism where the fundamental field is a vector $A_\mu$ and the equation of motion becomes

$$\partial_\mu\left(\frac{\partial \mathcal{L}}{\partial \partial_\mu A_\nu}\right) - \frac{\partial \mathcal{L}}{\partial A_\nu} = \partial_\mu F^{\mu\nu} = 0. \tag{2.103}$$

# Chapter 3

# Quantum Mechanics

Historically quantum mechanics was constructed rather than logically developed. The mathematical procedure of quantisation was later rigorously developed by mathematicians and physicists, for example by Weyl; Kohn and Nirenberg; Becchi, Rouet, Stora and Tyutin (BRST quantisation for quantising a field theory); Batalin and Vilkovisky (BV field-antifield formalism) as well as many other significant contributions and research into quantisation methods continues to this day. The original development of quantum mechanics due to Heisenberg is called the canonical quantisation and it is the approach we will follow here.

Atomic spectra are particular to specific elements, they are the fingerprints of atomic forensics. An atomic spectrum is produced by bathing atoms in a continuous spectrum of electromagnetic radiation. The electrons in the atom make only discrete jumps as the electromagnetic energy is absorbed. This can be seen in the atomic spectra by the absence of specific frequencies in the outgoing radiation and by recalling that $E = h\nu$ where $E$ is energy, $h$ is Planck's constant and $\nu$ is the frequency.

In 1925 Heisenberg was working with Born in Gottingen. He was contemplating the atomic spectra of hydrogen but not making much headway and he developed the most famous bout of hayfever in theoretical physics. Complaining to Born he was granted a two-week holiday and escaped the pollen-filled inland air for the island of Helgoland. He continued to work and there in a systematic fashion. He arranged all the known frequencies for the spectral lines of hydrogen into an array, or matrix, of frequencies $\nu_{ij}$. He was also able to write out matrices of numbers corresponding to the transition rates between energy levels. Armed with this organisation of the data, but with no knowledge of matrices, Heisenberg developed a correspondence between the harmonic oscillator and the idea of an electron orbitting in an extremely eccentric orbit. Having arrived at a consistent theory of observable quanitites, Heisenberg climbed a rock overlooking the sea and watched the sun rise in a moment of triumph. Heisenberg's triumph was short-lived as he quickly realised that his theory was based around non-commuting variables. One can imagine his shock realising that everything worked so long as the multiplication was non-Abelian, nevertheless Heisenberg persisted with his ideas. It was soon pointed out to him by Born that the theory would be consistent if the variables were matrices, to which Heisenberg replied that "I do not even know what a matrix is". The oddity that matrices were seen as an unusual mathematical formalism and not

a natural setting for physics played an important part in the development of quantum mechanics. As we will see a wave equation describing the quantum theory was developed by Schrödinger in apparent competition to Heisenberg's formulation. This was, in part, a reaction to the appearance of matrices in the fundamental theory as well as a rejection of the discontinuities inherent in Heisenberg's quantum mechanics. Physicists much more readily adopted Schrödinger's wave equation which was written in the language of differential operators with which physicists were much more familiar. In this chapter we will consider both the Heisenberg and Schrödinger pictures and we will see the equivalence of the two approaches.

## 3.1 Canonical Quantisation

We commence by recalling the structures used in classical mechanics. Consider a classical system described by $n$ generalised coordinates $q_i$ of mass $m_i$ subject to a potential $V(q_i)$ and described by the Lagrangian

$$\mathcal{L} = \sum_{i=1}^{n} \frac{1}{2} m_i \dot{q}_i^2 - \sum_{i=1}^{n} V(q_i) \tag{3.1}$$

where $V(q) = V(q_1, q_2, \ldots q_n)$. The equations of motion are:

$$m_i \ddot{q}_i + \frac{\partial V}{\partial q_i} = 0 \qquad \Rightarrow \qquad F_i = m_i \ddot{q}_i. \tag{3.2}$$

The Hamiltonian is

$$\mathcal{H} = \sum_{i=1}^{n} p_i \dot{q}_i - \mathcal{L} = \frac{p_i^2}{2m_i} + V(q) \tag{3.3}$$

and the Hamiltonian equations make explicit that there exists a natural antisymmetric (symplectic) structure on the phase space, the Poisson brackets:

$$\{q_i, p_j\} = \delta_{ij} \tag{3.4}$$

with all other brackets being trivial.

Canonical quantisation is the promotion of the positions $q_i$ and momenta $p_i$ to operators (which we denote with a hat):

$$(q_i, p_i) \longrightarrow (\hat{q}_i, \hat{p}_i) \tag{3.5}$$

together with the promotion of the Poisson bracket to the commutator by

$$\{A, B\} \longrightarrow \frac{1}{i\hbar}[\hat{A}, \hat{B}] \tag{3.6}$$

where $A$ and $B$ indicate arbitrary functions on phase space, while $\hat{A}$ and $\hat{B}$ are operators. For example we have

$$[\hat{q}_i, \hat{p}_j] = i\hbar\, \delta_{ij} \tag{3.7}$$

where $\hbar \equiv \frac{h}{2\pi}$ and $h$ is Planck's constant. In particular the classical Hamiltonian becomes under this promotion

$$\mathcal{H} \longrightarrow \hat{\mathcal{H}} = \sum_{i=1}^{n} \frac{\hat{p}_i^2}{2m_i} + \sum_i V(\hat{q}_i). \tag{3.8}$$

While the classical $q_i$ and $p_i$ collect to form vectors in phase space, the quantum operators $\hat{q}_i$ and $\hat{p}_i$ belong to a Hilbert space. In quantum mechanics physical observables are represented by operators which act on the Hilbert space of quantum states. The states include eigenstates for the operators and the corresponding eigenvalue represents the value of a measurement. For example we might denote a position eigenstate with eigenvalue $q$ for the position operator $\hat{q}$ by $|q\rangle$ so that:

$$\hat{q}|q\rangle = q|q\rangle \tag{3.9}$$

we will meet the bra-ket notation more formally later on, but it is customary to label an eigenstate by its eigenvalue hence the eigenstate is denoted $|q\rangle$ here. More general states are formed from superpositions of eigenstates e.g.

$$|\psi\rangle = \int dx \psi(x)|x\rangle \qquad \text{or} \qquad |\psi\rangle = \sum_i \psi_i |q_i\rangle \tag{3.10}$$

where we have taken $|x\rangle$ as a continuous basis for the Hilbert space while $|q_i\rangle$ is a discrete basis.

If we work using the eigenfunctions of the positon operator as a basis for the Hilbert space it is customary to refer to states in the 'position space'. By expressing states as a superposition of position eigenfunctions we determine an expression for the momentum operator in the position space. For simplicity, consider a single particle state described by a single coordinate given by $\psi = c(q)|q\rangle$, where $|q\rangle$ is the eigenstate of the position operator $\hat{q}$ and $\hat{q}\psi = q\psi$. The commutator relation $[\hat{q}, \hat{p}] = i\hbar$ fixes the momentum operator to be

$$\hat{p} = -i\hbar \frac{\partial}{\partial q} \tag{3.11}$$

as

$$\begin{aligned} [\hat{q}, \hat{p}]\psi &= (\hat{q}\hat{p} - \hat{p}\hat{q})c|q\rangle \\ &= \hat{q}\hat{p}c|q\rangle - \hat{p}qc|q\rangle \\ &= -i\hbar\hat{q}\frac{\partial c}{\partial q}|q\rangle + i\hbar\frac{\partial (qc)}{\partial q}|q\rangle \\ &= i\hbar\psi \end{aligned} \tag{3.12}$$

For many-particle systems we may take the position eigenstates as a basis for the Hilbert space and the state and momentum operator generalise to

$$\psi \equiv \sum_i c_i(q)|q_i\rangle \qquad \text{and} \qquad \hat{p}_i \equiv -i\hbar\frac{\partial}{\partial q_i}. \tag{3.13}$$

Note that the Hamiltonian operator in the position space becomes

$$\hat{\mathcal{H}} = \sum_i -\frac{\hbar^2}{2m_i}\frac{\partial^2}{\partial q_i^2} + \sum_i V(\hat{q}_i). \tag{3.14}$$

### 3.1.1 The Hilbert Space and Observables.

**Definition** A *Hilbert space H* is a complex vector space equipped with an inner product $<,>$ satisfying:

(i.) $< \phi, \psi >= \overline{< \psi, \phi >}$

(ii.) $< \phi, a_1\psi_1 + a_2\psi_2 >= a_1 < \phi, \psi_1 > + a_2 < \phi, \psi_2 >$

(iii.) $< \phi, \phi > \geq 0 \quad \forall \quad \phi \in H$ where equality holds only if $\phi = 0$.

where $\overline{\psi}$ indicates the complex conjugate of $\psi$

Note that as the inner product is linear in its second entry, it is conjugate linear in its first entry as

$$\begin{aligned} < a_1\phi_1 + a_2\phi_2, \psi > &= \overline{< \psi, a_1\phi_1 + a_2\phi_2 >} \\ &= a_1^*\overline{< \psi, \phi_1 >} + a_2^*\overline{< \psi, \phi_2 >} \\ &= a_1^* < \phi_1, \psi > + a_2^* < \phi_2, \psi > \end{aligned} \tag{3.15}$$

where we have used $a_1^*$ to indicate the complex-conjugate of $a_1$. The physical states in a system are described by normalised vectors in the Hilbert space, i.e. those $\psi \in H$ such that $< \psi, \psi >= 1$.

Observables are represented by Hermitian operators in $H$. Hermitian operators are self-adjoint.

**Definition** An operator $\hat{A}^*$ is the *adjoint* operator of $\hat{A}$ if

$$< \hat{A}^*\phi, \psi >=< \phi, \hat{A}\psi > . \tag{3.16}$$

From the definition it is rapidly observed that

- $\hat{A}^{**} = \hat{A}$
- $(\hat{A} + \hat{B})^* = \hat{A}^* + \hat{B}^*$
- $(K\hat{A})^* = K^*\hat{A}^*$
- $(\hat{A}\hat{B})^* = \hat{B}^*\hat{A}^*$
- If $\hat{A}^{-1}$ exists then $(\hat{A}^{-1})^* = (\hat{A}^*)^{-1}$.

A self-adjoint operator satisfies $A^* = A$. The prototype for the adjoint is the Hermitian conjugate of a matrix $M^\dagger \equiv (M^T)^*$.

**Example 1: $L^2$ as a Hilbert Space**

Let $H = L^2(\mathbb{R})$ i.e. $\psi \in H \Rightarrow < \psi, \psi >< \infty$ and the inner product is

$$< \phi, \psi >\equiv \int_{\mathbb{R}} dq\, \phi^*(q)\psi(q). \tag{3.17}$$

Using this inner product the momentum operator is a self-adjoint operator as

$$\begin{aligned} < \phi, \hat{p}\psi > &= \int_{\mathbb{R}} dq\, \phi^*(q)\bigg( - i\hbar\frac{\partial}{\partial q}\psi(q)\bigg) \\ &= \int_{\mathbb{R}} dq\, i\hbar\bigg(\frac{\partial}{\partial q}\phi^*(q)\bigg)\psi(q) \\ &= \int_{\mathbb{R}} dq\, \bigg( - i\hbar\frac{\partial}{\partial q}\phi(q)\bigg)^*\psi(q) \\ &=< \hat{p}\,\phi, \psi > \end{aligned} \tag{3.18}$$

N.B. we have assumed that $\phi \to 0$ and $\psi \to 0$ at $q = \pm\infty$ such that the boundary term from the integration by parts vanishes.

**Example 2: Hermitian Matrices on $\mathbb{C}^n$ as self-adjoint operators.**

On $\mathbb{C}^n$ the natural inner product is

$$< \mathbf{x}, \mathbf{y} >\equiv \mathbf{x}^\dagger \mathbf{y}. \tag{3.19}$$

Let $\hat{A}$ denote a self-adjoint matrix and we will show that $\hat{A}^* = \hat{A}^\dagger$:

$$< \mathbf{x}, \hat{A}\mathbf{y} >= \mathbf{x}^\dagger \hat{A}\mathbf{y} = (\hat{A}^\dagger \mathbf{x})^\dagger \mathbf{y} =< \hat{A}^\dagger \mathbf{x}, \mathbf{y} > . \tag{3.20}$$

### 3.1.2 Eigenvectors and Eigenvalues

In this section we will prove some simple properties of eigenvalues of self-adjoint operators.

Let $\mathbf{u} \in H$ be an eigenvector for the operator $\hat{A}$ with eigenvalue $\alpha \in \mathbb{C}$ such that

$$\hat{A}\mathbf{u} = \alpha \mathbf{u}. \tag{3.21}$$

The eigenvalues of a self-adjoint operator are real:

$$\begin{aligned} &< \mathbf{u}, \hat{A}\mathbf{u} >=< \mathbf{u}, \alpha\mathbf{u} >= \alpha < \mathbf{u}, \mathbf{u} > \\ &=< \hat{A}\mathbf{u}, \mathbf{u} >=< \alpha\mathbf{u}, \mathbf{u} >= \alpha^* < \mathbf{u}, \mathbf{u} > \end{aligned} \tag{3.22}$$

hence $\alpha = \alpha^*$ and $\alpha \in \mathbb{R}$.

Eignevectors which have different eigenvalues for a self-adjoint operator are orthogonal. Let

$$\hat{A}\mathbf{u} = \alpha\mathbf{u} \qquad \text{and} \qquad \hat{A}\mathbf{u}' = \alpha'\mathbf{u}' \tag{3.23}$$

where $\hat{A}$ is a self-adjoint operator and so $\alpha, \alpha' \in \mathbb{R}$. Then we have

$$< \mathbf{u}, \hat{A}\mathbf{u}' >=< \mathbf{u}, \alpha'\mathbf{u}' >= \alpha' < \mathbf{u}, \mathbf{u}' > \tag{3.24}$$

$$=< \hat{A}\mathbf{u}, \mathbf{u}' >=< \alpha\mathbf{u}, \mathbf{u}' >= \alpha < \mathbf{u}, \mathbf{u}' > \tag{3.25}$$

Therefore,

$$(\alpha' - \alpha) < \mathbf{u}, \mathbf{u}' >= 0 \quad \Rightarrow \quad < \mathbf{u}, \mathbf{u}' >= 0 \qquad \text{if } \alpha \neq \alpha'. \tag{3.26}$$

**Theorem 3.1.1.** *For every self-adjoint operator there exists a complete set of eigenvectors (i.e. a basis of the Hilbert space $H$).*

The basis may be countable[1] or continuous.

[1] Countable means it can be put in one=to-one correspondence with the natural numbers.

### 3.1.3 A Countable Basis.

Let $\{\mathbf{u}_n\}$ denote the eigenvectors of a self-adjoint operator $\hat{A}$, i.e.

$$\hat{A}\mathbf{u}_n = \alpha_n \mathbf{u}_n. \tag{3.27}$$

By the theorem above $\{\mathbf{u}_n\}$ form a basis of $H$, let us suppose that it is a countable basis. Let $\{\mathbf{u}_n\}$ be an orthonormal set such that

$$< \mathbf{u}_n, \mathbf{u}_m >= \delta_{nm}. \tag{3.28}$$

Any state may be written $\psi$ as a linear superposition of eigenvectors

$$\psi = \sum \psi_n \mathbf{u}_n \tag{3.29}$$

so that

$$< \mathbf{u}_m, \psi >=< \mathbf{u}_m, \sum \psi_n \mathbf{u}_n >= \psi_m. \tag{3.30}$$

Let us now adopt the useful bra-ket notation of Dirac where the inner product is denoted by

$$< \mathbf{u}_n, \psi > \to \langle u_n|\psi\rangle \tag{3.31}$$

so that, for example in $\mathbb{C}^n$, vectors are denoted by "kets" e.g.

$$\mathbf{u}_n \to |u_n\rangle \qquad \text{and} \qquad \psi \to |\psi\rangle \tag{3.32}$$

while adjoint vectors become "bras":

$$\mathbf{u}_n^\dagger \to \langle u_n| \qquad \text{and} \qquad \psi^\dagger \to \langle\psi|. \tag{3.33}$$

One advantage of this notation is that, being based around the Hilbert space inner product, it is universal for all explicit realisations of the Hilbert space. However its main advantage is how simple it is to use.

Using equation (3.30) we can rewrite equation (3.29) in the bra-ket notation as

$$\begin{aligned} |\psi\rangle = \sum_n \langle u_n|\psi\rangle |u_n\rangle = \sum_n |u_n\rangle\langle u_n|\psi\rangle \\ \Rightarrow \sum_n |u_n\rangle\langle u_n| = \mathbb{I}_H \end{aligned} \tag{3.34}$$

where $\mathbb{I}_H$ is known as the completenes operator. It is worth comparing with $\mathbb{R}^n$ where the identity matrix can be written $\sum_n \mathbf{e}_n \mathbf{e}_n^T = \mathbb{I}$ where $\mathbf{e}_n$ are the usual orthonormal basis vectors for $\mathbb{R}^n$ with zeroes in all compenents except the $n$'th which is one.

Using the properties of the Hilbert space inner product we observe that

$$\psi^* = \overline{\langle u_n|\psi\rangle} = \langle\psi|u_n\rangle \tag{3.35}$$

and further note that this is consistent with the insertion of the completeness operator between two states

$$\langle\phi|\psi\rangle = \sum_n \langle\phi|u_n\rangle\langle u_n|\psi\rangle = \sum_n \phi_n^*\psi_n. \tag{3.36}$$

We may insert a general operator $\hat{B}$ between two states:

$$< \phi, \hat{B}\psi >= \langle\phi|\hat{B}|\psi\rangle = \sum_{n,m}\langle\phi|u_n\rangle\langle u_n|\hat{B}|u_m\rangle\langle u_m|\psi\rangle = \sum_{n,m}\phi_n^* B^m{}_n\psi_m \tag{3.37}$$

where $B^m{}_n$ are the matrix components of the operator $\hat{B}$ written in the $\mathbf{u}_n$ basis. For example as $\mathbf{u}_n$ are eigenvectors of $\hat{A}$ with eigenvalues $\alpha_n$ then the matrix components $A^m{}_n$ are

$$\hat{A} = \begin{pmatrix} \alpha_1 & 0 & \dots & 0 \\ 0 & \alpha_2 & \dots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \dots & \alpha_n \end{pmatrix} \qquad \text{i.e.} \quad A^m{}_n = \alpha_n\delta_{nm}. \tag{3.38}$$

**Theorem 3.1.2.** *Given any two commuting self-adjoint operators $\hat{A}$ and $\hat{B}$ one can find a basis $\mathbf{u}_n$ such that $\hat{A}$ and $\hat{B}$ are simultaneously diagonalisable.*

*Proof.* As $\hat{A}$ is self-adjoint one can find a basis $\mathbf{u}_n$ such that

$$\hat{A}\mathbf{u}_n = \alpha_n\mathbf{u}_n. \tag{3.39}$$

Now

$$\hat{A}\hat{B}\mathbf{u}_n = \hat{B}\hat{A}\mathbf{u}_n = \alpha_n\hat{B}\mathbf{u}_n \tag{3.40}$$

as $[\hat{A}, \hat{B}] = 0$ and hence $\hat{B}\mathbf{u}_n$ is in the eigenspace of $\hat{A}$ (i.e. $\hat{B}\mathbf{u}_n = \sum_m \beta_m\mathbf{u}_m$) and has eigenvalue $\alpha_n$ hence

$$\hat{B}\mathbf{u}_n = \beta_n\mathbf{u}_n. \tag{3.41}$$

□

**Example: Position operators in $\mathbb{R}^3$.**

Let $(\hat{x}, \hat{y}, \hat{z})$ be the position operators of a particle moving in $\mathbb{R}^3$ then

$$[\hat{x}, \hat{y}] = 0, \quad [\hat{x}, \hat{z}] = 0 \quad \text{and} \quad [\hat{y}, \hat{z}] = 0 \tag{3.42}$$

using the canonical quantum commutation rules and hence are simultaneously diagonalisable. One can say the same for $\hat{p}_x$, $\hat{p}_y$ and $\hat{p}_z$.

**The Probabilistic Interpretation in a Countable Basis.**

The Born rule gives the probability that a measurement of a quantum system will yield a particular result. It was first evoked by Max Born in 1926 and it was principally for this work that in 1954 he was awarded the Nobel prize. It states that if an observable associated with a self-adjoint operator $\hat{A}$ then the measured result will be one of the eigenvalues $\alpha_n$ of $\hat{A}$. Further it states that the probability that the measurement of $|\psi\rangle$ will be $\alpha_n$ is given by

$$P(\psi, \mathbf{u}_n) \equiv \frac{\langle\psi|\hat{P}_n|\psi\rangle}{\langle\psi|\psi\rangle} \tag{3.43}$$

where $\hat{P}_n$ is a projection onto the eigenspace spanned by the normalised eigenvector $\mathbf{u}_n$ of $\hat{A}$, i.e. $\hat{P}_n = |u_n\rangle\langle u_n|$ giving

$$P(\psi, \mathbf{u}_n) \equiv \frac{\langle\psi|u_n\rangle\langle u_n|\psi\rangle}{\langle\psi|\psi\rangle} = \frac{|\langle\psi|u_n\rangle|^2}{\langle\psi|\psi\rangle}. \tag{3.44}$$

Note that if the state $\psi$ was an eigenstate of $\hat{A}$ (i.e. $\psi = \psi_n \mathbf{u}_n$) then $P(\psi, \mathbf{u}_n) = 1$. Following a measurement of a state the wavefunction "collapses" to the eigenstate that was measured. Given the probability of measuring a system in a particular eigenstate one can evaluate the expected value when measuring an observable. The expected value is a weighted average of the measurements (eigenvalues) where the weighting is in proportion to the probability of observing each eigenvalue. That is we may measure the observable associated with the operator $\hat{A}$ of a state $\psi$ and find that $\alpha_n$ occurs with probability $P(\psi, \mathbf{u_n})$ then the expected value for measuring $\hat{A}$ is

$$\langle \hat{A} \rangle_\psi = \sum_n \alpha_n P(\psi, \mathbf{u_n}) \tag{3.45}$$

Now given that $\hat{A}|u_n\rangle = \alpha_n |u_n\rangle$ we have that the expectation value of a measurement of the observable associated to $\hat{A}$ is

$$\langle \hat{A} \rangle_\psi = \sum_n \alpha_n \frac{|\langle \psi | u_n \rangle|^2}{\langle \psi | \psi \rangle} = \sum_{n,m} \frac{\langle \psi | u_n \rangle \langle u_n | \hat{A} | u_m \rangle \langle u_m | \psi \rangle}{\langle \psi | \psi \rangle} = \frac{\langle \psi | \hat{A} | \psi \rangle}{\langle \psi | \psi \rangle} \tag{3.46}$$

where we have used $\langle u_n | u_m \rangle = \delta_{nm}$. If $\psi$ is a normalised state then $\langle \hat{A} \rangle_\psi = \langle \psi | \hat{A} | \psi \rangle$. The next most reasonable question we should ask ourselves at this point is what is the probability of measuring the observable of a self-adjoint operator $\hat{B}$ which does not share the eigenvectors of $\hat{A}$, i.e. what does the Born rule say about measuring observables of operators which do not commute? The answer will lead to Heisenberg's uncertainty principle, which we relegate to a (rather long) problem.

**Problem 3.1.1.** *The expectation (or average) value of a self-adjoint operator $\hat{A}$ acting on a normalised state $|\psi\rangle$ is defined by*

$$A_{avg} = \langle \hat{A} \rangle \equiv \langle \psi | \hat{A} | \psi \rangle. \tag{3.47}$$

*The uncertainty in the measurement of $\hat{A}$ on the state $|\psi\rangle$ is the average value of its deviation from the mean and is defined by*

$$\Delta A \equiv \sqrt{\langle (A - A_{avg})^2 \rangle} = \sqrt{\langle \psi | (\hat{A} - A_{avg} \hat{\mathbb{I}})^2 | \psi \rangle} \tag{3.48}$$

*where $\hat{\mathbb{I}}$ is the completeness operator.*

*(a.) Show that for any two self-adjoint operators $\hat{A}$ and $\hat{B}$*

$$|\langle \psi | \hat{A}\hat{B} | \psi \rangle|^2 \leq \langle \psi | \hat{A}^2 | \psi \rangle \langle \psi | \hat{B}^2 | \psi \rangle. \tag{3.49}$$

*Hint: Use the Schwarz inequality: $| < x, y > |^2 \leq < x, x >< y, y >$ where $x, y$ are vectors in a space with inner product $<, >$.*

*(b.) Show that $\langle AB + BA \rangle$ is real and $\langle AB - BA \rangle$ is imaginary when $\hat{A}$ and $\hat{B}$ are self-adjoint operators.*

*(c.) Prove the triangle inequality for two complex numbers $z_1$ and $z_2$:*

$$|z_1 + z_2|^2 \leq (|z_1| + |z_2|)^2. \tag{3.50}$$

*(d.) Use the triangle inequality and the inequality from part (a.) to show that*

$$|\langle\psi|[\hat{A},\hat{B}]|\psi\rangle|^2 \leq 4\langle\psi|\hat{A}^2|\psi\rangle\langle\psi|\hat{B}^2|\psi\rangle. \tag{3.51}$$

*(e.) Define the operators $\hat{A}' \equiv \hat{A} - \alpha\hat{\mathbb{I}}$ and $\hat{B}' \equiv \hat{B} - \beta\hat{\mathbb{I}}$ where $\alpha, \beta \in \mathbb{R}$. Show that $\hat{A}'$ and $\hat{B}'$ are self-adjoint and that $[\hat{A}', \hat{B}'] = [A, B]$.*

*(f.) Use the results to show the uncertainty relation:*

$$(\Delta A)(\Delta B) \geq \frac{1}{2}|\langle\psi|[\hat{A},\hat{B}]|\psi\rangle| \tag{3.52}$$

*What does this give when $\hat{A} = \hat{q}$ and $\hat{B} = \hat{p}$?*

### 3.1.4 A Continuous Basis.

If an operator $\hat{A}$ has eigenstates $\mathbf{u}_\alpha$ where the eigenvalue $\alpha$ is a continuous variable then an arbitrary state in the Hilbert space is

$$|\psi\rangle \equiv \int d\alpha\psi_\alpha|u_\alpha\rangle. \tag{3.53}$$

Then

$$\langle u_\beta|\psi\rangle = \int d\alpha\langle u_\beta|u_\alpha\rangle\psi_\alpha = \psi_\beta. \tag{3.54}$$

The mathematical object that satisfies the above statement is the Dirac delta function:

$$\langle u_\alpha|u_\beta\rangle \equiv \delta(\alpha - \beta). \tag{3.55}$$

Formally the Dirac delta function is a distributon or measure that is equal to zero everywhere apart from 0 when $\delta(0) = \infty$. Its defining property is that its integral over $\mathbb{R}$ is one. One may regard it as the limit of a sequence of Gaussian functions of width $a$ having a maximum at the origin, i.e.

$$\delta_a(x) \equiv \frac{1}{a\sqrt{\pi}}\exp(-\frac{x^2}{a^2}) \tag{3.56}$$

so that as $a \to 0$ the limit of the Gaussians is the Dirac delta function as

$$\int_{-\infty}^{\infty}\delta_a(x)dx = \int_{-\infty}^{\infty}\frac{1}{a\sqrt{\pi}}\exp(-\frac{x^2}{a^2})dx = (\frac{1}{a\sqrt{\pi}})\sqrt{\pi}a = 1 \tag{3.57}$$

which is unchanged when we take the limit $a \to 0$ and so in the limit has the properties of the Dirac delta function. We recall that the Gaussian integral

$$I \equiv \int_{-\infty}^{\infty}dx\exp(-\frac{x^2}{a^2}) \tag{3.58}$$

gives

$$I^2 \equiv \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}dxdy\exp(-\frac{x^2+y^2}{a^2}) = \int_0^{2\pi}\int_0^{\infty}rdrd\theta\exp(-\frac{r^2}{a^2}) \tag{3.59}$$

$$= \int_0^{2\pi}d\theta\Big[-\frac{a^2}{2}\exp(-\frac{r^2}{a^2})\Big]_0^{\infty} \tag{3.60}$$

$$= \int_0^{2\pi}d\theta\frac{a^2}{2} \tag{3.61}$$

$$= \pi a^2 \tag{3.62}$$

hence

$$I = a\sqrt{\pi}. \tag{3.63}$$

As a consequence the eigenstate $|u_\alpha\rangle$ on its own is not correctly normalised to be a vector in the Hilbert space as

$$\langle u_\alpha|u_\beta\rangle = \delta(\alpha-\beta) \Rightarrow \langle u_\alpha|u_\alpha\rangle = \infty \tag{3.64}$$

however used within an integral it is a normalised eigenvector for $\hat{A}$ in the Hilbert space:

$$\int d\alpha \, \langle u_\alpha|u_\alpha\rangle = 1. \tag{3.65}$$

We can show that the continuous eigenvectors form a complete basis for the Hilbert space as

$$\begin{aligned}
\langle\phi|\psi\rangle &= \int\int d\alpha \, d\beta \, \langle u_\alpha|\phi_\alpha^* \psi_\beta|u_\beta\rangle \\
&= \int\int d\alpha \, d\beta \, \langle u_\alpha|\langle\phi|u_\alpha\rangle\langle u_\beta|\psi\rangle|u_\beta\rangle \\
&= \int\int d\alpha \, d\beta \, \langle u_\alpha|u_\beta\rangle\langle\phi|u_\alpha\rangle\langle u_\beta|\psi\rangle \\
&= \int\int d\alpha \, d\beta \, \delta(\alpha-\beta)\langle\phi|u_\alpha\rangle\langle u_\beta|\psi\rangle \\
&= \int d\alpha \langle\phi|u_\alpha\rangle\langle u_\alpha|\psi\rangle
\end{aligned} \tag{3.66}$$

hence we find the completeness relation for a continuous basis:

$$\int d\alpha |u_\alpha\rangle\langle u_\alpha| = \mathbb{I}_H \tag{3.67}$$

**The Probabilistic Interpretation in a Continuous Basis.**

The formulation of Born's rule is only slightly changed in a continuous basis. It now is stated as the probability of finding a system described by a state $|\psi\rangle$ to lie in the range of eigenstates between $|u_\alpha\rangle$ and $|u_{\alpha+\Delta\alpha}\rangle$ is

$$P(\psi, \mathbf{u}_\alpha) = \int_\alpha^{\alpha+\Delta\alpha} d\alpha \frac{\langle\psi|u_\alpha\rangle\langle u_\alpha|\psi\rangle}{\langle\psi|\psi\rangle} = \int_\alpha^{\alpha+\Delta\alpha} d\alpha \frac{|\psi_\alpha|^2}{\langle\psi|\psi\rangle} \tag{3.68}$$

**Transformations between Different Bases**

We finish this section by demonstrating how a state $|\psi\rangle \in H$ may be expressed using different bases for $H$ by using the completeness relation. In particular we show how one may relate a discrete basis of eigenstates to a continuous basis of eigenstates.

Let $\{|u_n\rangle\}$ be a countable basis for $H$ and let $\{|v_\alpha\rangle\}$ be a continuous basis, then:

$$\langle u_n|\psi\rangle = \psi_n \qquad \text{and} \qquad \langle v_\alpha|\psi\rangle = \psi_\alpha. \tag{3.69}$$

Hence we may expand each expression using the completeness operator for the alternative basis to find:

$$\begin{aligned}
\psi_\alpha &= \langle v_\alpha|\psi\rangle \\
&= \sum_n \langle v_\alpha|u_n\rangle\langle u_n|\psi\rangle \\
&= \sum_n u_n(\alpha)\psi_n
\end{aligned} \tag{3.70}$$

where $u_n(\alpha) \equiv \langle v_\alpha | u_n \rangle$, and similarly,

$$\begin{aligned} \psi_n &= \langle u_n | \psi \rangle \\ &= \int d\alpha \, \langle u_n | v_\alpha \rangle \langle v_\alpha | \psi \rangle \\ &= \int d\alpha \, u_n^*(\alpha) \psi_\alpha. \end{aligned} \tag{3.71}$$

## 3.2 The Schrödinger Equation.

Schrödinger developed a wave equation for quantum mechanics by building upon de Broglie's wave-particle duality. Just as the (dynamical) time-evolution of a system represented in phase space is given by Hamilton's equations, so the time evolution of a quantum system is described by Schrödinger's equation:

$$i\hbar \frac{\partial \psi}{\partial t} = \hat{\mathcal{H}} \psi \tag{3.72}$$

A typical Hamiltonian in position space has the form

$$\hat{\mathcal{H}} = -\frac{\hbar^2}{2} \sum_{i=1}^{n} \frac{1}{m_i} \frac{\partial^2}{\partial q_i^2} + \sum_{i=1}^{n} V_i(q) \tag{3.73}$$

where $V(q) = V(q_1, q_2, \ldots q_n)$ and is Hermitian[2]. We will make use of the Hamiltonian in this form in the following.

**Theorem 3.2.1.** *The inner product on the Hilbert space is time-indpendent.*

*Proof.* We will prove this for the $L^2$ norm and use the form of the Hamiltonian $\hat{\mathcal{H}}$ given above. As

$$\langle \psi | \phi \rangle = \int_{\mathbb{R}^k} d^k q \, \psi_q^* \phi_q \tag{3.74}$$

we have

$$\begin{aligned} \frac{\partial}{\partial t} \langle \psi | \phi \rangle &= \int_{\mathbb{R}^k} d^k q \left( \frac{\partial \psi_q^*}{\partial t} \phi_q + \psi_q^* \frac{\partial \phi_q}{\partial t} \right) \\ &= \int_{\mathbb{R}^k} d^k q \left( \frac{i}{\hbar} (\hat{\mathcal{H}}^* \psi_q^*) \phi_q - \frac{i}{\hbar} \psi_q^* (\hat{\mathcal{H}} \phi_q) \right) \end{aligned} \tag{3.75}$$

where we have used Schrödinger's equation and its complex conjugate: $-i\hbar \frac{\partial \psi^*}{\partial t} = \hat{\mathcal{H}}^* \psi^*$.

[2]This guarantees that the energy eigenstates have real eigenvalues and form a basis of the Hilbert space. We will only consider Hermitian Hamiltonians in this course. However while it is conventional to consider only Hermitian Hamiltonians it is by no means a logical consequence of canonical quantisation and one should be aware that non-Hermitian Hamiltonians are discussed occasionally at research level see for example the recent work of Professor Carl Bender.

As $\hat{\mathcal{H}}$ is Hermitian we have $\hat{\mathcal{H}}^* = \hat{\mathcal{H}}$ and so,

$$\begin{aligned}
\frac{\partial}{\partial t}\langle\psi|\phi\rangle &= \int_{\mathbb{R}^k} d^k q \left(\frac{i}{\hbar}(-\frac{\hbar^2}{2}\sum_{i=1}^{n}\frac{1}{m_i}\frac{\partial^2\psi_q^*}{\partial q_i^2} + \sum_{i=1}^{n} V_i(q)\psi_q^*)\phi_q \right. \\
&\qquad \left. - \frac{i}{\hbar}\psi_q^*(-\frac{\hbar^2}{2}\sum_{i=1}^{n}\frac{1}{m_i}\frac{\partial^2\phi_q}{\partial q_i^2} + \sum_{i=1}^{n} V_i(q)\phi_q)\right) \qquad (3.76) \\
&= -\frac{i\hbar}{2}\int_{\mathbb{R}^k} d^k q \sum_{i=1}^{n}\frac{1}{m_i}\left(\frac{\partial^2\psi_q^*}{\partial q_i^2}\phi_q - \psi_q^*\frac{\partial^2\phi_q}{\partial q_i^2}\right) \\
&= -\frac{i\hbar}{2}\int_{\mathbb{R}^k} d^k q \sum_{i=1}^{n}\frac{1}{m_i}\left(-\frac{\partial\psi_q^*}{\partial q_i}\frac{\partial\phi_q}{\partial q_i} + \frac{\partial\psi_q^*}{\partial q_i}\frac{\partial\phi_q}{\partial q_i}\right) \\
&\qquad - \frac{i\hbar}{2}\left[\sum_{i=1}^{n}\frac{1}{m_i}\left(\frac{\partial\psi_q^*}{\partial q_i}\phi_q - \psi_q^*\frac{\partial\phi_q}{\partial q_i}\right)\right]_{\mathbb{R}^k} \\
&= -\frac{i\hbar}{2}\left[\sum_{i=1}^{n}\frac{1}{m_i}\left(\frac{\partial\psi_q^*}{\partial q_i}\phi_q - \psi_q^*\frac{\partial\phi_q}{\partial q_i}\right)\right]_{\mathbb{R}^k} \\
&= 0
\end{aligned}$$

if the boundary term vanishes: typically well-behaved wavefunctions which have compact support on $H$ will vanish at $\pm\infty$. So to complete the proof we have assumed that both the wavefunctions go to zero while their first-derivatives remain finite at infinity. □

From the calculation above we see that the probability density $\rho \equiv \psi^*\psi$ (N.B. just the integrand above) for a wavefuntion $\psi$, which was used to normalise the probability expressed by Born's rule, is conserved, up to a probability current $J^i$ corresponding to the boundary term above:

$$\frac{\partial\rho}{\partial t} = \frac{\partial}{\partial q_i}\left[-\frac{i\hbar}{2}\sum_{i=1}^{n}\frac{1}{m_i}\left(\frac{\partial\psi_q^*}{\partial q_i}\psi_q - \psi_q^*\frac{\partial\psi_q}{\partial q_i}\right)\right] \equiv -\sum_{i=1}^{n}\frac{\partial J^i}{\partial q_i} \qquad (3.77)$$

where $J^i$ is called the probability current and is defined by

$$J^i \equiv \frac{i\hbar}{2m_i}\left(\frac{\partial\psi_q^*}{\partial q_i}\psi_q - \psi_q^*\frac{\partial\psi_q}{\partial q_i}\right). \qquad (3.78)$$

Consequently we arrive at the continuity equation for quantum mechanics

$$\frac{\partial\rho}{\partial t} + \nabla\cdot\mathbf{J} = 0 \qquad (3.79)$$

where $\mathbf{J}$ is the vector whose components are $J^i$.

While the setting was different, we note the similarity in the construction of the equations to the derivation of a conserved charge in Noether's theorem as presented in section 1.3.1.

### 3.2.1 The Heisenberg and Schrödinger Pictures.

Initially the two formulations of quantum mechanics were not understood to be identical. The matrix mechanics of Heisenberg was widely thought to be mathematically abstract while the formulation of a wave equation by Schrödinger although it appeared later was much more quickly accepted as the community of physicists were much more familiar

with wave equations than non-commuting matrix variables. However both formulations were shown to be identical. Here we will discuss the two "pictures" and show the transformations which transform them into each other.

### The Schrödinger Picture

In the Schrödinger picture the states are time-dependent $\psi = \psi(q,t)$ but the operators are not $\frac{d\hat{A}}{dt} = 0$. One can find the time-evolution of the states from the Schrödinger equation:

$$i\hbar\frac{\partial}{\partial t}|\psi(t)\rangle_S = \hat{\mathcal{H}}|\psi(t)\rangle_S \tag{3.80}$$

which has a formal solution

$$|\psi(t)\rangle_S = e^{-\frac{i\hat{\mathcal{H}}t}{\hbar}}|\psi(t)\rangle_S\Big|_{t=0} = e^{-\frac{i\hat{\mathcal{H}}t}{\hbar}}|\psi(0)\rangle_S \tag{3.81}$$

Using the energy eigenvectors (the eigenvectors of the Hamiltonian) as a countable basis for the Hilber space we have

$$|\psi(t)\rangle_S = \sum_n |E_n\rangle\langle E_n|\psi(0)\rangle_S e^{-\frac{i\hat{\mathcal{E}}t}{\hbar}} \tag{3.82}$$

i.e. we have taken $E$ to be the eigenvalue for the Hamiltonian of $\psi(0)_S$: $\hat{\mathcal{H}}|\psi(0)\rangle_S = \sum_n \psi_n^0 E_n|E_n\rangle \equiv E|\psi(0)\rangle_S$ so that $\psi(t) = e^{-\frac{iEt}{\hbar}}|\psi(0)\rangle_S$.

### The Heisenberg Picture

In the Heisenberg picture the states are time-independent but the operators are time-dependent:

$$|\psi\rangle_H = e^{\frac{i\hat{\mathcal{H}}t}{\hbar}}|\psi(t)\rangle_S = |\psi(0)\rangle_S \tag{3.83}$$

while

$$\hat{A}_H(t) = e^{\frac{i\hat{\mathcal{H}}t}{\hbar}}\hat{A}_S e^{-\frac{i\hat{\mathcal{H}}t}{\hbar}}. \tag{3.84}$$

Note that the dynamics in the Heisenberg picture is described by

$$\frac{\partial}{\partial t}\hat{A}_H(t) = \frac{i\hat{\mathcal{H}}}{\hbar}\hat{A}_H(t) - \hat{A}_H(t)\frac{i\hat{\mathcal{H}}}{\hbar} = \frac{i}{\hbar}[\hat{\mathcal{H}}, \hat{A}_H(t)] \tag{3.85}$$

and we note the parallel with the statement from Hamiltonian mechanics that $\frac{df}{dt} = \{f, \mathcal{H}\}$ for a function $f(q,p)$ on phase space.

**Theorem 3.2.2.** *The picture changing transformations leave the inner product invariant.*

*Proof.*

$$_H\langle\phi|\psi\rangle_H =_S \langle\phi|e^{-\frac{i\hat{\mathcal{H}}t}{\hbar}}e^{\frac{i\hat{\mathcal{H}}t}{\hbar}}|\psi\rangle_S =_S \langle\phi|\psi\rangle_S \tag{3.86}$$

□

**Theorem 3.2.3.** *The operator matrix elements are also invariant under teh picture-changing transformations.*

*Proof.*

$$\begin{aligned} {}_H\langle\phi|\hat{A}_H(t)|\psi\rangle_H &=_S \langle\phi|e^{-\frac{i\hat{\mathcal{H}}t}{\hbar}}\hat{A}_H(t)e^{\frac{i\hat{\mathcal{H}}t}{\hbar}}|\psi\rangle_S \\ &=_S \langle\phi|e^{-\frac{i\hat{\mathcal{H}}t}{\hbar}}e^{\frac{i\hat{\mathcal{H}}t}{\hbar}}\hat{A}_S e^{-\frac{i\hat{\mathcal{H}}t}{\hbar}}e^{\frac{i\hat{\mathcal{H}}t}{\hbar}}|\psi\rangle_S \\ &=_S \langle\phi|\hat{A}_S|\psi\rangle_S \end{aligned} \tag{3.87}$$

□

**Example** The Quantum Harmonic Oscillator. The Lagrangian for the harmonic oscillator is

$$\mathcal{L} = \frac{1}{2}m\dot{q}^2 - \frac{1}{2}kq^2 \tag{3.88}$$

The equation of motion is

$$\ddot{q} = -\frac{k}{m}q \tag{3.89}$$

whose solution is

$$q = A\cos(\omega t) + B\sin(\omega t) \tag{3.90}$$

where $\omega = \sqrt{\frac{k}{m}}$. The Legendre transform give the Hamiltonian:

$$\mathcal{H} = \frac{p^2}{2m} + \frac{k}{2}q^2 = \frac{1}{2}m\omega^2 q^2 + \frac{p^2}{2m}. \tag{3.91}$$

The canoonical quantisation procedure gives the quantum hamiltonian for the harmonic oscillator:

$$\hat{\mathcal{H}} = \frac{1}{2}m\omega^2\hat{q}^2 + \frac{\hat{p}^2}{2m}. \tag{3.92}$$

Let us make an inspired change of variables and reqrite the Hamiltonian in terms of

$$\begin{aligned} \alpha &= \sqrt{\frac{m\omega}{2\hbar}}\left(\hat{q} + \frac{i}{m\omega}\hat{p}\right) \\ \alpha^\dagger &= \sqrt{\frac{m\omega}{2\hbar}}\left(\hat{q} - \frac{i}{m\omega}\hat{p}\right) \end{aligned} \tag{3.93}$$

so that

$$\hat{q} = \sqrt{\frac{\hbar}{2m\omega}}\left(\alpha + \alpha^\dagger\right) \qquad \text{and} \qquad \hat{p} = -i\sqrt{\frac{\hbar m\omega}{2}}\left(\alpha - \alpha^\dagger\right). \tag{3.94}$$

Therefore,

$$\begin{aligned} \hat{\mathcal{H}} &= \frac{1}{2}m\omega^2\frac{\hbar}{2m\omega}\left(\alpha + \alpha^\dagger\right)\left(\alpha + \alpha^\dagger\right) - \frac{1}{2m}\frac{\hbar m\omega}{2}\left(\alpha - \alpha^\dagger\right)\left(\alpha - \alpha^\dagger\right) \\ &= \frac{\hbar\omega}{4}\left(\alpha\alpha + \alpha\alpha^\dagger + \alpha^\dagger\alpha + \alpha^\dagger\alpha^\dagger - \alpha\alpha + \alpha\alpha^\dagger + \alpha^\dagger\alpha - \alpha^\dagger\alpha^\dagger\right) \\ &= \frac{\hbar\omega}{2}\left(\alpha\alpha^\dagger + \alpha^\dagger\alpha\right) \end{aligned} \tag{3.95}$$

**Problem 3.2.1.** *Show that* $[\alpha, \alpha^\dagger] = 1$.

Using $[\alpha, \alpha^\dagger] = 1$ we find that

$$\hat{\mathcal{H}} = \hbar\omega\left(\frac{1}{2} + \alpha^\dagger\alpha\right). \tag{3.96}$$

The Hilbert space of states may be constructed as follows. Let $|n\rangle$ be an orthonormal basis such $\hat{\mathcal{H}}$ is diagonalised - i.e. these are the energy eignestates:

$$\hat{\mathcal{H}}|n\rangle \equiv E_n|n\rangle. \tag{3.97}$$

Now we note that

$$\begin{aligned}[\hat{\mathcal{H}}, \alpha^\dagger] &= \frac{1}{2}\omega\hbar\alpha^\dagger + \omega\hbar\alpha^\dagger\alpha\alpha^\dagger - \frac{1}{2}\omega\hbar\alpha^\dagger - \omega\hbar\alpha^\dagger\alpha^\dagger\alpha \\ &= \omega\hbar\alpha^\dagger[\alpha, \alpha^\dagger] \\ &= \omega\hbar\alpha^\dagger\end{aligned} \tag{3.98}$$

and, similarly,

$$[\hat{\mathcal{H}}, \alpha] = -\omega\hbar\alpha. \tag{3.99}$$

Consequently we may deduce that $alpha^\dagger$ raises the eignevalue of the energy eigenstate, while $\alpha$ lowers the energy eigenstates:

$$\begin{aligned}\hat{\mathcal{H}}\alpha^\dagger|n\rangle &= (\alpha^\dagger\hat{\mathcal{H}} + \omega\hbar\alpha^\dagger)|n\rangle = (E_n + \omega\hbar)\alpha^\dagger|n\rangle \\ \hat{\mathcal{H}}\alpha|n\rangle &= (\alpha\hat{\mathcal{H}} - \omega\hbar\alpha)|n\rangle = (E_n - \omega\hbar)\alpha|n\rangle\end{aligned} \tag{3.100}$$

consequently $\alpha^\dagger$ is called the creation operator while $\alpha$ is called the annihilation operator. Together $\alpha$ and $\alpha^\dagger$ are sometimes called the ladder operators.

It would appear that given a single eigenstate the ladder operators create an infinite set of eigenstates, however due to the postive definitieness of the Hilbert space inner product we see that the infinite tower of states must terminate at some point. Consider the length squared of the state $\alpha|n\rangle$:

$$0 \leq \langle n|\alpha^\dagger\alpha|n\rangle = \langle n|\frac{1}{\omega\hbar}\hat{\mathcal{H}} - \frac{1}{2}|n\rangle = \left(\frac{E_n}{\omega\hbar} - \frac{1}{2}\right) \tag{3.101}$$

hence $E_n \geq \frac{1}{2}\omega\hbar$. However the energy eigenvalues of the states $\alpha^k|n\rangle$ are

$$\hat{\mathcal{H}}\alpha^k|n\rangle = (E_n - k\omega\hbar)\alpha^k|n\rangle \tag{3.102}$$

where $k \in \mathbb{Z}$ and $k > 0$. We see that the eigenvalues of the states are continually reduced, but we know that a minimum energy exists ($\frac{1}{2}\omega\hbar$) beyond which the eigenstates will have negative length squared. Consequently we conclude there must exist a ground state eigenfunction $|0\rangle$ such that $\alpha|0\rangle = 0$. In fact if $\alpha|0\rangle = 0$ then

$$\langle 0|\alpha^\dagger\alpha|0\rangle = 0 \quad \Rightarrow \quad E_0 = \frac{1}{2}\omega\hbar. \tag{3.103}$$

Finally we comment on the normalisation of the energy eigenstates. Our aim is to find the normalising constant $\lambda$ where

$$|n-1\rangle = \lambda\alpha|n\rangle. \tag{3.104}$$

Then as both $|n-1\rangle$ and $|n\rangle$ are normalised we have:

$$1 = \langle n-1|n-1\rangle = |\lambda|^2\langle n|\alpha^\dagger\alpha|n\rangle = |\lambda|^2 n\langle n|n\rangle = |\lambda|^2 n \tag{3.105}$$

where we have used the observation that $\alpha^\dagger\alpha$ is the number operator.

**Problem 3.2.2.** *Let the state $|n\rangle$ be interpreted as an n-particle eigenstate with energy $E_n = \frac{1}{2}\omega\hbar + n\omega\hbar$. Show that the number operator $\hat{N} \equiv \alpha^\dagger\alpha$ satisfies:*

$$\langle n|\hat{N}|n\rangle = n \tag{3.106}$$

Hence $\lambda = \frac{1}{\sqrt{n}}$ and $\alpha|n\rangle = \sqrt{n}|n-1\rangle$.

**Problem 3.2.3.** *Show that $\alpha^\dagger|n\rangle = \sqrt{n+1}|n+1\rangle$.*

# Chapter 4

# Group Theory

The first investigations of groups are credited to the famously dead-at-twenty Evariste Galois, who was killed in a duel in 1832. Groups were first used to map solutions of polynomial equations into each other. For example the quadratic equation

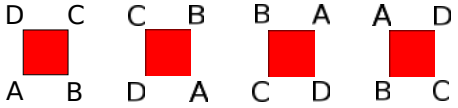
$$y = ax^2 + bx + c \tag{4.1}$$

is solved when $y = 0$ by

$$x = \frac{1}{2a}(-b \pm \sqrt{b^2 - 4ac}). \tag{4.2}$$

It has two solutions ($\pm$) which may be mapped into each other by a $\mathbb{Z}_2$ reflection which swaps the $+$ solution for the $-$ solution. The "$\mathbb{Z}_2$" is the cyclic group of order two (which is sometimes denoted $C_2$ and similarly there exist groups which map the roots of a more general polynomial equation into each other. Groups have a geometrical meaning too. The symmetries which leave unchanged the $n$-polygons under rotation are also the cyclic groups, $\mathbb{Z}_n$ (or $C_n$). For example $\mathbb{Z}_3$ rotates an equilateral triangle into itself using rotations of $\frac{2\pi}{3}$, $\frac{4\pi}{3}$ and $\frac{6\pi}{3} = 2\pi$ about the centre of the triangle and $\mathbb{Z}_4$ is the group of rotations of the square onto itself.

The cyclic groups are examples of discrete symmetry groups. The action of the discrete group takes a system (e.g. the square in $\mathbb{R}^2$) and rotates it onto itself without passing through any of the suspected intervening orientations. The $\mathbb{Z}_4$ group includes the rotation by $\frac{\pi}{2}$ but it does not include any of the rotations through angles less than $\frac{\pi}{2}$ and greater than 0. One may imagine that under the action of $\mathbb{Z}_4$ the square jumps between orientations:




(4.3)

On the other hand continuous groups (such as the rotation group in $\mathbb{R}^2$ move the square continuously about the centre of rotation. The rotation is parameterised by a continuous angle variable, often denoted $\theta$. The Norwegian Sophus Lie began the study of continuous groups, also known as Lie groups, in the second half of the $19^{\text{th}}$ century. Rather than thinking about geometry Sophus Lie was interested in whether there were some groups equivalent to Galois groups which mapped solutions of differential equations

into each other[1]. Such groups were identified, classified and named Lie groups. The rotation group $SO(n)$ is a Lie group.

In the wider context groups may act on more than algebraic equations or geometric shapes in the plane and the action of the group may be encoded in different ways. The study of the ways groups may be represented is aptly named representation theory. It is believed and successfully tested (at the present energies of expereiments) that the constiuent objects in the universe are invariant under certain symmetries. The standard model of particle physics holds that all known particles are representations of $SU(3) \otimes SU(2) \otimes U(1)$. More simply, Einstein's special theory of relativity may be studied as the theory of Lornetz groups.

We will make contact with most of these topics in this chapter and we begin with the preliminaries of group theory: just what is a group?

## 4.1 The Basics

**Definition** A group $G$ is a set of elements $\{g_1, g_2, g_3 \ldots\}$ with a composition law $(\circ)$ which maps $G \times G \to G$ by $(g_1, g_2) \to g_1 \circ g_2$ such that:

(i) $g_1 \circ (g_2 \circ g_3) = (g_1 \circ g_2) \circ g_3 \qquad \forall \quad g_1, g_2, g_3 \in G \qquad$ ASSOCIATIVE

(ii) $\exists \quad e \in G$ such that $e \circ g = g \circ e = g \qquad \forall \quad g \in G \qquad$ IDENTITY

(iii) $\exists \quad g^{-1} \in G$ such that $g \circ g^{-1} = g^{-1} \circ g = e \qquad \forall \quad g \in G \qquad$ INVERSES

Consequently the most trivial group consists of just the identity element $e$. Within the definition above, together with the associative proprty of the group multiplication, the existence of an identity element and an inverse element $g^{-1}$ for each $g$, there is what we might call the zeroth property of a group. namely the closure of the group (that $g_1 \circ g_2 \in G$.

Let us now define some of the most fundamental ideas in group theory.

**Definition** A group $G$ is called *commutative* or *ableian* if $g_1 \circ g_2 = g_2 \circ g_1 \, \forall \, g_1, g_2 \in G$.

**Definition** The *centre* $Z(G)$ of a group is:

$$Z(G) \equiv \{g_1 \in G \,|\, g_1 \circ g_2 = g_2 \circ g_1 \, \forall \, g_2 \in G\} \tag{4.4}$$

The centre of a group is the subset of elements in the group which commute with all other elements in $G$. Trivially $e \in G$ as $e \circ g = g \circ e \, \forall \, g \in G$.

**Definition** The *order* $|G|$ of a group $G$ is the number of elements in the set $\{g_1, g_2, \ldots\}$.

For example the order of the group $\mathbb{Z}_2$ is $|\mathbb{Z}_2| = 2$, we have also seen $|\mathbb{Z}_3| = 3$, $|\mathbb{Z}_4| = 4$ and in general $|\mathbb{Z}_n| = n$, where the elements are the rotations $\frac{m2\pi}{n}$ where $m \in \mathbb{Z} \mod n$.

**Definition** For each $g \in G$ the *conjugacy class* $C_g$ is the subset

$$C_g \equiv \{h \circ g \circ h^{-1} \,|\, h \in G\} \subset G. \tag{4.5}$$

[1]Very loosely, as each solution to a differential equation is correct "up to a constant", the solutions contain a continuous parameter: the constant.

***Exercise*** Show that the identity element of a group $G$ is unique.

***Solution*** Suppose $e$ and $f$ are two distinct identity elements in $G$. Then $e \circ g = f \circ g \Rightarrow e \circ (g \circ g^{-1}) = f \circ (g \circ g^{-1}) \Rightarrow e = f$. Contrary to the supposition.

## 4.2 Common Groups

A list of groups is shown in table 4.2.1, where the set and the group multiplication law have been highlighted.

A few remarks are in order.

- (1,6-10) are finite groups satisfying $|G| < \infty$.
- (14-20) are called the classical groups.
- Groups can be represented by giving their multiplication table. For example consider $\mathbb{Z}_3$:

| | $e$ | $g$ | $g^2$ |
|---|---|---|---|
| $e$ | $e$ | $g$ | $g^2$ |
| $g$ | $g$ | $g^2$ | $e$ |
| $g^2$ | $g^2$ | $e$ | $g$ |

- Arbitrary combinations of group elements are sometimes called words.

### 4.2.1 The Symmetric Group $S_n$

The Symmetric group $S_n$ is the group of permutations of $n$ elements. For example $S_2$ has order $|S_2| = 2!$ and acts on the two elements $((1,2),(2,1))$. The group action is defined element by element and may be written as a two-row matrix with $n$ columns, where the permutation is defined per column with the label in row one being substituted for the label in row two. For $S_2$ consider the group element

$$g_1 \equiv \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}. \tag{4.6}$$

This acts on the elements as

$$g_1 \circ (1,2) = (2,1) \qquad g_1 \circ (2,1) = (1,2) \tag{4.7}$$

$$g_1^2 \circ (1,2) = (1,2) \qquad g_1^2 \circ (2,1) = (2,1) \tag{4.8}$$

hence $g_1 = g_1^{-1}$ and $g_1^2 = e$ and $S_2 \equiv \{e, g_1\}$. It is identical to $\mathbb{Z}_2$.

More generally for the group $S_n$ having $n!$ elements it is denoted by a permutation $P$ such as:

$$P \equiv \begin{pmatrix} 1 & 2 & 3 & \dots & n \\ p_1 & p_2 & p_3 & \cdots & p_n \end{pmatrix} \tag{4.9}$$

where $p_1, p_2, p_3, \dots p_n \in \{1, 2, 3, \dots n\}$. The permutation $P$ takes $(1, 2, 3, \dots, n)$ to $(p_1, p_2, p_3, \dots, p_n)$. In general successive permutations do not commute. For example consider $S_3$ and let

$$P \equiv \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix} \qquad \text{and} \qquad Q \equiv \begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix}. \tag{4.10}$$

| | | |
|---|---|---|
| 1 | $G = \{e\}$ | Under multiplication. |
| 2 | $\{\mathbb{F}\}$ where $\mathbb{F} = \mathbb{Z}, \mathbb{Q}, \mathbb{R}, \mathbb{C}$ | Under addition. |
| 3 | $\{\mathbb{F}^\times \equiv \mathbb{F}\backslash 0\}$ where $\mathbb{F} = \mathbb{Q}, \mathbb{R}, \mathbb{C}$ | Under multiplication. |
| 4 | $\{\mathbb{F}_{>0}\}$ where $\mathbb{F} = \mathbb{Q}, \mathbb{R}$ | An abelian group under multiplication. |
| 5 | $\{0, \pm n, \pm 2n, \pm 3n, \ldots\} \equiv n\mathbb{Z}$ where $n \in \mathbb{Z}$. | An abelian group under addition. |
| 6 | $\{0, 1, 2, 3, \ldots, (n-1)\}$. | Addition $\mod(n)$, e.g. $a + b = c \mod n$. |
| 7 | $\{-1, 1\}$. | Under multiplication. |
| 8 | $\{e, g, g^2, g^3, \ldots g^{n-1}\}$. | With $g^k \circ g^l = g^{(k+l) \mod n}$.<br>This is the cyclic group of order $n$, $\mathbb{Z}_n$. |
| 9 | $S_n$ the symmetric group or<br>permutation group of $n$ elements. | Under the composition of permutations. |
| 10 | $D_n$ the dihedral group.<br>The group of rotations and reflections<br>of an $n$-sided polygon with undirected edges. | Under the composition of permutations.<br>Composition of transformations. |
| 11 | Bijections $f : X \to X$ where $X$ is a set. | Composition of maps. |
| 12 | $GL(V) \equiv \{f : V \to V \mid f$ is linear and invertible$\}$.<br>$V$ is a vector space. | Composition of maps. |
| 13 | A vector space, $V$. | An abelian group under vector addition. |
| 14 | $GL(n, \mathbb{F}) \equiv \{M \in n \times n$ matrices $\mid M$ is invertible.$\}$<br>The general linear group, with matrix entries in $\mathbb{F}$. | Matrix multiplication. |
| 15 | $SL(n, \mathbb{F}) \equiv \{M \in GL(n, \mathbb{F}) \mid \det M = 1\}$<br>The special linear group. | Matrix multiplication. |
| 16 | $O(n) \equiv \{M \in GL(n, \mathbb{R}) \mid M^T M = \mathbb{I}_n\}$<br>The orthogonal group. | Matrix multiplication. |
| 17 | $SO(n) \equiv \{M \in GL(n, \mathbb{R}) \mid \det M = 1\}$<br>The special orthogonal group. | Matrix multiplication. |
| 18 | $U(n) \equiv \{M \in GL(n, \mathbb{C}) \mid M^\dagger M = \mathbb{I}_n\}$<br>The unitary group. | Matrix multiplication. |
| 19 | $SU(n) \equiv \{M \in U(n) \mid \det M = 1\}$<br>The special unitary group. | Matrix multiplication. |
| 20 | $Sp(2n) \equiv \{M \in GL(2n, \mathbb{R}) \mid M^T \mathcal{J} M = \mathcal{J}\}$<br>Where $\mathcal{J} \equiv \begin{pmatrix} 0_n & \mathbb{I}_n \\ -\mathbb{I}_n & 0_n \end{pmatrix}$.<br>The symplectic group. | Matrix multiplication. |
| 21 | $O(p, q) \equiv \{M \in GL(p+q, \mathbb{R}) \mid M^T \eta_{p,q} M = \eta_{p,q}\}$<br>Where $\eta_{p,q} \equiv \begin{pmatrix} \mathbb{I}_p & 0_{p\times q} \\ 0_{p\times q} & -\mathbb{I}_q \end{pmatrix}$. | Matrix multiplication. |
| 22 | $SL(2, \mathbb{Z}) \equiv \{\begin{pmatrix} a & b \\ c & d \end{pmatrix} \mid a, b, c, d \in \mathbb{Z},\ ad - bc = 1\}$<br>The modular group. | Matrix multiplication. |

Table 4.2.1: A list of commonly occurring groups.

Then,

$$Q \circ P = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix} \circ \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix} \tag{4.11}$$

while

$$P \circ Q = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix} \circ \begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix}. \tag{4.12}$$

Hence $P \circ Q \neq Q \circ P$ and $S_3$ is non-abelian.

Alternatively one may denotes each permutation by its disjoint cycles of labels formed by multiple actions of that permutation. For example consider $P \in S_3$ as defined above. Under successive actions of $P$ we see that the label 1 is mapped as:

$$1 \xrightarrow{P} 2 \xrightarrow{P} 3 \xrightarrow{P} 1. \tag{4.13}$$

We may denote this cycle as $(1, 2, 3)$ and it defines $P$ entirely. On the other hand $Q$, as defined above, may be described by two disjoint cycles:

$$1 \xrightarrow{Q} 1 \tag{4.14}$$

$$2 \xrightarrow{Q} 3 \xrightarrow{Q} 2. \tag{4.15}$$

We may write $Q$ as two disjoint cycles $(1), (2, 3)$. In this notation $S_3$ is written

$$\{(), (1, 2), (1, 3), (2, 3), (1, 2, 3), (1, 3, 2)\} \tag{4.16}$$

where () denotes the trivial identity permutation. $S_3$ is identical to the dihedral group $D_3$. The dihedral group $D_n$ is sometimes defined as the symmetry group of rotations of an n-sided polygon with undirected edges - this definition requires a bit of thought, as the rotations may be about an axis through the plane of the polygon and so are reflections. The dihedral group should be compared with cyclic groups $\mathbb{Z}_n$ which are the rotation symmetries of an n-polygon with directed edges, while $D_n$ includes the reflections in the plane as well. For example if we label the vertices of an equilateral triangle by 1, 2 and 3 we could denote $D_3$ as the following permutations of the vertices

$$\begin{aligned} \{ &\begin{pmatrix} 1 & 2 & 3 \\ 1 & 2 & 3 \end{pmatrix}, \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix}, \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix}, \\ &\begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix}, \begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{pmatrix}, \begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{pmatrix} \} \\ &= \{(), (1, 2), (1, 3), (2, 3), (1, 2, 3), (1, 3, 2)\}. \end{aligned} \tag{4.17}$$

So we see that $D_3$ is identical to $S_3$. We see that there are three reflections and three rotations within $D_3$ (the identity element is counted as a rotation for this purpose). In general $D_n$ contains the $n$ rotations of $\mathbb{Z}_n$ as well as reflections. For even $n$ there is an axis in which the reflection is a symmetry which passes through each pair of opposing vertices ($\frac{n}{2}$ and also reflections in the line through the centre of each opposing edge $\frac{n}{2}$. For odd $n$ there are again $n$ lines about which reflection is a symmetry, however these lines now join a vertex to the middle of an opposing edge. In both even and odd cases there are therefore $n$ rotations and $n$ reflections. Hence $|D_n| = 2n$.

We may wonder if all dihedral groups $D_n$ are identical to the permutation groups $S_n$. The answer is no, it was a coincidence that $S_3 \cong D_3$. We can convince ourselves of these by considering the order of $S_n$ and $D_n$. As we have already observed $|S_n| = n!$ while $|D_n| = 2n$. For the groups to be identical we at least require their orders to match and we note that we can only satisfy $n! = 2n$ for $n = 3$.

Returning to the symmetric group we will mention a third important notation for permutations which is used to define symmetric and anti-symmetric tensors. Each permutation P can be written as combinations of elements called transpositions $\tau_{ij}$ which swap elements $i$ and $j$ but leave the remainder untouched. Consequently each transposition may be written as a 2-cycle $\tau_{i,j} = (i, j)$. For example,

$$P \equiv \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix} = \tau_{2,3} \circ \tau_{1,3}. \tag{4.18}$$

If there are $N$ transpositions required to replicate a permutation $P \in S_n$ then the sign of the permuation is defined by

$$\text{Sign}(P) \equiv (-1)^N. \tag{4.19}$$

You should convince yourself that this operation is well-defined and that each permutation $P$ has a unique value of Sign($P$) - this is not obvious as there are many different combinations of the transpositions which give the same overall permutation. The canonical way to decompose permutations into transpositions is to consider only transpositions which interchange consecutive labels, e.g $\tau_{1,2}, \tau_{2,3}, \ldots \tau_{n-1,n}$. A general $r$-cycle may be decomposed (not in the canonical way) into $r - 1$ transpositions:

$$(n_1, n_2, n_3, \ldots n_r) = (n_1, n_2)(n_2, n_3) \ldots (n_{r-1}, n_r) = \tau_{n_1,n_2}\tau_{n_2,n_3} \ldots \circ \tau_{n_{r-1},n_r}. \tag{4.20}$$

Consequently an $r$-cycle corresponds to a permutation $R$ such that $\text{Sign}(R) = (-1)^{(r-1)}$. Therefore the elements of $S_3 \cong D_3$ may be partitioned into those elements of sign 1 $(), (1, 2, 3), (1, 3, 2)$, which geometrically correspond to the rotations of the equilateral triangle in the plane, and those of sign -1 $(1, 2), (2, 3), (1, 3)$ which are the reflections in the plane. The subset of permutations $P \in S_n$ which have Sign($P$)=1 form a sub-group of $S_n$ which is called the *alternating group* and denoted $A_n$.

We finish our discussion of the symmetric group by mentioning Cayley's theorem. It states that every finite group of order $n$ can be considered as a subgroup of $S_n$. Since $S_n$ contains all possible permutations of $n$ labels it is not a surprising theorem.

**Problem 4.2.1.** *$D_n$ is the dihedral group the set of rotation symmetries of an n-polygon with undirected edges.*

*(i.) Write down the multiplication table for $D_3$ defined on the elements {e,a,b} by $a^2 = b^3 = (ab)^2 = e$. Give a geometrical interpretation in terms of the transformations of an equilateral triangle for a and b.*

*(ii.) Rewrite the group multiplication table of $D_3$ in terms of six disjoint cycles given by repeated action of the basis elements on the identity until they return to the identity, e.g. $e \to e$ under the action of e, $e \to a \to e$ under the action of a.*

*(iii.) Label the vertices of the equilateral triangle by* $(1,2,3)$*. Denote the vertices of the triangle by* $(1,2,3)$ *and give permutations of* $\{1,2,3\}$ *for* $e$*,* $a$ *and* $b$ *which match the defining relations of* $D_3$*.*

*(iv.) Rewrite each of the cycles of part (b.) in cyclic notation on the vertices* $(1,2,3)$ *to show this gives all the permutations of* $S_3$*.*

### 4.2.2 Back to Basics

**Definition** A *subgroup* $H$ of a group $G$ is a subset of $G$ such that $e \in H$, if $g_1, g_2 \in H$ then $g_1 \circ g_2 \in H$ and if $g \in H \Rightarrow g^{-1} \in H$ where $g, g_1, g_2, g^{-1} \in G$.

The identity element $\{e\}$ and $G$ itself are called the trivial subgroups of G. If a subgroup $H$ is not one of these two trivial cases then it is called a proper subgroup and this is denoted $H < G$. For example $S_2 < S_3$ as:

$$\begin{aligned} S_2 &= \{(), (1,2)\} \qquad \text{and} \\ S_3 &= \{(), (1,2), (1,3), (2,3), (1,2,3), (1,3,2)\}. \end{aligned} \tag{4.21}$$

**Definition** Let $H < G$. The subsets $g \circ H \equiv \{g \circ h \in G \,|\, h \in H\}$ are called *left-cosets* while the subsets $H \circ g \equiv \{h \circ g \in G \,|\, h \in H\}$ are called *right-cosets*.

The left-coset $g \circ H$ where $g \in G$ contains the elements

$$\{g \circ h_1, g \circ h_2, \ldots, g \circ h_r\} \tag{4.22}$$

where $r \equiv |H|$ and $\{h_1, h_2, \ldots, h_r\}$ are the distinct elements of $H$. One might suppose that $r < |H|$ which could occur if two or more elements of $g \circ H$ were identical, but if that were the case we would have

$$g \circ h_1 = g \circ h_2 \qquad \Rightarrow \qquad h_1 = h_2 \tag{4.23}$$

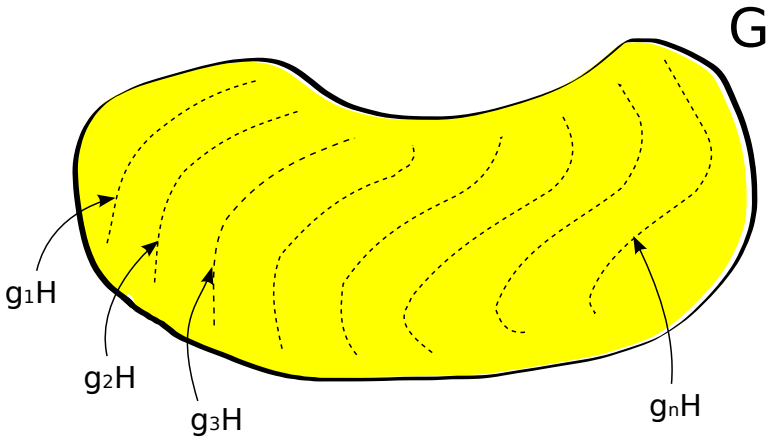
but $h_1$ and $h_2$ are defined to be distinct. Hence all cosets of $G$ have the same number of elements which is $|H|$, the order of $H$.

Consequently any two cosets are either disjoint or coincide. For example, consider the two left-cosets $g_1 \circ H$ and $g_2 \circ H$ and suppose that there existed some element $g$ in the intersection of both cosets, i.e. $g \in g_1 \circ H \cap g_2 \circ H$. In this case we would have $g = g_1 \circ h_1 = g_2 \circ h_2$ for some $h_1, h_2 \in H$. Then,

$$g_1 \circ H = (g \circ h_1^{-1}) \circ H = g \circ H = g \circ (h_2^{-1} \circ H) = g_2 \circ H. \tag{4.24}$$

Hence either the cosets are disjoint or if they do have a non-zero intersection they are

in fact coincident. This means that the cosets provide a disjoint partition of $G$




(4.25)

hence

$$|G| = n|H| \tag{4.26}$$

for some $n \in \mathbb{Z}$. This statement is known as Lagrange's theorem which states that the order of any subgroup of $G$ must be a divisor of $|G|$.

A corollary of Lagrange's theorem is that groups of prime order have no proper subgroups (e.g. $\mathbb{Z}_n$ where $n$ is prime).

**Definition** $H < G$ is called a *normal subgroup* of $G$ if

$$g \circ H = H \circ g \tag{4.27}$$

$\forall\, g \in G$. This is denoted $H \lhd G$.

The definition of a normal subgroup is equivalent to saying that $g \circ H \circ g^{-1} = H$, or that $g$ commutes with all of $H$. In particular the largest non-trivial normal subgroup is the centre $Z(G)$ of $G$, i.e. $Z(G) \lhd G$.

**Definition** $G$ is called a *simple group* is it has no non-trivial normal subgroups (i.e. besides $\{e\}$ and $G$ itself).

**Theorem 4.2.1.** *If $H \lhd G$ then the set of cosets $\frac{G}{H}$ is itself a group with composition law*

$$(g_1 \circ H) \circ (g_2 \circ H) = (g_1 \circ g_2) \circ H \qquad \forall\ g_1, g_2 \in G. \tag{4.28}$$

This group is called the quotient group, or factor group, and denoted $\frac{G}{H}$.

*Proof.* Evidently it is closed as the group action takes $g \circ H \times g \circ H \to g \circ H$. Let us check the three axioms that define a group.

(i.) Associativity:

$$\begin{aligned}(g_1 \circ H) \circ ((g_2 \circ H) \circ (g_3 \circ H)) &= (g_1 \circ H) \circ (g_2 \circ g_3) \circ H \\ &= (g_1 \circ (g_2 \circ g_3)) \circ H \\ &= ((g_1 \circ g_2) \circ g_3) \circ H \\ &= ((g_1 \circ g_2) \circ H) \circ (g_3 \circ H) \\ &= ((g_1 \circ H) \circ (g_2 \circ H)) \circ (g_3 \circ H)\end{aligned} \tag{4.29}$$

(ii.) Identity. The coset $e \circ H$ acts as the identity element:

$$(e \circ H) \circ (g \circ H) = (g \circ H) \circ (e \circ H) = (e \circ g) \circ H = g \circ H \tag{4.30}$$

(iii.) Inverse. The inverse of the coset $g \circ H$ is the coset $g^{-1} \circ H$ as:

$$(g \circ H) \circ (g^{-1} \circ H) = e \circ H = H \tag{4.31}$$

□

N.B. that the group composition law arises as $H \lhd G$ so $g_1 \circ H \circ g_2 \circ H = g_1 \circ g_2 \circ H$.

## 4.3 Group Homomorphisms

Maps between groups are incredibly useful in recognising similar groups and constructing new groups.

**Definition** A *group homomorphism is a map* $f : G \to G'$ *between two groups* $(G, \circ)$ *and* $(G', \circ')$ *such that*

$$f(g_1 \circ g_2) = f(g_1) \circ' f(g_2) \qquad \forall\, g_1, g_2 \in G \tag{4.32}$$

**Definition** A group *isomorphism is an invertible group homomorphism.*

If an isomorphism exists between $G$ and $G'$ we write $G \cong G'$ and say that '$G$ is isomorphic to $G'$'.

**Definition** A group *automorphism is an isomorphism* $f : G \to G$.

**Problem 4.3.1.** *If* $f : G \to G'$ *is a group homomorphism between the groups* $G$ *and* $G'$*, show that*

*(i.)* $f(e) = e'$*, where* $e$ *and* $e'$ *are the identity elements of* $G$ *and* $G'$ *respectively, and*

*(ii.)* $f(g^{-1}) = (f(g))^{-1}$.

**Theorem 4.3.1.** *If* $f : G \to G'$ *is a group homomorphism then the kernel of* $f$*, defined as* $Ker(f) \equiv \{g \in G | f(g) = e'\}$ *is a normal subgroup of* $G$.

**Problem 4.3.2.** *Prove Theorem 4.3.1.*

The theorem above can be used to prove that $\frac{G}{Ker(f)} \cong G'$ for a given group homomorphism $f : G \to G'$, or conversely given an isomorphism between $\frac{G}{Ker(f)}$ and $G'$ to identify the group homomorphism $f$ (see section 4.3.1). A corollary of the theorem above is that simple groups, having no non-trivial sub-groups, admit only trivial homomorphisms, i.e. those for which $Ker(f) = G$ or $Ker(f) = \{e\}$.

**Comments**

- $(n\mathbb{Z}, +)$ are abelian groups and hence normal subgroups of $\mathbb{Z}$: $n\mathbb{Z} \lhd \mathbb{Z}$.
- $(\mathbb{F}_{>0}, \times) \lhd (\mathbb{F}^{\times}, \times)$.
- Group 6 in table 4.2.1 $(\{0, 1, 2, 3, \ldots, (n-1)\}, +\mod(n))$ is isomorphic to group 8 $(\{e, g, g^2, g^3, \ldots g^{n-1}\}, g^k \circ g^l = g^{(k+l) \mod n})$, with the group isomorphism being $f(1) = g$.
- $D_n < S_n$ and $D_n$ is not a normal subgroup in general.
- $\text{Sign}(P \in S_n) \to \mathbb{Z}_2$ is a group homomorphism. Consequently the alternating group $A_n \equiv (P \in S_n, \text{Sign}(P) = 1)$ is a normal subgroup of $S_n$ as $A_n \equiv Ker(\text{Sign})$.
- The determinant, Det is a group homomorphism: $\text{Det}(GL(n, \mathbb{F})) \to (\mathbb{F}^{\times}, \times)$. Hence:

  - $SL(n, \mathbb{F}) \lhd GL(n, \mathbb{F})$ as $SL(n, \mathbb{F}) \equiv Ker(\text{Det})$,
  - $SO(n) \lhd O(n)$ and
  - $SU(n) \lhd U(n)$.

  And so

  - $\frac{GL(n,\mathbb{F})}{SL(n,\mathbb{F})} \cong (\mathbb{F}^{\times}, \times)$,
  - $\frac{O(n)}{SO(n)} \cong \mathbb{Z}_2$ and
  - $\frac{U(n)}{SU(n)} \cong U(1) \equiv \{z \in \mathbb{C}, |z| = 1\}$.
- The centre of $SU(2)$ denoted $Z(SU(2)) = \mathbb{Z}_2$ and one can show that the coset group $\frac{SU(2)}{\mathbb{Z}_2} \cong SO(3)$.

There are a number of simple ways to create new groups from known groups for example:

(1.) Given a group $G$, identify a subgroup $H$. If these are normal $H \lhd G$ then $\frac{G}{H}$ is a group.

(2.) Given two groups $G$ and $G'$, find a group homomorphism $F : G \to G'$ such that $Ker(f) \lhd G$ then $\frac{G}{Ker(f)} \cong G'$ and we observe as a corollary that $Ker(f)$ is a group.

(3.) One can form the direct product of groups to create more complicated groups. The direct product of two groups $G$ and $H$ is denoted $G \times H$ and has composition law:

$$(g_1, h_1) \circ' (g_2, h_2) \equiv (g_1 \circ_G g_2, h_1 \circ_H h_2) \tag{4.33}$$

where $g_1, g_2 \in G$, $h_1, h_2 \in H$, $\circ_G$ is the composition law on $G$ and $\circ_H$ is the composition law on $H$. E.g. the direct product $\mathbb{R} \times \mathbb{R}$ has the compsition law corresponding to two-dimensional real vector addition, i.e. $(x_1, y_1) + (x_2, y_2) = (x_1 + x_2, y_1 + y_2)$. The direct product of a group $G$ with itself $G \times G$ has a natural subgroup $\Delta(G)$ called *the diagonal* and defined by $\Delta(G) \equiv \{(g, g) \in G \times G | g \in G\}$.

(4.) If $X$ is a set and $G$ a group such that there exists a map $f : X \to G$ then the functions $f$ with the composition law

$$f_1 \circ' f_2(x) \equiv f_1(x) \circ_G f_2(x) \tag{4.34}$$

where $x \in X$ form a group. For example if $X = S^1$ the set of maps of $X$ into $G$ form the 'loop group' of $G$.

There are only a finite number of finite simple groups. The quest to identify them all is universally accepted as having been completed in the 1980's. In addition to groups such as the cyclic groups $\mathbb{Z}_n$, the symmetric group $S_n$, the dihedral group $D_n$ and the alternating group $A_n$ there are fourteen other infinite series and twenty-six other 'sporadic groups'. These include:

- The Matthieu groups (e.g. $|M_{24}| = 2^1 0.3^3.5.7.11.23 = 244,823,040$),
- the Janko groups (e.g. $|J_4| \approx 8.67 \times 10^{19}$),
- the Conway groups (e.g. $|Co_1| \approx 4.16 \times 10^{18}$),
- the Fischer groups (e.g. $|Fi_{24}| \approx 1.26 \times 10^{24}$) and
- the Monster group ($|M| \approx 8.08 \times 10^{53}$).

**Definition** Let $G$ be a group and $X$ be a set. The (left) action of $G$ on $X$ is a map taking $G \times X \to X$ and denoted[2]

$$(g, x) \to g \circ x \equiv T_g(x) \tag{4.35}$$

that satisfies

(i.) $(g_1 \circ g_2) \circ x = g_1 \circ (g_2 \circ x) \quad \forall g_1, g_2 \in G, x \in X$

(ii.) $e \circ x = x \quad \forall x \in X$ where $e$ is the identity element in $G$.

The set $X$ is called a (left) $G$-set.

**Definition** The *orbit* of $x \in X$ under the $G$-action is

$$G \circ x \equiv \{x' \in X | x' = g \circ x \quad \forall g \in G\}. \tag{4.36}$$

**Definition** The *stabiliser subgroup* of $x \in X$ is the group of all $g \in G$ such that $g \circ x = x$, i.e.

$$G_x \equiv \{g \in G | g \circ x = x\}. \tag{4.37}$$

**Definition** The *fundamental domain* is the subset $X_F \subset X$ such that

(i.) $x \in X_F \quad \Rightarrow g \circ x \notin X_F \quad g \in G \backslash \{e\}$ and

(ii.) $X = \cup_{g \in G} g \circ X_F$.

[2]Here we use $T_g$ to denote the left-translation by $g$, but we could similarly define the right-translation with the group element acting on the set from the right-hand-side.

**Examples**

(1.) $S_n$ acts on the set $\{1, 2, 3, \ldots n\}$.

(2.) A group $G$ can act on itself in three canonical ways:

(i.) left translation: $T^{(L)}_{g_1}(g_2) = g_1 \circ g_2$,

(ii.) right translation: $T^{(R)}_{g_1}(g_2) = g_2 \circ g_1$ and

(iii.) by conjugation[3]: $T^{(R)}_{g_1^{-1}} T^{(L)}_{g_1}(g_2) = g_1 \circ g_2 \circ g_1^{-1} \equiv Ad_{g_1}(g_2)$.

(3.) $SL(2, \mathbb{Z})$ acts on the set of points in the upper half-plane $H \equiv \{z \in \mathbb{C} | \mathrm{Im}(z) > 0\}$ by the Möbius transformations:

$$\left( \begin{pmatrix} a & b \\ c & d \end{pmatrix}, z \right) \rightarrow \frac{az + b}{cz + d} \in H \tag{4.38}$$

**Problem 4.3.3.** *Consider the Klein four-group, $V_4$, (named after Felix Klein) consisting of the four elements $\{e, a, b, c\}$ and defined by the relations:*

$$a^2 = b^2 = c^2 = e, \qquad ab = c, \qquad bc = a \qquad \text{and} \qquad ac = b$$

*(i.) Show that $V_4$ is abelian.*

*(ii.) Show that $V_4$ is isomorphic to the direct product of cyclic groups $\mathbb{Z}_2 \times \mathbb{Z}_2$. To do this choose a suitable basis of $\mathbb{Z}_2 \times \mathbb{Z}_2$ and group composition rule and use it to show that the basis elements of $\mathbb{Z}_2 \times \mathbb{Z}_2$ have the same relations as those of $V_4$.*

### 4.3.1 The First Isomomorphism Theorem

The first isomomorphism theorem combines many of the observations we have made in the preceeding section.

**Theorem 4.3.2.** *(The First Isomorphism Theorem) Let $G$ and $G'$ be groups and let $f : G \rightarrow G'$ be a group homomorphism. Then the image of $f$ is isomorphic to the coset group $\frac{G}{Ker(f)}$. If $f$ is a surjective map then $G' \cong \frac{G}{Ker(f)}$.*

*Proof.* Let $K$ denote the kernel of $f$ and $H$ denote the image of $f$. Define a map $\phi : \frac{G}{K} \rightarrow H$ by

$$\phi(g \circ K) = f(g) \tag{4.39}$$

where $g \in G$. Let us check that $\phi$ is well-defined in that it maps different elements in a coset $gK$ to the same image $f(g)$. Suppose that $g_1 K = g_2 K$ then $g_1^{-1} \cdot g_2 \in K$ and

$$\begin{aligned} \phi(g_1 \circ K) &= f(g_1) \\ &= f(g_1) \circ' e' \\ &= f(g_1) \circ' f(g_1^{-1} \circ g_2) \\ &= f(g_1 \circ g_1^{-1} \circ g_2) \\ &= f(g_2) \\ &= \phi(g_2 \circ K). \end{aligned} \tag{4.40}$$

[3]The conjugate action is also called the group adjoint action

$\phi$ is a group homomorphism as

$$\begin{aligned}\phi(g_1 \circ K) \circ' \phi(g_2 \circ K) &= f(g_1) \circ' f(g_2) \\ &= f(g_1 \circ g_2) \\ &= \phi((g_1 \circ g_2) \circ K) \\ &= \phi((g_1 \circ K) \circ (g_2 \circ K))\end{aligned} \tag{4.41}$$

as $K \lhd G$. To prove that $\phi$ is an isomorphism we must show it is surjective (onto) and injective (one-to-one). For any $h \in H$ we have by the definition of $H$ that there exists $g \in G$ such that $f(g) = h$, hence $h = f(g) = \phi(g \circ K)$ and $\phi$ is surjective. To show that $\phi$ is injective let us assume the contrary statement that two distinct cosets $(g_1 \circ K \neq g_2 \circ K)$ are mapped to the same element $f(g_1) = f(g_2)$. As $f$ is a homorphism $f(g_1^{-1} \circ g_2) = e'$, hence $g_1^{-1} \circ g_2 \in K$ and so $g_1 \circ K = g_1 \circ (g_1^{-1} \circ g_2 \circ K) = g_2 \circ K$ contradicting our assumption that $g_1 \circ K \neq g_2 \circ K$. Hence $\phi$ is injective. As $\phi$ is both surjective and injective it is a bijection. The inverse map $\phi^{-1}(f(g)) = g \circ K$ is also a homomorphism:

$$\begin{aligned}\phi^{-1}(f(g_1) \circ' f(g_2)) &= \phi^{-1}(f(g_1 \circ g_2)) \\ &= (g_1 \circ g_2) \circ K \\ &= (g_1 \circ K) \circ (g_2 \circ K) \\ &= \phi^{-1}(g_1 \circ K) \circ' \phi^{-1}(g_2 \circ K))\end{aligned} \tag{4.42}$$

as well as a bijection. Hence $\phi$ is a group isomorphism and $\frac{G}{Ker(f)} \cong H$. If $f$ is surjective onto $G'$ then $H = G'$ and $\frac{G}{Ker(f)} \cong G'$. □

## 4.4 Some Representation Theory

**Definition** A *representation* of a group on a vector space $V$ is a group homomorphism $\Pi : G \to GL(V)$.

In other words a representation $\Pi$ is a way to write the group $G$ as matrices acting on a vector space which preserves the group composition law. Many groups are naturally written as matrices e.g. $GL(n, \mathbb{F})$, $SL(n, \mathbb{F})$, $SO(n)$, $O(n)$, $U(n)$, $SU(n)$ etc. (where $\mathbb{F}$ stands for $\mathbb{Z}$, $\mathbb{R}$, $\mathbb{Q}$, $\mathbb{C}$ ...) however there may be numerous ways to write the group elements as matrices. In addition not all groups can be represented as matrices e.g. $S_\infty$ (the infinite symmetric group) - try writing out an $\infty \times \infty$ matrix! Similarly $GL(\infty, \mathbb{F})$, $SL(\infty, \mathbb{F})$, ... for that matter. Here $V$ is called the representation space and the dimension of the representation is the dimension of the vector space $V$, i.e. $Dim(V)$.

**Definition** If a representation $\Pi$ is such that $Ker(\Pi) = e$ where $e$ is the identity element of $G$ then $\Pi$ is a *faithful* representation.

That $Ker\Pi$ is trivial indicates that $\Pi$ is injective (one-to-one), as suppose $\Pi$ was not injective so that $\Pi(g_1) = \Pi(g_2)$ where $g_1 \neq g_2$ for $g_1, g_2 \in G$ then as $\Pi$ is a homomorphism

$$\Pi(g_2^{-1} \circ g_1) = \mathbb{I} \tag{4.43}$$

where $\mathbb{I}$ is the identity matrix acting on $V$. Hence $g_2^{-1} \circ g_1 \in Ker(\Pi)$ and the kernel would be non-trivial.

**Definition** A representation $\Pi_1(G) \in GL(V_1)$ is *equivalent* to a second representation $\Pi_2(G) \in GL(V_2)$ if there exists an invertible linear map $T : V_1 \to V_2$ such that

$$T\Pi_1(g) = \Pi_2(g)T \qquad \forall \quad g \in G \tag{4.44}$$

The map $T$ is called the *intertwiner* of the representations $Pi_1$ and $Pi_2$.

**Definition** $W \subset V$ is an *invariant subspace* of a representation $\Pi : G \to GL(V)$ if $\Pi(g)W \subset W$ for all $g \in G$.

$W$ is called a subrepresentation space and if such an invariant subspace exists evidently one can trivially construct a representation of $G$ whose dimension is smaller than that of $\Pi$ (as $Dim(W) < Dim(V)$) by restricting the action of $\Pi$ to its action on $W$. The representations which possess no invariant subspaces are special.

**Definition** An *irreducible representation* $\Pi : G \to GL(V)$ contains no non-trivial invariant sub-spaces in $V$.

That is there do not exist any subspaces $W \subset V$ such that $\Pi(g)W \subset W \quad \forall \quad g \in G$ except $W = V$ or $W = \{e\}$. The irreducible represesntations are often referred to by the shorthand "irrep" and they are the basic building blocks of all the other "reducible" representations of $G$. They are the prime numbers of representation theory.

### 4.4.1 Schur's Lemma

**Theorem 4.4.1.** *(Schur's lemma first form) Let $\Pi_1 : G \to GL(V)$ and $\Pi_2 : G \to GL(W)$ be irreducible representations of $G$ and let $T : V \to W$ be an intertwining map between $\Pi_1$ and $\Pi_2$. Then either $T = 0$ (the zero map) or $T$ is an isomorphism.*

*Proof.* $T$ is an intertwining map so $T\Pi_1(g) = \Pi_2(g)T$ for all $g \in G$. First we show that $Ker(T)$ is an invariant subspace of $V$ as if $v \in Ker(T)$ then $Tv = \mathbf{0}$ (as the identity element on the vector space is the zero vector under vector addition), therefore

$$T\Pi_1(g)v = \Pi_2(g)T(v) = \mathbf{0} \quad \Rightarrow \Pi_1(g)v \in Ker(T) \quad \forall \quad v \in Ker(T). \tag{4.45}$$

Hence $Ker(T)$ is an invariant subspace of $V$ under the action of $\Pi_1(G)$. As $\Pi_1(G)$ is an ireducible representation of $G$ then $Ker(T) = \{\mathbf{0}\}$ or $V$. If $Ker(T) = V$ then $T$ is a map sending all $v \in V$ to $\mathbf{0} \in W$ (the zero map) and $T = 0$. If $Ker(T) = \mathbf{0} \in V$ then $T$ is an injective map. If $T$ is injective and in addition surjective then it is an isomorphism, so it remains for us to show that if $T$ is not the zero map it is a surjective map. We will do this by proving that the image of $T$ is an invariant subspace of $W$. Let the image of a vector $v \in V$ be denoted $w \in W$, i.e. $T(v) = w$ then

$$\Pi_2(g)w = \Pi_2(g)T(v) = T(\Pi_1(g)v) \in Im(T) \quad \forall \quad g \in G \tag{4.46}$$

and so the image of $T$ is an invariant subspace of $W$. As $\Pi_2$ is an irreducible representation then it has no non-trivial invariant subspaces, hence $Im(T) = \{\mathbf{0}\}$ or $W$. If the

image of $T$ is the zero vector then $T$ is the zero map, otherwise if the image of $T$ is $W$ then $T$ is a surjective map. Consequently either $T = 0$ or $T$ is an isomorphism between $V$ and $W$. □

**Theorem 4.4.2.** *(Schur's lemma second form) If $T : V \to V$ is an intertwiner from an irreducible representation $\Pi$ to itself and $V$ is a finite-dimensional complex vector space then $T = \lambda\mathbb{I}$ for some $\lambda \in \mathbb{C}$.*

*Proof.* We have $T\Pi(g) = \Pi(g)T$ and as $V$ is a complex vector space then one can always solve the equation $det(T - \lambda\mathbb{I}) = 0$ to find a complex eigenvalue $\lambda$[4]. Hence $Tv = \lambda v$ where $v$ is an eigenvector of $T$ and

$$T\Pi(g)v = \Pi(g)Tv = \lambda\Pi(g)v \quad \forall \quad g \in G \tag{4.47}$$

So $\Pi(g)v$ is another eigenvector for $T$ with eigenvalue $\lambda$. Hence the $\lambda$-eigenspace of $T$ is an invariant subspace of $\Pi(G)$. As $\Pi$ is an irreducible representation then the $\lambda$-eigenspace of $T$ is either $\{\mathbf{0}\}$ or $V$ itself. If we assume $V$ to be non-trivial then at least one eigenvalue exists and so the $\lambda$-eigenspace of $T$ is $V$ itself. Therefore

$$Tv = \lambda v \quad \forall \quad v \in V \qquad \Rightarrow \quad T = \lambda\mathbb{I}. \tag{4.48}$$

□

A corollary of Schur's lemma is that if there exist a pair of intertwining maps $T_1 : V \to W$ and $T_2 : V \to W$ which are both non-zero then $T_1 = \lambda T_2$ for some $\lambda \in \mathbb{C}$. For if $T_2$ is non-zero then it is an isomorphism of $V$ and $W$ and its inverse map $T_2^{-1} : W \to V$ is also an interwtwiner. Now

$$T_1T_2^{-1}\Pi_2(g) = T_1\Pi_1(g)T_2^{-1} = \Pi_2(g)T_1T_2^{-1} \tag{4.49}$$

hence $T_1T_2^{-1} : W \to W$ and by Schur's lemma (second form) we have $T_1T_2^{-1} = \lambda\mathbb{I}$ and so $T_1 = \lambda T_2$ for some $\lambda \in \mathbb{C}$.

**Problem 4.4.1.** *If $\Pi(G)$ is a finite-dimensional representation of a group $G$, show that the matrices $\Pi^*(g)$ also form a representation, where $\Pi^*(g)$ is the complex-conjugate of $\Pi(g)$.*

**Problem 4.4.2.** *The representation $\Pi^*(g)$ may or may not be equivalent to $\Pi(g)$. If they are equivalent then there exists an intertwining map, $T$, such that:*

$$\Pi^*(g) = T^{-1}\Pi(g)T$$

*Show that if $\Pi(g)$ is irreducible then $TT^* = \lambda\mathbb{I}$*

**Problem 4.4.3.** *If $\Pi(g)$ is a unitary representation on $\mathbb{C}^n$ show that $TT^\dagger = \mu\mathbb{I}$. (Hint: Make use of the fact that the inner product on $\mathbb{C}^n$ is $< v, w >= v^\dagger w$ where $v$, $w \in \mathbb{C}^n$ to find a relation between $\Pi^\dagger$ and $\Pi$.) Show that $T$ may be redefined so that $\mu = 1$ and that $T$ is either symmetric or antisymmetric.*

[4]This gives a polynomial in $\lambda$ which always has a solution over $\mathbb{C}$, or indeed over any algebraically closed field.

**Problem 4.4.4.** *Let $G$ be an abelian group. Show that*

$$\Pi(g_2) = \Pi(g_1)^{-1}\Pi(g_2)\Pi(g_1)$$

*where $g_1$, $g_2 \in G$ and $\Pi$ is an irreducible representation of $G$. Hence show that every complex irreducible representation of an abelian group is one-dimensional by proving that $\Pi(g) = \lambda\mathbb{I}$ for all $g \in G$ where $\lambda \in \mathbb{C}$.*

**Problem 4.4.5.** *Prove that a representation of $G$ of dimension $n+m$ having the form:*

$$\Pi(g) = \begin{pmatrix} A(g) & C(g) \\ 0 & B(g) \end{pmatrix} \qquad \forall\ g \in G$$

*is reducible. Here $A(g)$ is an $n \times n$ matrix, $B(g)$ is an $m \times m$ matrix, $C(g)$ is an $n \times m$ matrix and $0$ is an empty $m \times n$ matrix where $n$ and $m$ are integers and $n > 0$.*

**Problem 4.4.6.** *The affine group consists of affine transformations $(A, b)$ which act on a $D$-dimensional vector $x$ as:*

$$(A, b)x = Ax + b$$

*Find, with justification, a $(D + 1)$-dimensional reducible representation of the affine group of transformations.*

**Definition** Let $V$ be a vector space endowed with an inner product $<,>$. A representation $\Pi : G \to GL(V)$ is called *unitary* if $\Pi(g)$ are unitary operators i.e.

$$< \Pi(g)v, \Pi(g)w >=< v, w > \qquad \forall g \in G, \quad v, w \in V. \tag{4.50}$$

**Definition** Let $\Pi : G \to GL(V)$ be a representation on a finite-dimensional vector space $V$, then the *character* of $\Pi$ is the function $\chi_\Pi : G \to \mathbb{C}$ defined by

$$\chi_\Pi(g) = Tr(\Pi(g)) \tag{4.51}$$

where $Tr$ is the trace.

Notice that $\chi_\Pi(e) = Tr(\Pi(e)) = Tr(\mathbb{I}) = Dim(V)$ is the dimension of the representation. The character is constant on the conjugacy classes of a group $G$ as

$$\begin{aligned} \chi_\Pi(g \circ h \circ g^{-1}) &= Tr(\Pi(g \circ h \circ g^{-1})) \\ &= Tr(\Pi(g)\Pi(h)\Pi(g^{-1})) \\ &= Tr(\Pi(h)) \\ &= \chi_\Pi(h). \end{aligned} \tag{4.52}$$

where we have used the cyclicty of the trace. Any function which is invariant over the conjugacy class is called a 'class function'. If $\Pi$ is a unitary representation then

$$\chi_\Pi(g^{-1}) = Tr(\Pi(g^{-1})) = Tr(\Pi(g)^{-1}) = Tr(\Pi(g)^\dagger) = \chi_{\Pi^\dagger}(g) = \overline{\chi_\Pi(g)}. \tag{4.53}$$

If $\Pi_1$ and $\Pi_2$ are equivalent representations (with intertwinging map $T$) then they have the same characters as

$$\begin{aligned}\chi_{\Pi_1}(g) &= Tr(\Pi_1(g)) \\ &= Tr(T^{-1}\Pi_2(g)T) \\ &= Tr(\Pi_2(g)) \\ &= \chi_{\Pi_2}(g)\end{aligned} \tag{4.54}$$

and conversely if two representations of $G$ have the same characters for all $g \in G$ then they are equivalent representations.

### 4.4.2 The Direct Sum and Tensor Product

Given two representations $\Pi_1 : G \to GL(V_1)$ and $\Pi_2 : G \to GL(V_2)$ of a group $G$ one can form two important representations:

1. **The direct sum**, $\Pi_1 \oplus \Pi_2 : G \to GL(V_1 \oplus V_2)$ such that $(\Pi_1 \oplus \Pi_2)(g) = \Pi_1(g) \oplus \Pi_2(g)$. This is a homomorphism as

$$\begin{aligned}(\Pi_1 \oplus \Pi_2)(g_1 \circ g_2) &= \begin{pmatrix} \Pi_1(g_1 \circ g_2) & 0 \\ 0 & \Pi_2(g_1 \circ g_2) \end{pmatrix} \\ &= \begin{pmatrix} \Pi_1(g_1)\Pi_1(g_2) & 0 \\ 0 & \Pi_2(g_1)\Pi_2(g_2) \end{pmatrix} \\ &= \begin{pmatrix} \Pi_1(g_1) & 0 \\ 0 & \Pi_2(g_1) \end{pmatrix}\begin{pmatrix} \Pi_1(g_2) & 0 \\ 0 & \Pi_2(g_2) \end{pmatrix} \\ &= (\Pi_1 \oplus \Pi_2)(g_1)(\Pi_1 \oplus \Pi_2)(g_2)\end{aligned} \tag{4.55}$$

   If $V_1$ is the vector space with basis $\{\mathbf{e}_1, \mathbf{e}_2, \ldots \mathbf{e}_n\}$ and $V_2$ is the vector space with basis $\{\mathbf{f}_1, \mathbf{f}_2, \ldots \mathbf{f}_m\}$ then $V_1 \oplus V_2$ has the basis $\{\mathbf{e}_1, \mathbf{e}_2, \ldots \mathbf{e}_n, \mathbf{f}_1, \mathbf{f}_2, \ldots \mathbf{f}_m\}$, i.e. we can write this using the direct product as $V_1 \oplus V_2 \equiv \{(v_1, v_2) \in V_1 \times V_2 | v_1 \in V_1, v_2 \in V_2\}$ with vector addition and scalar mulitplication acting as

$$\begin{aligned}(v_1, v_2) + (v_1', v_2') &= (v_1 + v_1', v_2 + v_2') \\ a(v_1, v_2) &= (av_1, av_2)\end{aligned} \tag{4.56}$$

   where $v_1, v_1' \in V_1$, $v_2, v_2' \in V_2$ and $a$ is a constant. In this notation the basis of $V_1 \oplus V_2$ is

$$\{(\mathbf{e}_1, \mathbf{0}), (\mathbf{e}_2, \mathbf{0}), \ldots (\mathbf{e}_n, \mathbf{0}), (\mathbf{0}, \mathbf{f}_1), (\mathbf{0}, \mathbf{f}_2), \ldots (\mathbf{0}, \mathbf{f}_m)\} \cong \{\mathbf{e}_1, \mathbf{e}_2, \ldots \mathbf{e}_n, \mathbf{f}_1, \mathbf{f}_2, \ldots \mathbf{f}_m\}.$$

   Hence $Dim(V_1 \oplus V_2) = Dim(V_1) + Dim(V_2) = n + m$.

   **Example** Let $G$ be $\mathbb{Z}_2 \equiv \{e, g | e = Id, g^2 = e\}$ with $V_1 = \mathbb{R}^1$ and $V_2 = \mathbb{R}^2$ so that

$$\begin{aligned}&\Pi_1(e) = 1, && \Pi_1(g) = -1 \\ &\Pi_2(e) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, && \Pi_2(g) = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}\end{aligned} \tag{4.57}$$

now $V_1 \oplus V_3 = \mathbb{R}^3$ with

$$(\Pi_1 \oplus \Pi_2)(e) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \qquad \Pi_2(g) = \begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix}. \tag{4.58}$$

2. **The tensor product**, $\Pi_1 \otimes \Pi_2 : G \to GL(V_1 \otimes V_2)$ such that $(\Pi_1 \otimes \Pi_2)(g) = \Pi_1(g) \otimes \Pi_2(g)$. The tensor product is the most general blinear product and so its defintion may seem obscure at first sight. This is a homomorphism as

$$\begin{aligned}(\Pi_1 \otimes \Pi_2)(g_1 \circ g_2) &= \Pi_1(g_1 \circ g_2) \otimes \Pi_2(g_1 \circ g_2) \\ &= \Pi_1(g_1)\Pi_1(g_2) \otimes \Pi_2(g_1)\Pi_2(g_2) \\ &= (\Pi_1 \otimes \Pi_2)(g_1)(\Pi_1(g_2) \otimes \Pi_2(g_2)) \\ &= (\Pi_1 \otimes \Pi_2)(g_1)(\Pi_1 \otimes \Pi_2)(g_2)\end{aligned} \tag{4.59}$$

If $V_1$ is the vector space with basis $\{\mathbf{e}_1, \mathbf{e}_2, \ldots \mathbf{e}_n\}$ and $V_2$ is the vector space with basis $\{\mathbf{f}_1, \mathbf{f}_2, \ldots \mathbf{f}_m\}$ then $V_1 \otimes V_2$ has the basis

$$\{\mathbf{e}_1 \otimes \mathbf{f}_1, \mathbf{e}_1 \otimes \mathbf{f}_2, \ldots \mathbf{e}_1 \otimes \mathbf{f}_m, \mathbf{e}_2 \otimes \mathbf{f}_1, \mathbf{e}_2 \otimes \mathbf{f}_2, \ldots \mathbf{e}_2 \otimes \mathbf{f}_m, \ldots, \mathbf{e}_n \otimes \mathbf{f}_1, \mathbf{e}_n \otimes \mathbf{f}_2, \ldots \mathbf{e}_n \otimes \mathbf{f}_m\}$$

i.e. the basis is $\{\mathbf{e}_i \otimes \mathbf{e}_j | i = 1, 2, \ldots Dim(V_1), j = 1, 2, \ldots Dim(V_2)\}$. Hence $Dim(V_1 \otimes V_2) = Dim(V_1) \times Dim(V_2) = nm$. The tensor product of two vector spaces $V$ and $W$ satisfies

$$\begin{aligned}(v_1 + v_2) \otimes w_1 &= v_1 \otimes w_1 + v_2 \otimes w_1 \\ v_1 \otimes (w_1 + w_2) &= v_1 \otimes w_1 + v_1 \otimes w_2 \\ av \otimes w &= v \otimes aw = a(v \otimes w)\end{aligned} \tag{4.60}$$

where $v, v_1, v_2 \in V$, $w, w_1, w_2 \in W$ and $a$ is a constant.

**Example** As for the direct sum consider the example where $G$ is $\mathbb{Z}_2$ and $\Pi_1$ and $\Pi_2$ are the representations given explicitly in equation (4.57) above. Then the basis elements for $V_1 \otimes V_2$ are $\{\mathbf{e}_1 \otimes \mathbf{f}_1, \mathbf{e}_1 \otimes \mathbf{f}_2\}$ where $\mathbf{e}_1$ is the basis vector for $\mathbb{R}$ and $\{\mathbf{f}_1, \mathbf{f}_2\}$ are the basis vectors for $\mathbb{R}^2$ and the tensor product representation is

$$(\Pi_1 \otimes \Pi_2)(e) = 1 \otimes \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \qquad (\Pi_1 \otimes \Pi_2)(g) = -1 \otimes \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}.$$

These act on $\mathbb{R} \otimes \mathbb{R}^2$ by

$$\begin{aligned}(\Pi_1 \otimes \Pi_2)(e)(v_1 \otimes v_2) &= v_1 \otimes v_2, \\ (\Pi_1 \otimes \Pi_2)(g)(v_1 \otimes v_2) &= -v_1 \otimes \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix} v_2 = v_1 \otimes v_2\end{aligned} \tag{4.61}$$

which is the trivial representation acting on the two-dimensional vector space $\mathbb{R} \otimes \mathbb{R}^2 \cong \mathbb{R}^2$. A slightly less trivial example involves the representation $\Pi_3$ of $\mathbb{Z}_2$ on $\mathbb{R}^2$ given by

$$\Pi_3(e) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \qquad \Pi_3(g) = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}. \tag{4.62}$$

The tensor product representation $\Pi_1 \otimes \Pi_3$ acts on $\mathbb{R}^2$ as

$$(\Pi_1 \otimes \Pi_3)(e) = 1 \otimes \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \qquad (\Pi_1 \otimes \Pi_3)(g) = -1 \otimes \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}$$

these act on $\mathbb{R} \otimes \mathbb{R}^2$ by

$$\begin{aligned} (\Pi_1 \otimes \Pi_3)(e)(v_1 \otimes v_2) &= v_1 \otimes v_2, \\ (\Pi_1 \otimes \Pi_3)(g)(v_1 \otimes v_2) &= -v_1 \otimes \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} v_2 = v_1 \otimes \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} v_2 \end{aligned} \tag{4.63}$$

which is non-trivial.

One may introduce scalar products on the direct sum and tensor product spaces:

$$\begin{aligned} &< v_1 \oplus w_1, v_2 \oplus w_2 >_{V \oplus W} \equiv < v_1, v_2 >_V + < w_1, w_2 >_W \\ &< v_1 \otimes w_1, v_2 \otimes w_2 >_{V \otimes W} \equiv < v_1, v_2 >_V < w_1, w_2 >_W \end{aligned} \tag{4.64}$$

as well as the character function:

$$\begin{aligned} \chi_{\Pi_1 \oplus \Pi_2}(g) &= Tr(\Pi_1(g)) + Tr(\Pi_2(g)) \\ \chi_{\Pi_1 \otimes \Pi_2}(g) &= Tr_V(\Pi_1(g)) Tr_W(\Pi_2(g)). \end{aligned} \tag{4.65}$$

One might think that all the information about these product representations is contained already in $V$ and $W$. However consider the endomorphisms (the homomorphisms from a vector space to itself[5]) of $V \oplus W$, denoted $End(V \oplus W)$. Any $A \in End(V \oplus W)$ may be written

$$A = \begin{pmatrix} A_{VV} & A_{VW} \\ A_{WV} & A_{WW} \end{pmatrix} \tag{4.66}$$

where $A_{VV} : V \to V$, $A_{VW} : V \to W$ etc. that is $A_{VV} \in End(V)$ and $A_{WW} \in EndW$ do not generate all the endomorphisms of $V \oplus W$ (note that if $Dim(V) = n$ and $Dim(W) = m$ then $Dim(End(V \oplus W)) = (n+m)^2 \geq n^2 + m^2 = Dim(End(V)) + Dim(End(W))$. On the other hand the endomorphisms of $V$ and $W$ do generate all the endomorphisms of the tensor product space $V \otimes W$ as $Dim(End(V \otimes W)) = n^2 m^2 = Dim(End(V)) Dim(End(W))$.

The direct sum never gives an irreducible representation, having two non-trivial subspaces $V \oplus 0 \cong V$ and $0 \oplus W \cong W$. It is less straightforward with the tensor product to discover whether or not it gives an irreducible representation. Frequently one is interested in decomposing the tensor product into direct sums of irreducible subrepresentations:

$$V \otimes W = U_1 \oplus U_2 \oplus \ldots \oplus U_n. \tag{4.67}$$

To do this one must find an endomorphism (a change of basis) of $V \otimes W$ such that

$$T(\Pi_1 \otimes \Pi_2(g))T^{-1} = \hat{\Pi}_1(g) \oplus \hat{\Pi}_2(g) \oplus \ldots \hat{\Pi}_n(g) \tag{4.68}$$

where $T \in End(V \otimes W)$. The decomposition

$$\Pi(G) \otimes \Pi(G) = \sum_i a_i \hat{\Pi}_i(G) \tag{4.69}$$

[5]If an endomorphism is invertible then the map is an automorphism.

is called the Clebsch-Gordan decomposition. This is not always possible. One can achieve this decomposition for one example central to quantum mechanics $G = SU(2)$. It is a fact (which we will not prove here) that $SU(2)$ has only one unitary irreducible representation for each vector space of dimension $Dim(V) \equiv n+1$. This $n+1$-dimensional representation is isomorphic to a representation of the irreducible representations of $SO(3)$ associated to angular momentum in quantum mechanics due to the group isomorphism $\frac{SU(2)}{\mathbb{Z})2} \cong SO(3)$ which will be shown explicitly later in this chapter. In summary representations of $SU(2)$ may be labelled by $Dim(V) = n+1$ and the equivalent $SO(3)$ representation is labelled by spin $j$. In fact $j = \frac{n}{2}$ hence as $n \in \mathbb{Z}^+$ then $j$ may take half-integer (fermions) as well as integer (bosons) values. When $j = 0$ then $n = 0$ so $Dim(V) = 1$ is the trivial representation of $SU(2)$; $j = \frac{1}{2}$ then $n = 1$ and $Dim(V) = 2$ giving the "fundamental" or standard representation of $SU(2)$ as a two-by-two matrix; and when $j = 1$ then $n = 2$ giving $Dim(V) = 3$ is called the "adjoint" representation of $SU(2)$. The Clebsch-Gordan decomposition rewrites the tensor product of two $SU(2)$ irreducible representations $[j_1]$ and $[j_2]$, labelled using the spin, as a direct sum of irreducible representations:

$$[j_1] \otimes [j_2] = [j_1 + j_2] \oplus [j_1 + j_2 - 1] \oplus \ldots \oplus [|j_1 - j_2|]. \tag{4.70}$$

Some simple examples are

$$[0] \otimes [j] = [j] \tag{4.71}$$

One can quickly check that the tensor product has the same dimension as the direct sum. Note that $Dim[j] = Dim(V) = n + 1 = 2j + 1$ so that $Dim([0] \otimes [j]) = 1 \times (2j + 1) = Dim[j]$. Another example short example is

$$[\frac{1}{2}] \otimes [j] = [\frac{1}{2} + j] \oplus [-\frac{1}{2} + j] \tag{4.72}$$

where we have $Dim([\frac{1}{2}] \otimes [j]) = (2\frac{1}{2} + 1)(2j + 1) = 4j + 2$ while the direct sum of representations has $Dim([\frac{1}{2}+j]\oplus[-\frac{1}{2}+j]) = (2(\frac{1}{2}+j)+1)+(2(-\frac{1}{2}+j)+1) = 4j+2$. Notice that the tensor products of the "fundamental" representation $[\frac{1}{2}]$ with itself generates all the other irreducible representations of $SU(2)$ that is

$$\begin{aligned} [\frac{1}{2}] \otimes [\frac{1}{2}] &= [1] \oplus [0] \\ \text{Dimensions:} \quad 2 \ &\times \ 2 \ = \ 3 \ + \ 1 \\ [1] \otimes [\frac{1}{2}] &= [\frac{3}{2}] \oplus [\frac{1}{2}] \\ \text{Dimensions:} \quad 3 \ &\times \ 2 \ = \ 4 \ + \ 2. \end{aligned} \tag{4.73}$$

For other groups the decomposition theory is more involved. To work out the Clebsch-Gordan coefficients one must know the inequivalent irreducible representations ofthe group, its conjugacy classes and its character table. If a representation of a group itself may be rewritten as a sum of representations it is by definition not an irreducible representation - it is called a reducible representation.

**Definition** A representation $\Pi : G \to GL(V_n \oplus V_m)$ on a vector space of dimension $n + m$ is *reducible* if $\Pi(g)$ has the form

$$\Pi(g) = \begin{pmatrix} A(g) & C(g) \\ 0 & B(g) \end{pmatrix} \qquad \forall \quad g \in G \tag{4.74}$$

where $A$ is an $n \times n$ matrix, $B$ is an $m \times m$ matrix, $C$ is an $n \times m$ matrix and 0 is the empty $m \times n$ matrix.

Notice that

$$\begin{pmatrix} A(g) & C(g) \\ 0 & B(g) \end{pmatrix} \begin{pmatrix} v_n \\ 0_m \end{pmatrix} = \begin{pmatrix} A(g)v_n \\ 0_m \end{pmatrix} \tag{4.75}$$

where $0_m \in V_m$ is the $m$-dimensional zero vector and $v_n \in V_n$ is an $n$-dimensional vector. So we see that $V_n$ is an invariant subspace of $\Pi$ and so $\Pi$ is reducible. Furthermore if we multiply two such matrices together we have

$$\begin{aligned} \Pi(g_1)\Pi(g_2) &= \begin{pmatrix} A(g_1) & C(g_1) \\ 0 & B(g_1) \end{pmatrix} \begin{pmatrix} A(g_2) & C(g_2) \\ 0 & B(g_2) \end{pmatrix} \\ &= \begin{pmatrix} A(g_1)A(g_2) & A(g_1)C(g_2) + C(g_1)B(g_2) \\ 0 & B(g_1)B(g_2) \end{pmatrix} \\ &= \Pi(g_1 \circ g_2) \\ &= \begin{pmatrix} A(g_1 \circ g_2) & C(g_1 \circ g_2) \\ 0 & B(g_1 \circ g_2) \end{pmatrix} \end{aligned} \tag{4.76}$$

hence we see that $A(g_1 \circ g_2) = A(g_1)A(g_2)$ and $A(g)$ is representation of $G$ on the invariant subspace $V_n$. For finite groups the matrix $C$ is equivalent to the null matrix (by Maschke's theorem "all reducible representations of a finite group are completely reducible"). In this case the representation $\Pi$ is said to be completely reducible:

$$\Pi(g) = A(g) \oplus B(g). \tag{4.77}$$

It does not follow that $A(G)$ and $B(G)$ are themselves irreducible, but if they are not then the process may be repeated until $\Pi(G)$ is expressed as a direct sum of irreducible representations.

## 4.5 Lie Groups

Many of the groups we have met so far have been parameterised by discrete variables e.g. $\{e, g, g^2\}$ for $\mathbb{Z}_3$ but frequently a number of group actions we have met, e.g. $SO(n)$, $SU(n)$, $U(n)$, $Sp(n)$, have been described by continuous parameters. For example $SO(2)$ describing rotations of $S^1$ is parameterised by $\theta$ which takes values in the continuous set $[0, 2\pi)$ and for each value of $\theta$ we find an element of $SO(2)$:

$$R(\theta) = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix} \tag{4.78}$$

(one may check that $R(\theta)R^T(\theta) = \mathbb{I}$ and $Det(R(\theta)) = 1$). $R(\theta)$ is a two-dimensional representation of the abstract group $SO(2)$. We may check that is a faithful representation of $SO(2)$: $R(0) = \mathbb{I}$ and the kernel of the representation is trivial for $\theta \in [0, 2\pi)$. Incidentally the two-dimensional representation is irreducible over $\mathbb{R}$ but it is reducible over $\mathbb{C}$. Over $\mathbb{C}$ we take as column vector $\begin{pmatrix} z \\ z^* \end{pmatrix} = \begin{pmatrix} x+iy \\ x-iy \end{pmatrix} = \begin{pmatrix} re^{i\theta} \\ re^{-i\theta} \end{pmatrix}$ and an

$SO(2)$ rotation takes

$$\begin{pmatrix} z \\ z^* \end{pmatrix} \to \begin{pmatrix} z' \\ z'^* \end{pmatrix} = \begin{pmatrix} re^{i(\theta+\phi)} \\ re^{-i(\theta+\phi)} \end{pmatrix} \tag{4.79}$$

that is

$$R(\phi, \mathbb{C}) = \begin{pmatrix} e^{i\phi} & 0 \\ 0 & e^{-i\phi} \end{pmatrix} \tag{4.80}$$

There is a qualitative difference when we move from $\mathbb{R}$ to $\mathbb{C}$ as this matrix is block diagonal and hence reducible into two one-dimensional complex representations of $U(1) \cong SO(2)$. Geometrically the parameter defining the rotation parameterises the circle $S^1$. For other continuous groups we may also make an identification with a geometry e.g. $\mathbb{R}\backslash 0$ under multiplication is associated with two open half-lines (the real line with zero removed), a second example is $SU(2) = \{\begin{pmatrix} \alpha & -\beta^* \\ \beta & \alpha^* \end{pmatrix} ||\alpha|^2 + |\beta|^2 = 1\}$ which as a set parameterises $S^3$. The proper notion for the geometric setting is the manifold and each group discussed above is a manifold. Any geometric space one can imagine can be embedded in some Euclidean $\mathbb{R}^n$ as a surface of some dimensions less than or equal to $n$. For example the circle $S^1 \subset \mathbb{R}^2$ and in general $S^{n-1} \subset \mathbb{R}^n$. No matter how extraordinary the curvature of the surface (so long as it remains well-defined) a manifold will have the appearance of being a Euclidean space at a sufficiently local scale. Consider $S^1 \subset \mathbb{R}^2$ sufficiently close to a point on $S^1$, the segment of $S^1$ appears identical to $\mathbb{R}^1$. The geometry of a manifold is found by piecing together these open and locally-Euclidean stes. Each open neighbourhood is called a chart and is equipped with a map $\phi$ that converts points $p \in \mathcal{M}$, where $\mathcal{M}$ is the manifold, to local Euclidean coordinates. Using these local coordinates one can carry out all the usual mathematics in $\mathbb{R}^n$. The global structure of a manifold is defined by how these open sets are glued together. Since a manifold is a very well-defined structure these transition functions, encoding the gluing, are smooth. The study of manifolds is the beginning of learning about differential geometry.

**Definition** A *Lie group* is a differentiable manifold $G$ which is also a group such that the group product $G \times G \to G$ and the inverse map $g \to g^{-1}$ are differentiable.

We will restrict our interest to matrix Lie groups in this foundational course, these are those Lie groups which are written as matrices e.g. $SL(n, \mathbb{F}), SO(n), SU(n), Sp(n)$.

**Definition** A matrix Lie group $G$ is *connected* if given any two matrices $A$ and $B$ in $G$, there exists a continuous path $A(t)$ with $0 \leq t \leq 1$ such that $A(0) = A$ and $A(1) = B$.

A matrix Lie group which is not connected can be decomposed into several connected pieces.

**Theorem 4.5.1.** *If $G$ is a matrix Lie group then the component of $G$ connected to the identity is a subgroup of $G$. It is denoted $G_0$.*

*Proof.* Let $A(t), B(t) \in G_0$ such that $A(0) = \mathbb{I}$, $A(1) = A$, $B(0) = \mathbb{I}$ and $B(1) = B$ are continuous paths. Then $A(t)B(t)$ is a continuous path from $\mathbb{I}$ to $AB$. Hence $G_0$ is closed and evidently $\mathbb{I} \in G_0$. Also $A^{-1}(t) = A(-t)$ is a continuous path from $\mathbb{I}$ to $A^{-1} \in G_0$ defined by $A(-t)A(t) = \mathbb{I}$. □

The groups $GL(n,\mathbb{C})$, $SL(n\mathbb{C}$, $SL(n,\mathbb{R})$, $SO(n)$, $U(n)$ and $SU(n)$ are connected groups. While $GL(n,\mathbb{R}$ and $O(n)$ are not connected. For example one can convince oneself that $O(n)$ is not connected by supposing that $A, B \in O(n)$ such that $Det(A) = +1$ and $Det(B) = -1$. Then any path $A(t)$ such that $A(0) = A$ and $A(1) = B$ would give a continuous function $Det(A(t))$ passing from 1 to $-1$. Since $A \in O(n)$ satisfy $Det(A) = \pm 1$ then no such set of matrices forming a continuous path from $A$ to $B$ exist. A similar argument can be made for $GL(n,\mathbb{R})$ splitting it into components with $Det > 0$ and $Det < 0$.

### 4.5.1 Infinitesimal Generators and the Invariance of Physical Law

Newton's second law $\mathbf{F} = m\mathbf{a}$ may be read as a relation between the acceleration $\mathbf{a}$ and a restoring force $\mathbf{F}$. If $\mathbf{F}$ happens to take a particular form then the equation of motion for the system might be a wave equation:

$$\left(\frac{1}{c^2}\frac{\partial^2}{\partial t^2} - \nabla^2\right)u = 0 \tag{4.81}$$

where $\nabla^2 u$ measures the spatial curvature of the wavefunction $u$. Differential operators in physics are generally associated to generators of infinitesimal translations in space and time. Consider an infinitesimal time translation by $dt$ of a function $F(t)$:

$$\begin{aligned} f(t+dt) &= f(t) + \frac{df}{dt}dt + \mathcal{O}(dt^2) \\ &= (1 + dt\frac{d}{dt})f(t) + \mathcal{O}(dt^2). \end{aligned} \tag{4.82}$$

Repeating the timetranslation gives

$$\begin{aligned} f(t+2dt) &= (1 + dt\frac{d}{dt})f(t+dt) + \mathcal{O}(dt^2) \\ &= (1 + dt\frac{d}{dt})^2 f(t) + \mathcal{O}(dt^2). \end{aligned} \tag{4.83}$$

Repeating the infinitesimal time translation $n$ times and letting $n \to \infty$ to give a finite time translation $\tau$ we have

$$\begin{aligned} f(t+\tau) &= \lim_{n\to\infty}\left[(1 + dt\frac{d}{dt})^n f(t) + \mathcal{O}(dt^2)\right] \\ &= \lim_{n\to\infty}\left[(1 + \frac{\tau}{n}\frac{d}{dt})^n f(t) + \mathcal{O}((\frac{\tau}{n})^2)\right] \\ &= \exp(\tau\frac{d}{dt})f(t) \end{aligned} \tag{4.84}$$

Where we have used the binomial expansion

$$\begin{aligned} \left(1 + \frac{\tau}{n}\frac{d}{dt}\right)^n &= 1 + \binom{n}{1}\left(\frac{\tau}{n}\frac{d}{dt}\right) + \binom{n}{2}\left(\frac{\tau}{n}\frac{d}{dt}\right)^2 + \dots \\ &= 1 + \frac{n\tau}{n}\frac{d}{dt} + \frac{n(n-1)\tau^2}{2n^2}\left(\frac{d}{dt}\right)^2 + \dots \end{aligned} \tag{4.85}$$

so that

$$\lim_{n\to\infty}\left[\left(1 + \frac{\tau}{n}\frac{d}{dt}\right)^n\right] = 1 + \tau\frac{d}{dt} + \frac{\tau^2}{2!}\left(\frac{d}{dt}\right)^2 + \dots = \exp(\tau\frac{d}{dt}). \tag{4.86}$$

It is conventional for physicists to write this operator in an unusual way. We may write:

$$f(t+\tau) = \exp(\tau\frac{d}{dt})f(t) = \exp(-i\tau W)f(t) \tag{4.87}$$

where $W \equiv i\frac{d}{dt}$ is called the infinitesimal generator of time translations. By acting with $\exp(-i\tau W)$ on $f(t)$ we find $f(t+\tau)$. If we act twice, writing $T(\tau_i) \equiv \exp(-i\tau_i W))$, we find

$$\begin{aligned} T(\tau_2)T(\tau_1)f(t) &= \exp(-i\tau_2 W)\exp(-i\tau_1 W)f(t) \\ &= \exp(-i(\tau_2+\tau_1)W)f(t) \\ &= f(t+\tau_1+\tau_2) \\ &= T(\tau_1+\tau_2)f(t) \end{aligned} \tag{4.88}$$

hence the time translation generators satisfy the closure axiom of a group. One can confirm that the time translations do indeed form a group axiom by axiom.

Suppose that the function $f(t)$ has the explicit form $f(t) = \exp(-i\omega t)$, so that

$$W(f(t)) = i\frac{d}{dt}\exp(-i\omega t) = \omega\exp(-i\omega t) = \omega f(t) \tag{4.89}$$

and $\omega$ is an eigenvalue of $W$ with eigenfunction $f(t)$. Then,

$$f(t+\tau) = \exp(-i\tau W)f(t) = \exp(-i\tau\omega)f(t). \tag{4.90}$$

If, at $t=0$, $f(0)=1$ we have $f(\tau) = \exp(-i\tau\omega)$ as required. Now $f(t) = \exp(-it\omega)$ describes a system oscillating with frequency $\omega$ and $W$ may be referred to as the frequency operator. Additionally we note that

$$[W,t]f(t) = Wtf(t) - tWf(t) = if(t) + tWf(t) - tWf(t) = if(t) \tag{4.91}$$

i.e. $[W,t] = i$.

Similarly we might consider a translation in space:

$$\phi(\mathbf{r}+d\mathbf{r}) = \phi(\mathbf{r}) + d\mathbf{r}\cdot\nabla(\phi) + \mathcal{O}(d\mathbf{r}^2). \tag{4.92}$$

In the limit $n\to\infty$ consider a finite translation $\boldsymbol{\rho} \equiv nd\mathbf{r}$:

$$\phi(\mathbf{r}+\boldsymbol{\rho}) = \lim_{n\to\infty}\left[(1+\frac{\boldsymbol{\rho}}{n}\cdot\boldsymbol{\nabla})^n\right]\phi(\mathbf{r}) = \exp(\boldsymbol{\rho}\cdot\boldsymbol{\nabla})\phi(\mathbf{r}) \equiv T_{\mathbf{r}}(\boldsymbol{\rho})\phi(\mathbf{r}) \tag{4.93}$$

Where $T_{\mathbf{r}}(\boldsymbol{\rho})$ generates a translation of $\boldsymbol{\rho}$. As with the time translation the physicist's convention is to define:

$$T_{\mathbf{r}}(\boldsymbol{\rho}) \equiv \exp(i\boldsymbol{\rho}\cdot\mathbf{K}) \tag{4.94}$$

where $\mathbf{K} \equiv -i\boldsymbol{\nabla}$ is the generator of spatial translations. Its eigenfunction is $\exp(i\mathbf{k}\cdot\mathbf{r})$ with eigenvalue $\mathbf{k}$. Its components satisfy the non-trivial commutation relations:

$$\begin{aligned} [K_1,r_1]\phi(\mathbf{r}) &= [-i\frac{\partial}{\partial x},x]\phi(\mathbf{r}) = -i\phi(\mathbf{r}) + x(-i\frac{\partial\phi(\mathbf{r})}{\partial x}) - x(-i\frac{\partial\phi(\mathbf{r})}{\partial x}) = -i\phi(\mathbf{r}) \\ [K_2,r_2]\phi(\mathbf{r}) &= [-i\frac{\partial}{\partial y},y]\phi(\mathbf{r}) = -i\phi(\mathbf{r}) \\ [K_3,r_3]\phi(\mathbf{r}) &= [-i\frac{\partial}{\partial z},z]\phi(\mathbf{r}) = -i\phi(\mathbf{r}) \end{aligned} \tag{4.95}$$

or $[K_i,r_j] = -i\delta_{ij}$.

**Example** Calculate the electrostatic potential of a point charge at $\mathbf{r}_s \equiv \begin{pmatrix} 0 \\ 0 \\ d \end{pmatrix}$ which is denoted by $\phi(\mathbf{r}-\mathbf{r}_s)$ if $\phi(\mathbf{r}) = \frac{Q}{|\mathbf{r}|}$.
So,

$$\begin{aligned} \phi(\mathbf{r}-\mathbf{r}_s) &= \exp(-i\mathbf{r}_s\cdot\mathbf{K})\phi(\mathbf{r}) \\ &= \exp(-d\frac{\partial}{\partial z})\phi(\mathbf{r}) \\ &= (1 - d\frac{\partial}{\partial z} + \frac{d^2}{2!}\frac{\partial^2}{\partial z^2} - \ldots)\phi(\mathbf{r}) \end{aligned} \tag{4.96}$$

Now,

$$\begin{aligned} \frac{\partial}{\partial z}\phi(\mathbf{r}) &= -Q\frac{z}{r^3} = -Q\frac{\cos\theta}{r^2} \\ \frac{\partial^2}{\partial z^2}\phi(\mathbf{r}) &= \frac{\partial}{\partial z}(-Q\frac{z}{r^3}) = -\frac{Q}{r^3} + \frac{3Qz^2}{r^5} = -\frac{Q}{r^3} + \frac{3Q\cos^2\theta}{r^3} \end{aligned} \tag{4.97}$$

where we have defined $z$ in polar coordinates by $z = r\cos\theta$. Therefore

$$\phi(\mathbf{r}-\mathbf{r}_s) = \frac{Q}{r}\left(1 + \frac{d\cos\theta}{r} + \frac{d^2(3\cos^2\theta - 1)}{2!r^2} + \ldots\right). \tag{4.98}$$

These terms are referred to as the monopole, dipole and quadropole terms respectively and the expansion is called the multipole expansion.

Suppose that a system at position $\mathbf{r}$ and time $t$ is described by an equation of motion/state:

$$Df(\mathbf{r},t) = 0 \tag{4.99}$$

where $D$ is a differential operator. Under a time translation this becomes

$$\begin{aligned} 0 &= T_t(\tau)[Df(\mathbf{r},t)] \\ &= T_t(\tau)DT_t^{-1}(\tau)T_t(\tau)f(\mathbf{r},t) \\ &= (T_t(\tau)DT_t^{-1}(\tau))f(\mathbf{r},t+\tau). \end{aligned} \tag{4.100}$$

Consequently if the equation of motion/state is to hold at time $t+\tau$ then we see that $D$ must transform as

$$T_t(\tau)DT_t^{-1}(\tau) = D. \tag{4.101}$$

In other words $D$ must commute with $T_t$ the generator of time translations, or equivalently with the frequency operator $W$. Similarly if $Df(\mathbf{r},t) = 0$ is to be invariant under spatial translations then $D$ must commute with $\mathbf{K}$.

We might turn this line of argument on its head and attempt to generate equations of motion by finding the differential operators which such that $[D,\mathbf{K}] = 0$ and/or $[D,W] = 0$. For example if $D = \mathbf{K}$ trivially commutes with $\mathbf{K}$, however under a coordinate reflection $\mathbf{r} \to -\mathbf{r}$ the equation of motion would flip sign - it would not be parity invariant. Consequently we could try $D = \mathbf{K}^2$ giving a translation invariant equation of motion:

$$\mathbf{K}^2 f(\mathbf{r},t) = -\boldsymbol{\nabla}^2 f(\mathbf{r},t) = 0 \tag{4.102}$$

which is the Laplace equation. More generally we could pick $D = \mathbf{K}^2 - \mathbf{k}^2$, where $\mathbf{k}$ is a constant vector, giving

$$(\mathbf{K}^2 - \mathbf{k}^2)f(\mathbf{r},t) = 0 \quad \Rightarrow \quad -\boldsymbol{\nabla}^2 f(\mathbf{r},t) = \mathbf{k}^2 f(\mathbf{r},t) \tag{4.103}$$

which is the Helmholtz equation and has solution $f(\mathbf{r},t) = A\exp(\pm i\mathbf{k}\cdot\mathbf{r})$. Similarly a time invariant equation of motion which does not change sign under time reversal is

$$(W^2 - \omega^2)f(\mathbf{r},t) = 0 \quad \Rightarrow \quad -\frac{d^2 f(\mathbf{r},t)}{dt^2} = \omega^2 f(\mathbf{r},t) \tag{4.104}$$

which is solved by $f(\mathbf{r},t) = B\exp(\pm i\omega t)$. Putting these two invariances together we may find the wave equation:

$$(\mathbf{K}^2 - \frac{W^2}{c^2})u(\mathbf{r},t) = (\mathbf{k}^2 - \frac{\omega^2}{c^2})u(\mathbf{r},t) = 0 \tag{4.105}$$

and is satisfied is $|k| \equiv k = \pm\frac{\omega}{c}$, e.g. $u(\mathbf{r},t) = \exp(i\mathbf{k}\cdot\mathbf{r} - i\omega t)$ describes a wave with wavespeed $c$ and frequency $\omega$ in the $\mathbf{k}$ direction.

### 4.5.2 Lie Algebras

The generators of a Lie group are introduced by considering elements infnitesimally close to the identity. For $SO(2)$ with

$$R(\theta) = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix} \tag{4.106}$$

we can expand[6]

$$R(\theta) = \mathbb{I} - i\theta X + \mathcal{O}(\theta^2) \tag{4.107}$$

or equivalently

$$-iX = \left.\frac{dR(\theta)}{d\theta}\right|_{\theta=0} \tag{4.108}$$

that is the first term in the Taylor expansion of $R(\theta)$ about $R(0)$. Specifically we have

$$-iX = \left.\begin{pmatrix} -\sin(\theta) & -\cos(\theta) \\ \cos(\theta) & -\sin(\theta) \end{pmatrix}\right|_{\theta=0} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}. \tag{4.109}$$

The matrix $X$ is called the infinitesimal generator of $SO(2)$. Note that we could have derived that $X$ is skew-symmetric from the defining relation of $SO(2)$

$$\mathbb{I} = R(\theta)R^T(\theta) = (\mathbb{I} - i\theta X + \mathcal{O}(\theta^2))(\mathbb{I} - i\theta X^T + \mathcal{O}(\theta^2)) = \mathbb{I} - i\theta(X + X^T) + \mathcal{O}(\theta^2) \tag{4.110}$$

Hence $X = -X^T$. Also

$$1 = Det(R(\theta)) = Det(\mathbb{I} - i\theta X + \mathcal{O}(\theta^2)) = 1 - i\theta Tr(X) + \mathcal{O}(\theta^2) \tag{4.111}$$

and so $Tr(X) = 0$. One can learn the properties of the generators for other groups by imposing the definitions of the group on the exponential expansion.

The sub-group $G_0$ of all group elements continuously connected to $\mathbb{I}$ can be constructed from the exponentiation of the generators $X$ i.e. $e^{-i\theta X} \in G_0$. For connected groups one can reconstruct the full group from the generators or Lie algebra.

[6] We are using the physicist's conventions here where the generators $X$ are taken to be imaginary.

**Problem 4.5.1.** *$SO(2)$ is a connected group. Show that exponentiation of $-iX = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$ covers $SO(2)$. N.B. the matrix exponentiation is $e^{-i\theta X} = \sum_{n=0}^{\infty} \frac{1}{n!}(-i\theta X)^n$ where $X^n$ is the matrix multiplication of $X$ with itself $n$ times.*

**Definition** For each Lie group $G$ there exists a vector space called the *Lie algebra* of $G$, denoted $\mathfrak{g}$ or $Lie(G)$, such that $X \in \mathfrak{g}$ with $e^{-itX} = g \in G_0$ for all $t \in \mathbb{R}$.

We will revisit this definition of the Lie algebra after we have developed some of its properties.

**Examples of Lie algebras**

- $z \in U(1)$ then there exists $X \in \mathbb{R}$ such that $z = e^{-iX}$.
- $\mathbb{R}\backslash 0$ consists of two disconnected compoents. The positive half-line contains the multiplicative identity element so this is $G_0$. It can be written as $e^{-iX}$ where $-iX \in \mathbb{R}$.
- $SU(2)$ matrices may be written as $M = e^{i\boldsymbol{\alpha}\cdot\boldsymbol{\sigma}}$ where $\boldsymbol{\alpha} \in \mathbb{R}^3$ and $\boldsymbol{\sigma} = \begin{pmatrix} \sigma_1 \\ \sigma_2 \\ \sigma_3 \end{pmatrix}$ is a vector with matrix entries where

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{4.112}$$

  These are called the Pauli matrices.

  **Problem 4.5.2.** *Prove using the defining properties of $SU(2)$ that elements of its Lie algebra are traceless, Hermitian, two-by-two matrices.*

  **Problem 4.5.3.** *Show, by expanding the exponential map, that each element of $SU(2)$ can be written in the form*

$$g = \exp\left(i\hat{\mathbf{n}} \cdot \boldsymbol{\sigma}\theta\right)$$

  *where $\hat{\mathbf{n}}$ is a unit vector, $\boldsymbol{\sigma}$ is the vector whose components are the Pauli matrices $\sigma_i$ where $i = 1, 2, 3$ and $\theta$ is a continuous real parameter.*

- $SO(3)$ can be constructed from the rotations around the $x$, $y$ and $z$ axes of $\mathbb{R}^3$, namely,

$$R_x(\theta) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{pmatrix}, \quad R_y(\theta) = \begin{pmatrix} \cos\theta & 0 & \sin\theta \\ 0 & 1 & 0 \\ -\sin\theta & 0 & \cos\theta \end{pmatrix} \quad \text{and} \tag{4.113}$$

$$R_z(\theta) = \begin{pmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Note the different choice of sign convention adopted here for the rotation about the $y$-axis: this is chosen so that a rotation in positive $\theta$ rotates the posiitve half of the $z$-axis towards the positive half of the $x$-axis and that the rotation is governed by the right-hand-rule[7].

Hence the infinitesimal generators are found from $-iX \equiv \left.\frac{dR_x(\theta)}{d\theta}\right|_{\theta=0}$, $-iY \equiv \left.\frac{dR_y(\theta)}{d\theta}\right|_{\theta=0}$ and $-iZ \equiv \left.\frac{dR_z(\theta)}{d\theta}\right|_{\theta=0}$ and are given by

$$X = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{pmatrix}, \quad Y = \begin{pmatrix} 0 & 0 & i \\ 0 & 0 & 0 \\ -i & 0 & 0 \end{pmatrix} \quad \text{and} \quad Z = \begin{pmatrix} 0 & -i & 0 \\ i & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}. \tag{4.114}$$

**Theorem 4.5.2.** *Let $G$ be a matrix Lie group, $X$ be an element of $\mathfrak{g}$ and $A \in G$. Then $AXA^{-1} \in \mathfrak{g}$.*

*Proof.*

$$e^{it(AXA^{-1})} = \mathbb{I} + (it)(AXA^{-1}) + \frac{1}{2!}(it)^2(AXA^{-1})^2 + \ldots \tag{4.115}$$

but $(AXA^{-1})^n = (AX^nA^{-1})$ hence

$$e^{it(AXA^{-1})} = A(\mathbb{I} + (it)X + \frac{1}{2!}(it)^2X^2 + \ldots)A^{-1} = Ae^{itX}A^{-1} \in G \tag{4.116}$$

implying that $AXA^{-1} \in \mathfrak{g}$. □

The action $AXA^{-1}$ for $A \in G$, $X \in \mathfrak{g}$ defines a map on the Lie algebra acting as $G \times \mathfrak{g} \to \mathfrak{g}$. The group elements $A \in G$ act by conjugation on the Lie algebra to give an action moving through the algebra, this is called the adjoint action of the group. This is a representation of the Lie group on its own algebra and is known as the adjoint representation. Since $A \in G$ we may write $A = e^{itY}$ for some element $Y \in \mathfrak{g}$. We may use this to find how the adjoint group action descends to the Lie algebra directly. The adjoint action transforms $X \to e^{itY}Xe^{-itY} \in \mathfrak{g}$. This is a path $y(t) \equiv e^{itY}Xe^{-itY}$ through the algebra parameterised by $t$. The infinitesimal transformation is

$$\begin{aligned} -i\left.\frac{dy}{dt}\right|_{t=0} &= -i\Big(iYe^{itY}Xe^{-itY} + e^{itY}X(-iY)e^{-itY}\Big)\Big|_{t=0} \\ &= YX - XY \\ &= [Y, X] \in \mathfrak{g} \end{aligned} \tag{4.117}$$

We have rediscovered the Lie bracket.

**Definition** Given any two $n \times n$ matrices $A$ and $B$ the *Lie bracket* or commutator $[A, B]$ is defined to be

$$[A, B] = AB - BA. \tag{4.118}$$

[7]The right-hand rule can be used to give a consisitent definition of positive angle to a set of rotations of vector spaces. A rotation of positive angle about an axis is defined as rotating the space in the direction given by one's fingers when one wraps one's right hand around the axis with the thumb pointing along the positive direction of the axis. For example a positive rotation about the $z$-axis rotates the $+x$-axis towards the $+y$-axis, a positive rotation about the $x$-axis rotates the $+y$-axis towards the $+z$-axis and a positive rotation about the $y$-axis rotates the $+z$-axis towards the $+x$-axis.

This motivates a second definition of a Lie algebra.

**Definition** A *Lie algebra* $(V, [\bullet, \bullet])$ is a vector space $V$ together with a bilinear map $[\bullet, \bullet] : V \times V \to V$ called the Lie bracket such that for all $u, v, w \in V$:

(i.) $[u, v] = -[v, u]$ and

(ii.) $[u, [v, w]] + [v, [w, u]] + [w, [u, v]] = 0$.

By linearity if the Lie brackets are known on the basis elements of $\mathfrak{g}$ then they may be found for the full $\mathfrak{g}$. If $\{X_a\} \in \mathfrak{g}$ are a basis then

$$[X_a, X_b] = \sum_c f_{ab}^c X_c \tag{4.119}$$

where $f_{ab}^c = -f_{ba}^c$ are called the structure constants.

**Problem 4.5.4.** *Show that the Pauli matrices $\boldsymbol{\sigma}_i$ satisfy $[\boldsymbol{\sigma}_i, \boldsymbol{\sigma}_j] = 2i\epsilon_{ijk}\boldsymbol{\sigma}_k$.*

**Problem 4.5.5.** *Let $X_1 = X$, $X_2 = Y$ and $X_3 = Z$ where $X$, $Y$ and $Z$ are elements of the Lie algebra of $SO(3)$ as defined in equation (4.114). Show that $[X_i, X_j] = i\epsilon_{ijk}X_k$.*

### 4.5.3 Examples of interest in theoretical physics

**Example** The Double Cover of $SO(3)$ From problems 4.5.4 and 4.5.5 we have $[\boldsymbol{\sigma}_i, \boldsymbol{\sigma}_j] = 2i\epsilon_{ijk}\boldsymbol{\sigma}_k$ and $[X_i, X_j] = i\epsilon_{ijk}X_k$. Hence we can see there is an isomorphism between the Lie algebras of $SU(2)$ and $SO(3)$ given by $X_i = \frac{\sigma_i}{2}$, i.e.

$$\begin{aligned}[X_i, X_j] &= [\frac{\sigma_i}{2}, \frac{\sigma_j}{2}] \\ &= \frac{1}{4}(2i\epsilon_{ijk}\sigma_k) \\ &= i\epsilon_{ijk}X_k.\end{aligned} \tag{4.120}$$

Given the isomorphism between the two Lie algebras we may wonder whether the two groups $SU(2)$ and $SO(3)$ are isomorphic. To do this we look for a group homomorphism $\Phi : SU(2) \to SO(3)$ derived from the Lie algebra isomorphism $\phi(\frac{\sigma_i}{2}) = X_i$ and given by

$$\Phi(\exp{(\frac{i\alpha}{2}\hat{\mathbf{n}} \cdot \boldsymbol{\sigma})}) = \exp{(i\alpha\hat{\mathbf{n}} \cdot \mathbf{X})}$$

where $\mathbf{X}$ is the vector whose components are the matrices $X_i$ which form a basis for the Lie algebra of $SO(3)$. The matrix $\exp{(i\alpha\hat{\mathbf{n}} \cdot \mathbf{X})}$ is a rotation about the axis parallel with $\hat{\mathbf{n}}$ of angle $\alpha$. While we know from completing problem 4.5.3 that

$$\exp{(\frac{i\alpha}{2}\hat{\mathbf{n}} \cdot \boldsymbol{\sigma})} = \cos{(\frac{\alpha}{2})}\mathbb{I} + i\hat{\mathbf{n}} \cdot \boldsymbol{\sigma} \sin{(\frac{\alpha}{2})} \tag{4.121}$$

which covers the group elements of $SU(2)$ when $0 \leq \frac{\alpha}{2} < 2\pi$ i.e. when $0 \leq \alpha < 4\pi$. On the other hand this range of *alpha* corresponds to roatations with angle $0 \leq \alpha < 4\pi$ in $SO(3)$ under the homomorphism. That is the homomorphism gives a double-covering of $SO(3)$. The kernel of the homomorphism is non-trivial. Due to the geometrical intuition we have of the rotatations in $SO(3)$ we know that a rotation by $2\pi$ is the identity element - and we may quickly identify the kernel of $\Phi$. When $\alpha = 2\pi$ we have

$\cos(\frac{2\pi}{2})\mathbb{I} + i\hat{\mathbf{n}}\cdot\boldsymbol{\sigma}\sin(\frac{2\pi}{2}) = -\mathbb{I}$, hence the kernel of $\Phi$ is $\{\mathbb{I}, -\mathbb{I}\} \cong \mathbb{Z}_2$. So by the first isomorphism theorem we have

$$\frac{SU(2)}{\mathbb{Z}_2} \cong SO(3). \tag{4.122}$$

Let us summarise our observations. We commenced with an isomorphism between representations of two Lie algebras and we wondered whether it extended by the exponential map to an isomomorphism between the representations of the Lie groups. However the identification of the group representation (which is informed by the global group structure) with the exponentiation of the Lie algebra representation is only possible for a certain class of groups. Such groups are called simply-connnected and in addition to being connected, every closed loop on them may be continuously shrunk to a point. In this class of groups one can make deductions about the global group structure from the local knowledge of the Lie algebra. We will not discuss simple-connectedness in any detail here, but in the example above both $SU(2)$ and $SO(3)$ are connected but only $SU(2)$ is simply-connected. Hence for $SU(2)$ we may identify the representations of the group with those of the algebra but for $SO(3)$ we may not. A Lie algebra homomorphism does noy in general give a Lie group homomorphism. However if $G$ is a connected group then there always exists a related simply-connected group $\tilde{G}$ called the universal covering group for which the Lie algebra homomorphism does extend to a Lie group homomorphism. Above we see that $SU(2)$ is the universal covering group of $SO(3)$. The double cover of the group $SO(p,q)$ is the universal covering group of $SO(p,q)$ and is called $Spin(p,q)$, hence here we see that $Spin(3) \cong SU(2)$.

**Example** The Infinitesimal Generators of $SO(1,3)$. Recall that the Lorentz group $O(1,3)$ is defined by

$$O(1,3) \equiv \{\Lambda \in GL(4,\mathbb{R}) | \Lambda^T \eta \Lambda = \eta; \eta \equiv diag(1,-1,-1,-1)\}$$

In addition to rotations (in the three-dimensional spatial subspace parameterised by $\{x,y,z\}$ which are generated by $X_1$, $X_2$ and $X_3$ in the notation of the previous section) and reflections ($t \to -t$, $x \to -x$, $y \to -y$, $z \to -z$) the Lorentz group includes three Lorentz boosts. The proper Lorentz group consists of $\Lambda$ such that $Det(\Lambda) = 1$ and is the group $SO(1,3)$. The orthochoronous Lorentz group is the subgroup which preserves the direction of time, having ${\Lambda^0}_0 \geq 1$. The orthochronous proper Lorentz group is sometimes denoted $SO^+(1,3)$. The proper Lorentz group $SO(1,3)$ consists of just the rotations and boosts. The Lorentz boosts are the rotations which rotate each of $x$, $y$ and $z$ into the time direction and are represented by the generalisation of the matrix shown in equation (2.30):

$$\Lambda_1(\theta) = \begin{pmatrix} \cosh\theta & -\sinh\theta & 0 & 0 \\ -\sinh\theta & \cosh\theta & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad \Lambda_2(\theta) = \begin{pmatrix} \cosh\theta & 0 & -\sinh\theta & 0 \\ 0 & 1 & 0 & 0 \\ -\sinh\theta & 0 & \cosh\theta & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad \text{and}$$

$$\Lambda_3(\theta) = \begin{pmatrix} \cosh\theta & 0 & 0 & -\sinh\theta \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -\sinh\theta & 0 & 0 & \cosh\theta \end{pmatrix}. \tag{4.123}$$

Using

$$-iY_i \equiv \left.\frac{d\Lambda_i}{d\theta}\right|_{\theta=0} \tag{4.124}$$

we identify a basis for the Lorentz boosts in the Lie algebra $\mathfrak{so}(1,3)$:

$$Y_1 = \begin{pmatrix} 0 & -i & 0 & 0 \\ -i & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad Y_2 = \begin{pmatrix} 0 & 0 & -i & 0 \\ 0 & 0 & 0 & 0 \\ -i & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \quad \text{and} \quad Y_3 = \begin{pmatrix} 0 & 0 & 0 & -i \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ -i & 0 & 0 & 0 \end{pmatrix}. \tag{4.125}$$

The remainder of the Lie algebra of the proper Lorentz group is made up of the generators of rotations:

$$X_1 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -i \\ 0 & 0 & i & 0 \end{pmatrix}, \quad X_2 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & i \\ 0 & 0 & 0 & 0 \\ 0 & -i & 0 & 0 \end{pmatrix} \quad \text{and} \quad X_3 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & -i & 0 \\ 0 & i & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}. \tag{4.126}$$

Computation of the commutators gives (after some time...)

$$[X_i, X_j] = i\epsilon_{ijk}X_k, \quad [X_i, Y_j] = i\epsilon_{ijk}Y_k \quad \text{and} \quad [Y_i, Y_j] = -i\epsilon_{ijk}X_k. \tag{4.127}$$

It is worth observing that teh generators for the rotations are skew-symmetric matrices $X_i^T = -X_i$ while the boost generators are symmetric matrices $Y_i^T = Y_i$ for $i \in \{1,2,3\}$. This is a consequence of the rotations being an example of a compact transformation (all the components of the matrix representation of the rotation $(\cos\theta, \pm\sin\theta)$ in the group are bounded) while the Lorentz boosts are non-compact transformations (some of the components of the matrix representation of the boosts $(\cosh\theta, -\sinh\theta)$ in the group are unbounded - they may go to $\infty$.)

Notice that if one uses the combinations

$$W_i^{\pm} \equiv \frac{1}{2}(X_i \pm iY_i) \tag{4.128}$$

as a basis of the Lie algebra then the commutator relations simplify:

$$\begin{aligned} &[W_i^+, W_j^+] = i\epsilon_{ijk}W_k^+ && \mathfrak{su}(2) \\ &[W_i^-, W_j^-] = i\epsilon_{ijk}W_k^- && \mathfrak{su}(2) \\ &[W_i^+, W_j^-] = 0. \end{aligned} \tag{4.129}$$

Via a change of basis for the Lie algebra we recognise that it encodes two copies of the algebra $\mathfrak{su}(2)$:

$$\mathfrak{so}(1,3) \cong \mathfrak{su}(2) \oplus \mathfrak{su}(2). \tag{4.130}$$

The algebra isomorphism lifts to the group isomorphism $SO(1,3) \cong SL(2,\mathbb{C})$ (which we will exhibit in the next example) as $SL(2,\mathbb{C})$ is the double cover of the proper Lorentz group, i.e. $Spin(1,3) \cong SL(2,\mathbb{C})$. Consequently one can label the representations of $SO(1,3)$ by the dimensions, or spins, of the two $SU(2)$ irreducible representations. We note that the proper orthochronous Lorentz group $SO^+(1,3)$ is isomorphic to $\frac{SL(2,\mathbb{C})}{\mathbb{Z}_2}$.

**Example** The Proper Lorentz Group and $SL(2,\mathbb{C})$. Let us recall the Pauli matrices and introduce the identity matrix as $\sigma_0$:

$$\sigma_0 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{4.131}$$

Consider for each Lorentz vector $\mathbf{x} \in \mathbb{R}^{1,3}$ the map two-by-two matrix given by

$$X \equiv x^\mu \sigma_\mu = \begin{pmatrix} x^0 - x^3 & -x^1 + ix^2 \\ -x^1 - ix^2 & x^0 + x^3 \end{pmatrix} \tag{4.132}$$

so that

$$Det(X) = (x^0)^2 - (x^3)^2 - (x^1)^2 - (x^2)^2 = x^\mu x_\mu. \tag{4.133}$$

Consequently the transformations on $X$ which leave its determinant unaltered are the Lorentz transformations. One may confirm that matrices $A \in SL(2,\mathbb{C})$ transforming $X \to X'$ by the action

$$X \to X' \equiv AXA^\dagger \tag{4.134}$$

preserve $Det(X)$: $Det(X') = Det(AXA^\dagger) = Det(X)$ as $Det(A) = 1$. Hence each $A \in SL(2,\mathbb{C})$ encodes a proper Lorentz transformation on $\mathbf{x}$. To discover the precise transformation one considers the components of $\mathbf{x}$ which are simply related to $X$. By direct computation we can check that

$$\sigma_i \sigma_j = \begin{cases} i\epsilon_{ijk}\sigma_k & i \neq j \\ \delta_{ij}\sigma_0 & i = j \end{cases} \tag{4.135}$$

and

$$X\sigma_\nu = x^\mu \sigma_\mu \sigma_\nu = x^0\sigma_\nu + x^i\sigma_i\sigma_\nu = \begin{cases} x^0\sigma_0 + x^i\sigma_i & \nu = 0 \\ x^0\sigma_j + x^i\sigma_i\sigma_j & \nu = j \begin{cases} x^0\sigma_j + ix^i\epsilon_{ijk}\sigma_k & j \neq i \\ x^0\sigma_j + x^i\delta_{ij}\sigma_0 & j = i \end{cases} \end{cases}$$

As $Tr(\sigma_0) = 2$ while $Tr(\sigma_i) = 0$ we have

$$Tr(X\sigma_\nu) = 2x_\nu \qquad \Rightarrow \qquad x_\nu = \frac{1}{2}Tr(X\sigma_\nu) \tag{4.136}$$

and we have used the Minkowski metric to lower indices where necessary. We leave the exercise of finding the proper Lorentz transformation corresponding to each matrix of $SL(2,\mathbb{C})$ to the following problem.

**Problem 4.5.6.** *Let $X = x^\mu\sigma_\mu$ and show that the Lorentz transformation $x'^\mu = \Lambda^\mu{}_\nu x^\nu$ induced by $X' = AXA^\dagger$ has:*

$$\Lambda^\mu{}_\nu(A) = \frac{1}{2}Tr(A\sigma_\mu A^\dagger \sigma_\nu)$$

*thus defining a map $A \to \Lambda(A)$ from $SL(2,\mathbb{C})$ into $SO(1,3)$. Where $\sigma_0$ is the two-by-two identity matrix and $\sigma_i$ are the Pauli matrices as defined in question 4.2. (Method: show first that $Tr(X\sigma_\nu) = 2x^\nu$, then find the expression for the Lorentz transform of $x^\nu \to x'^\nu$ associated to $X \to X'$. Finally set $x$ to be the 4-vector with all components equal to zero*

*apart from the $x_\mu$ component which is equal to one.)*
*By considering a further transformation $X'' = BX'B^\dagger$ show that:*

$$\Lambda(BA) = \Lambda(B)\Lambda(A)$$

*so that the mapping is a group homomorphism. Identify the kernel of the homomorphism as the centre of $SU(2)$ i.e. $A = \pm\mathbb{I}$, thus showing that the map is two-to-one.*

**Example** The Poincaré Group. The Poincaré group is the group of isometries of Minkwski spacetime. It includes the translations in Minkowski space in addition to the Lorentz transformations:

$$\{(\Lambda, a)|\Lambda \in O(1,3), a \in \mathbb{R}^{1,3}\} \tag{4.137}$$

a general transformation of the Poincaé froup takes the form

$$x'^\mu = \Lambda^\mu{}_\nu x^\nu + a^\mu. \tag{4.138}$$

The Poincaré group is ten-dimensional and the abelian group of translations form a normal subgroup.

**Example** Representations of the Lorentz Group and Lorentz Tensors. The most simple representations of the Lorentz group are scalars. Scalar objects being devoid of free Lorentz indices form trivial representation of the Lorentz group (objects which are invariant under the Lorentz transformations). The standard vector represesntation of the Lorentz group on $\mathbb{R}^{1,3}$ acts as

$$x^\mu \to x'^\mu = \Lambda^\mu{}_\nu x^\nu. \tag{4.139}$$

This is the familiar vector action of $\Lambda$ on $\mathbf{x}$ and we shall denote it by $\Pi^{(1,0)}$.

Similarly one may define the contragredient, or co-vector, representation $\Pi^{(0,1)}$acting on co-vectors as

$$x_\mu \to x'_\mu = \Lambda^\nu{}_\mu x_\nu. \tag{4.140}$$

**Problem 4.5.7.** *Show that $\Pi^{(1,0)}$ and $\Pi^{(0,1)}$ are equivalent representations with the intertwining map being the Minkowski metric $\eta$.*

More general tensor representations are constructed from tensor products of the vector and co-vector representations of the Lorentz group and are called $(r, s)$-tensors:

$$\underbrace{\Pi^{(1,0)} \otimes \Pi^{(1,0)} \otimes \ldots \otimes \Pi^{(1,0)}}_{r} \otimes \underbrace{\Pi^{(0,1)} \otimes \Pi^{(0,1)} \otimes \ldots \otimes \Pi^{(0,1)}}_{s} \tag{4.141}$$

$(r, s)$-tensors have components with $r$ vector indices and $s$ co-vector indices

$$T^{\mu_1\mu_2\ldots\mu_r}{}_{\nu_1\nu_2\ldots\nu_s}$$

and under a Lorentz transformation $\Lambda$ the components transform as

$$T^{\mu_1\mu_2\ldots\mu_r}{}_{\nu_1\nu_2\ldots\nu_s} \to \Lambda^{\mu_1}{}_{\kappa_1}\Lambda^{\mu_2}{}_{\kappa_2}\ldots\Lambda^{\mu_r}{}_{\kappa_r}\Lambda^{\lambda_1}{}_{\nu_1}\Lambda^{\lambda_2}{}_{\nu_2}\ldots\Lambda^{\lambda_r}{}_{\nu_r}T^{\kappa_1\kappa_2\ldots\kappa_r}{}_{\lambda_1\lambda_2\ldots\lambda_s}. \tag{4.142}$$

There are two natural operations on the tensors that map them to other tensors:

(1.) One may act with the metric to raise and lower indices (raising an index maps an $(r,s)$ tensor to an $(r+1,s-1)$ tensor while lowering an index maps an $(r,s)$ tensor to an $(r-1,s+1)$ tensor):

$$\begin{aligned}\eta_{\rho\mu_k}T^{\mu_1\mu_2\ldots\mu_r}{}_{\nu_1\nu_2\ldots\nu_s} &= T^{\mu_1\mu_2\ldots\mu_{k-1}}{}_{\rho}{}^{\mu_{k+1}\ldots\mu_r}{}_{\nu_1\nu_2\ldots\nu_s} \\ \eta^{\rho\nu_k}T^{\mu_1\mu_2\ldots\mu_r}{}_{\nu_1\nu_2\ldots\nu_s} &= T^{\mu_1\mu_2\ldots\mu_r}{}_{\nu_1\nu_2\ldots\nu_{k-1}}{}^{\rho}{}_{\nu_{k+1}\ldots\nu_s}\end{aligned} \tag{4.143}$$

(2.) One can contract a pair of indices on an $(r,s)$ tensor to obtain an $(r-1,s-1)$ tensor:

$$T^{\mu_1\mu_2\ldots\mu_{r-1}\rho}{}_{\nu_1\nu_2\ldots\nu_{s-1}\rho} = T^{\mu_1\mu_2\ldots\mu_{r-1}}{}_{\nu_1\nu_2\ldots\nu_{s-1}}\,. \tag{4.144}$$

One may be interested in special subsets of tensors whose indices (or even a subset of indices) are symmetrised or antisymmetrised. Given a tensor one can always symmetrise or antisymmetrise a set of its indices:

- A symmetric set of indices is denoted explicitly by a set of ordinary brackets ( ) surrounding the symmetrised indices, e.g. a symmetric $(r,0)$ tensor is denoted $T^{(\mu_1\mu_2\ldots\mu_r)}$ and is constructed from the tensor $T^{\mu_1\mu_2\ldots\mu_r}$ using elements $P$ of the permutation group $S_r$:

$$T^{(\mu_1\mu_2\ldots\mu_r)} \equiv \frac{1}{r!}\sum_{P\in S_r} T^{\mu_{P(1)}\mu_{P(2)}\cdots\mu_{P(r)}} \tag{4.145}$$

  so that under an interchange of neighbouring indices the tensor is unaltered, e.g.

$$T^{(\mu_1\mu_2\ldots\mu_r)} = T^{(\mu_2\mu_1\ldots\mu_r)}. \tag{4.146}$$

  One may wish to symmetrise only a subset of indices, for example symmetrising only the first and last indices on the $(r,0)$ tensor is denoted by $T^{(\mu_1|\mu_2\ldots\mu_{r-1}|\mu_r)}$ and defined by

$$T^{(\mu_1|\mu_2\ldots\mu_{r-1}|\mu_r)} \equiv \frac{1}{2!}\sum_{P\in S_2} T^{\mu_{P(1)}\mu_2\cdots\mu_{r-1}\mu_{P(r)}} \tag{4.147}$$

  the pair of vertical lines indicates the set of indices omitted from the symmetrisation.

- An antisymmetric set of indices is denoted explicitly by a set of square brackets [ ] surrounding the antisymmetrised indices, e.g. an antisymmetric $(r,0)$ tensor is denoted $T^{[\mu_1\mu_2\ldots\mu_r)]}$ and is constructed from the tensor $T^{\mu_1\mu_2\ldots\mu_r}$ using elements $P$ of the permutation group $S_r$:

$$T^{[\mu_1\mu_2\ldots\mu_r]} \equiv \frac{1}{r!}\sum_{P\in S_r} Sign(P)T^{\mu_{P(1)}\mu_{P(2)}\cdots\mu_{P(r)}} \tag{4.148}$$

  so that under an interchange of neighbouring indices the tensor picks up a minus sign e.g.

$$T^{[\mu_1\mu_2\ldots\mu_r]} = -T^{[\mu_2\mu_1\ldots\mu_r]}. \tag{4.149}$$

Frequently in theoretical physics the symmetry or antisymmetry of teh indices on a tensor will be assumed and not written explicitly (which can cause confusion). For example we might define $g_{\mu\nu}$ to be a symmetric tensor which means that $g_{[\mu\nu]} = 0$ while $g_{(\mu\nu)} = g_{\mu\nu}$. Similarly for the Maxwell field strength $F_{\mu\nu}$ which was defined to be antisymmetric hence $F_{[\mu\nu]} = F_{\mu\nu}$ while $F_{(\mu\nu)} = 0$.

**Example** The symmetric and antisymmetric projections of a $(2,0)$ tensor.

$$\begin{aligned} T^{(\mu\nu)} &\equiv \frac{1}{2}(T^{\mu\nu} + T^{\nu\mu}) \qquad & \therefore\ T^{(\mu\nu)} &= T^{(\nu\mu)} \\ T^{[\mu\nu]} &\equiv \frac{1}{2}(T^{\mu\nu} - T^{\nu\mu}) \qquad & \therefore\ T^{[\mu\nu]} &= -T^{[\nu\mu]} \end{aligned} \tag{4.150}$$

N.B. the symmetric and antisymmetric projections recombine to form the original tensor.

$$T^{(\mu\nu)} + T^{[\mu\nu]} = T^{\mu\nu}. \tag{4.151}$$

**Problem 4.5.8.** *Consider the space of rank $(3,0)$-tensors $T^{\mu_1\mu_2\mu_3}$ forming a tensor representation of the Lorentz group $SO(1,3)$ which transforms under the Lorentz transformation $\Lambda$ as*

$$T'^{\nu_1\nu_2\nu_3} = \Lambda^{\nu_1}{}_{\mu_1}\Lambda^{\nu_2}{}_{\mu_2}\Lambda^{\nu_3}{}_{\mu_3}T^{\mu_1\mu_2\mu_3}.$$

*(a.) Prove that*

$$T^2 \equiv T_{\mu_1\mu_2\mu_3}T^{\mu_1\mu_2\mu_3}$$

*is a Lorentz invariant. The Einstein summation convention for repeated indices is assumed in the expression for $T^2$.*

*(b.) Give the definitions of the symmetric $(3,0)$-tensors and of antisymmetric $(3,0)$-tensors and show that they form two invariant subspaces under the Lorentz transformations.*

*(c.) Prove that the symmetric $(3,0)$-tensors form a reducible representation of the Lorentz group.*

# Chapter 5

# Special topics in Mathematical Physics

In this chapter we review briefly a number of subjects that will be useful throughout the study of theoretical physics. In a number of places we will follow closely the elegant presentation of Sadri Hassani in his excellent book *Mathematical physics - a modern introduction to its foundations.*

Before doing so it will be useful to remind ourselves of the basic differential equations that appear in physics. The search for solutions to these equations will motivate the rest of the chapter.

## 5.1 Famous Differential Equations in Physics

The most popular equation in physics is the wave equation which is almost ubiquitous. We can derive the wave equation by considering a string of $n$ one-dimensional simple harmonic oscillators where the $i$'th oscillator is attached at one end to the $(i-1)$'th and at the other end to the $(i+1)$'th oscillator. We will assume all masses are identical and have coordinate $q_i$ where $i = 1, 2, 3, \ldots n$. The force on the $i$'th mass gives its equations of motion:

$$\begin{aligned} -k(q_i - q_{i-1}) + k(q_{i+1} - q_i) &= m\ddot{q}_i \\ \therefore \frac{m}{k}\ddot{q}_i &= q_{i-1} - 2q_i + q_{i+1}. \end{aligned} \tag{5.1}$$

If we take the limit $(n \to \infty, k \to \infty, m \to 0, d \to 0)$ (where $d$ is the natural length of the springs in each oscillator) such that the total mass $M \equiv nm$, total length $L \equiv nd$ and also $\tau \equiv kd$ all remain finite. Why do we insist that these quantities are to remain finite? Consider first the natural length of the harmonic oscillator $d$, in our limit this will go to zero which will correspond to a continuum of masses along a string. Now we make the following observation:

$$\begin{aligned} \frac{\partial q}{\partial x} &= \lim_{\delta x \to 0}\left[\frac{q(x+\delta x) - q(x)}{\delta x}\right] \\ \therefore \frac{\partial^2 q}{\partial x^2} &= \lim_{\delta x \to 0}\left[\frac{1}{\delta x}\left(\frac{q(x+2\delta x) - q(x+\delta x)}{\delta x} - \frac{q(x+\delta x) - q(x)}{\delta x}\right)\right] \end{aligned} \tag{5.2}$$

where in our transformation from the discrete masses to the continuum we have taken $d = \delta x$. Hence we have that

$$\frac{\partial^2 q}{\partial x^2} = \lim_{d\to 0}\left[\frac{1}{d^2}\Big(q(x_{i+1}) - 2q(x_i) + q(x_{i-1})\Big)\right] \tag{5.3}$$

where $x_i = id$ and $q_i = q(x_i)$. So, returning to our equation of motion, we are motivated to pre-multiply both sides by $\frac{1}{d^2}$ and take the full continuum limit (i.e. $(n \to \infty, k \to \infty, m \to 0, d \to 0)$):

$$\begin{aligned}\lim\left[\frac{m}{kd^2}\frac{\partial q(x_i,t)}{\partial t^2}\right] &= \lim\left[\frac{1}{d^2}(q_{i-1} - 2q_i + q_{i+1})\right] \\ &= \frac{\partial^2}{\partial x^2}\Big(q(x,t)\Big).\end{aligned} \tag{5.4}$$

Now consider what happens to the term $\frac{m}{kd^2}$ when we take the continuum limit:

$$\lim\left[\frac{m}{kd^2}\right] = \lim\left[\frac{mn}{(kd)(nd)}\right] = \frac{M}{\tau L} \tag{5.5}$$

which is finite and gives the reason for insisting that $nm$, $nd$ and $kd$ remain finite in the continuum limit. Now in all we have the continuum equation of motion:

$$\frac{M}{\tau L}\frac{\partial^2 q(x,t)}{\partial t^2} = \frac{\partial^2}{\partial x^2}\Big(q(x,t)\Big). \tag{5.6}$$

Notice that the units of $\frac{M}{\tau L}$ are $\frac{1}{(ms^{-1})^2}$ which are the same as those of $\frac{1}{v^2}$ where $v$ is a velocity. If we define $\frac{1}{v^2} \equiv \frac{M}{\tau L}$ then we have

$$\frac{1}{v^2}\frac{\partial^2 q}{\partial t^2} - \frac{\partial^2 q}{\partial x^2} = 0 \tag{5.7}$$

which is the usual form of the one-dimensional wave equation. It has the generic solution

$$q(x,t) = f(x - vt) + g(x + vt) \tag{5.8}$$

where $f$ and $g$ are arbitrary functions. Let us check explicitly that this is a solution of the wave equation

$$\begin{aligned}\left(\frac{1}{v^2}\frac{\partial^2}{\partial t^2} - \frac{\partial^2}{\partial x^2}\right)q &= \frac{1}{v^2}\left(\frac{\partial^2 f}{\partial t^2} + \frac{\partial^2 g}{\partial t^2}\right) - \frac{\partial^2 f}{\partial x^2} - \frac{\partial^2 g}{\partial x^2} \\ &= \frac{1}{v^2}\Big(v^2(f'' + g'')\Big) - f'' - g'' \\ &= 0.\end{aligned} \tag{5.9}$$

Let us write $s = x - vt$ and set the function $g = 0$ then we have $q = f(s)$. For constant $s$ we have $0 = ds = dx - vdt \quad \Rightarrow \quad v = \frac{dx}{dt}$. Hence a point of constant $s$ moves with speed $v$ along $x$. The points of constant $s$ are called a wavefront of constant phase. Similarly if we had set the function $f = 0$ while $g$ remained a non-zero function of $x + vt$ then a wavefront of speed $-v$ is identified moving along $x$ when $x + vt$ is held constant.

Now consider the wave equation in three dimensions. We could find similar solutions of waves with wavefronts travelling along the $\pm x$. Now however the wavefront is the entire $yz$ plane at $x$. Consequently it is called a plane wave. We would also expect to

find plane waves moving in the $y$ and $z$ directions and we may wonder how the wave equation may be enhanced to permit these solutions in three dimensions. It becomes

$$\left(\frac{1}{v^2}\frac{\partial^2}{\partial t^2} - \frac{\partial^2}{\partial x^2} - \frac{\partial^2}{\partial y^2} - \frac{\partial^2}{\partial z^2}\right)u(\mathbf{r},t) = \left(\frac{1}{v^2}\frac{\partial^2}{\partial t^2} - \nabla^2\right)u(\mathbf{r},t) = 0 \tag{5.10}$$

where $\nabla^2 = \nabla \cdot \nabla \equiv \Delta$ is the Laplacian. The general solution to the three-dimensional wave equation is

$$u(\mathbf{r},t) = f(\mathbf{r} - \mathbf{v}t) + g(\mathbf{r} + \mathbf{v}t). \tag{5.11}$$

The wave equation is, perhaps, the most famous equation in physics. Other common differential equations include:

| | |
|---|---|
| Poisson's Equation (from electrostatics) | $\nabla^2\phi(r) = -4\pi\rho(r)$ |
| Laplace's Equation (when $\rho(r) = 0$) defines a harmonic function | $\nabla^2\phi(r) = 0$ |
| The Heat Equation | $\frac{\partial T}{\partial t} = a^2\nabla^2 T(r)$ |
| Helmholtz's Equation | $\nabla^2\psi = -k\psi$ |
| Schrödinger's Equation | $\hat{\mathcal{H}}\psi = i\hbar\frac{\partial\psi}{\partial t}$ |
| The 1D Wave Equation with Friction | $\left(\frac{1}{v^2}\frac{\partial^2}{\partial t^2} + \frac{1}{\kappa}\frac{\partial}{\partial t} - \frac{\partial^2}{\partial x^2}\right)u(x,t) = 0$ |

The term $\frac{1}{\kappa}\frac{\partial u}{\partial t}$ dampens the wave function as it oscilates through space-time. If the effective mass term vanishes above (compare with $vecF = m\mathbf{a}$) then we have

$$\left(\frac{1}{\kappa}\frac{\partial}{\partial t} - \frac{\partial^2}{\partial x^2}\right)u(x,t) = 0 \tag{5.12}$$

which is the diffusion equation and $\kappa$ is the diffusion constant when we are discussing the diffusion of atoms, but $\kappa$ is called the conductivity if we are describing heat conduction.

These equations can be solved in a number of ways. Some very useful methods involve the use of orthogonal polynomials, Greens' functions or the Fourier transform. We will not discuss the Fourier transform in what follows but we will investigate the other methods for solving differential equations.

## 5.2 Classical Orthogonal Polynomials

We have seen the appearance of $L^2$ functions in our review of quantum mechanics. It is natural to wonder whether there is a canonical basis of such functions which are orthogonal. If so we may expand a function $\psi(x)$ in such a basis $F_n(x)$,in bra-ket notation:

$$|\psi\rangle = \sum_n a_n |F_n\rangle \tag{5.13}$$

and quickly interrogate the coefficients $a_n$ using the orthogonality of $|F_n\rangle$,

$$\langle F_m|\psi\rangle = \sum_n a_n \langle F_m|F_n\rangle = a_m \langle F_m|F_m\rangle \quad \Rightarrow \quad a_m = \frac{\langle F_m|\psi\rangle}{\langle F_m|F_m\rangle}. \tag{5.14}$$

But what are the $F_n(x)$ and under which conditions do they exist. We will now follow a very general procedure and develop the classical orthogonal polynomials in a single expression.

**Theorem 5.2.1.** *Let*

$$F_n(x) = \frac{1}{w(x)} \frac{d^n}{dx^n}(w(x)s^n(x)) \qquad \text{for} \quad n = 0, 1, 2, \ldots \tag{5.15}$$

*where $F_1(x)$ is a first order polynomial in $x$, $s(x)$ is a polynomial in $x$ of degree two or less with only real roots and $w(x)$, the weight function, is a strictly positive function integrable in the region $(a, b)$ satisfying $w(a)s(a) = w(b)s(b) = 0$. Then $F_n(x)$ is a polynomial of degree $n$ in $x$ and*

$$\int_a^b p_k(x) F_n(x) w(x) dx = 0 \qquad \forall \quad k < n \tag{5.16}$$

*where $p_k(x)$ is any polynomial of degree $k$.*

The formulation of $F_n(x)$ above is sufficiently versatile to include all the classical orthogonal polynomials. By specifying precisely the function $s(x)$ we will be able to determine $w(x)$ and the range $(a, b)$ and for each choice we will find a new series of orthogonal polynomials which will be useful in different settings. These will be described after we have proved the theorem. In order to do this we will make use of two lemmas.

**Lemma 5.2.2.** *The following is an identity:*

$$\frac{d^m}{dx^m}(ws^n p_{\leq k}) = ws^{n-m} p_{\leq k+m} \qquad \forall \quad m \leq n \tag{5.17}$$

*Proof.* For $m = 1$ the we have

$$\frac{d}{dx}(ws^n p_{\leq k}) = \frac{dw}{dx} s^n p_{\leq k} + nws^{n-1} \frac{ds}{dx} p_{\leq k} + ws^n p_{\leq k-1} \tag{5.18}$$

where we emphasise that $p_{\leq k}$ is notation for any polynomial of degree $\leq k$, hence $\frac{dp_{\leq k}}{dx} = p_{\leq k-1}$. Looking at the above expression, and considering the lemma, we would like to count the polynomial degree of each of the three terms on the right hand side and check that we can factorise each term into terms of the form $ws^{n-1}p_{\leq k+1}$. We are

obstructed in doing this by the $\frac{dw}{dx}$ in the first term. However we can use the definition of the orthogonal polynomials in the main theorem to overcome this obstruction. We have

$$F_1(x) = \frac{1}{w}\frac{d}{dx}(ws) = \frac{1}{w}\frac{dw}{dx}s + \frac{ds}{dx} = p_1(x). \tag{5.19}$$

In the last equality we are simply noting that by definition $F_1(x)$ is a first order polynomial in $x$. Rearranging we have,

$$s\frac{dw}{dx} = wp_1 - w\frac{ds}{dx} \tag{5.20}$$

which we may use to rewrite equation (5.2.2) in a way that clearly factorsises according to the lemma. Substituting we find

$$\begin{aligned}\frac{d}{dx}(ws^n p_{\leq k}) &= (wp_1 - w\frac{ds}{dx})s^{n-1}p_{\leq k} + nws^{n-1}\frac{ds}{dx}p_{\leq k} + ws^n p_{\leq k-1} \\ &= wp_1 s^{n-1}p_{\leq k} - w\frac{ds}{dx}s^{n-1}p_{\leq k} + nws^{n-1}\frac{ds}{dx}p_{\leq k} + ws^n p_{\leq k-1} \\ &= ws^{n-1}p_{\leq k+1} + (n-1)ws^{n-1}\frac{ds}{dx}p_{\leq k} + ws^n p_{\leq k-1} \\ &= ws^{n-1}p_{\leq k+1}\end{aligned} \tag{5.21}$$

where in the last line we have used the fact that $s(x)$ is a polynomial of degree two or less, i.e. $s = p_{\leq 2}$, and hence $\frac{ds}{dx} = p_{\leq 1}$. Taking additional derivatives and repeating the process gives the proof of the lemma. We will need a second lemma whch is derived from lemma 5.2.2 above.

**Lemma 5.2.3.** *All the $m$ derivatives of $ws^n$ vanish at $x = a, b$ for all $m < n$, i.e.*

$$\frac{d^m}{dx^m}(ws^n)\bigg|_{x=a,b} = 0 \qquad \forall \quad m \leq n. \tag{5.22}$$

*Proof.* From lemma 5.2.2, when $k = 0$ we have

$$\frac{d^m}{dx^m}(ws^n) = ws^{n-m}p_{\leq m} \tag{5.23}$$

and at $x = a, b$ we have

$$\frac{d^m}{dx^m}(ws^n)\bigg|_{x=a,b} = (ws)\bigg|_{x=a,b}(s^{n-m-1}p_{\leq m})\bigg|_{x=a,b} = 0. \tag{5.24}$$

Armed with this lemma we can now prove the main theorem.

*Proof.*

$$\begin{aligned}\int_a^b p_k F_n w dx &= \int_a^b p_k \frac{d^n}{dx^n}(ws^n)dx \\ &= \int_a^b p_k \frac{d}{dx}(\frac{d^{n-1}}{dx^{n-1}}(ws^n))dx \\ &= \int_a^b -\frac{dp_k}{dx}\frac{d^{n-1}}{dx^{n-1}}(ws^n)dx + \underbrace{\left[p_k\frac{d^{n-1}}{dx^{n-1}}(ws^n)\right]_a^b}_{=0 \text{ by lemma 5.2.3}}\end{aligned} \tag{5.25}$$

Repeating the integration by parts $k$ times gives

$$\begin{aligned}\int_a^b p_k F_n w dx &= \int_a^b (-1)^k \frac{d^k p_k}{dx^k}\frac{d^{n-k}}{dx^{n-k}}(ws^n)dx \\ &= C\int_a^b \frac{d}{dx}(\frac{d^{n-k-1}}{dx^{n-k-1}}(ws^n))dx \\ &= C\left[\frac{d^{n-k-1}}{dx^{n-k-1}}(ws^n)\right]_a^b \\ &= 0\end{aligned} \tag{5.26}$$

where $C = (-1)^k \frac{d^k p_k}{dx^k}$ is a constant and we have used lemma 5.2.3 again in the final line. To prove that $F_n(x)$ is a polynomial of degree $n$ we use lemma 5.2.2 to note that $\frac{d^n}{dx^n}(ws^n) = wp_{\leq n}$ and so

$$F_n(x) = \frac{1}{w}\frac{d^n}{dx^n}(ws^n) = p_{\leq n}. \tag{5.27}$$

To show that $F_n$ is a polynomial of degree n we write

$$F_n(x) = a_n x^n + p_{\leq n-1} \tag{5.28}$$

where $a_n$ is a constant and we note that the repeated index does not indicate the summation convention here. Now consider the positive definite integral (and use the power of the orthogonal polynomials):

$$\begin{aligned}h_n \equiv \int_a^b F_n^2 w dx &= \int_a^b F_n(a_n x^n + p_{\leq n-1})w dx \\ &= \int_a^b F_n a_n x^n w dx + \underbrace{\int_a^b F_n p_{\leq n-1})w dx}_{=0}\end{aligned} \tag{5.29}$$

where we have used the second part of the theorem (which we have proved first) to see that the second integral vanishes. Consequently as the left hand side is positive definite then $a_n \neq 0$ and so $F_n(x) = p_n$. The classical orthogonal polynomials were not found as presented above. They were each found individually as solutions to particular differential equations. Consequently the historical definitions of orthogonal polynomials are all normalised individually with a factor of $\frac{1}{K_n}$, hence we will now adopt the normalised function as our definition of the orthogonal polynomial $F_n(x)$:

$$F_n(x) \equiv \frac{1}{K_n w}\frac{d^n}{dx^n}(ws^n). \tag{5.30}$$

The expression above is called the generalised Rodrigues formula after (Benjamin) Olinde Rodrigues, a Frenchman who first wrote down this formula in 1816. It was later independently discovered by Sir James Ivory in 1824 and Karl Jacobi in 1827. After writing down this formula Rodrigues became a banker. Some time later Rodrigues contributed again to mathematics and has a strong claim to discovering the quaternions three years before Hamilton and without the use of a bridge.

In table 5.2.1 we list the orthogonal polynomials and the classification has been organised according to the degree of $s(x)$. It is important to note the range of the

| Name of $F_n(x)$ | $s(x)$ | $w(x)$ | $[a,b]$ |
|---|---|---|---|
| The Hermite polynomials, $H_n$. | $p_0$ | $e^{-x^2}$ | $[-\infty,+\infty]$ |
| The Laguerre polynomials, $L_n^\nu(x)$. | $x$ | $x^\nu e^{-x}$, $\nu>-1$ | $[0,+\infty]$ |
| The Jacobi polynomials, $P_n^{\mu,\nu}(x)$. | $1-x^2$ | $(1-x)^\mu(1-x)^\nu$, $\mu,\nu>-1$ | $[-1,+1]$ |

Table 5.2.1: The classical series of orthogonal polynomials.

integration where these polynomials are orthogonal, this gives a good indication of where these polynomials may be useful in theoretical physics. For example, the Hermite polynomials, which we will consider in more detail, will be useful in Hilbert space and quantum mechanics, the Laguerre polynomials are useful when separating and solving the radial part of the Schrödinger equation for hydrogen, and the Jacobi polynomials occur in the study of rotation groups. The Laguerre and Jacobi polynomial series are paramterised by $\nu$ and $(\mu,\nu)$ respectively and specific vaues of these parameters select out interesting sub-series of the orthogonal polynomials. The most famous case of this are the Legendre polynomials which are a sub-class of the Jacobi polynomials having $\mu=0$, $\nu=0$ and hence $w(x)=1$. The Legendre polynomials occur when finding solutions of the Laplace equation in spherical coordinates (using the method of separation of variables), consequently they are closely related to spherical harmonics. The Legendre polynomials occur naturally in multipole expansions, recall this is the expansion resulting from translation of a spherically symmetric point source and the mutlipole exapnsion can be usefully rewritten as a sum of Legendre polynomials.

### 5.2.1 Recurrence Relations

Orthogonal polynomials are often defined via their recurrence relations, these are relations between neighbouring polynomials, e.g. relating $F_{n-1}$, $F_n$ and $F_{n+1}$. The recurrence relations are derived by first expanding out a generic orthogonal polynomial and singling out its $x^{k-1}$ and $x^k$ terms:

$$F_k(x)=p_{<k-2}+a_k'x^{k-1}+a_kx^k. \tag{5.31}$$

Next we note that

$$F_{n+1}-\frac{a_{n+1}}{a_n}xF_n=p_{\leq n} \tag{5.32}$$

and so we may use the fact that the orthogonal polynomials form a basis of the polynomials to write

$$F_{n+1}-\frac{a_{n+1}}{a_n}xF_n=\sum_{m=0}^{n}C_mF_m \tag{5.33}$$

where $C_m$ are constants. Now,

$$\int_a^b F_{n+1}F_pwdx-\frac{a_{n+1}}{a_n}\int_a^b xF_nF_pwdx=\sum_{m=0}^{n}C_m\int_a^b F_mF_pwdx. \tag{5.34}$$

If $p\leq n-2$ then both $\int_a^b F_{n+1}F_pwdx=0$ and $\int_a^b xF_nF_pwdx=0$ and hence

$$\sum_{m=0}^{n}C_m\int_a^b F_mF_pwdx=0 \qquad \forall \quad p\leq n-2. \tag{5.35}$$

The integral above is only non-zero when $m = p$ so we have,

$$C_p \int_a^b F_p^2 w dx = 0 \qquad \forall \quad p \leq n-2. \tag{5.36}$$

But the integral is postive by definition and does not vanish so we deduce that all the coefficients $C_p = 0$ for $p \leq n-2$. Substituting these coefficients into equation (5.33) we are left with

$$\begin{aligned} F_{n+1} - \frac{a_{n+1}}{a_n} x F_n &= C_{n-1} F_{n-1} + C_n F_n \\ \Rightarrow F_{n+1} &= \left( \frac{a_{n+1}}{a_n} x + C_n \right) F_n + C_{n-1} F_{n-1} \end{aligned} \tag{5.37}$$

which gives a general recurrence relation for a series of orthogonal polynomials. One can write the recurrence relation in a simpler form using the following definitions[1]

$$\begin{aligned} &h_k \equiv \int_a^b F_k^2 w dx, \\ &\alpha_n \equiv \frac{a_{n+1}}{a_n}, \quad \beta_n \equiv \alpha_n \left( \frac{a'_{n+1}}{a_{n+1}} - \frac{a'_n}{a_n} \right) \quad \text{and} \quad \gamma_n \equiv -\frac{h_n}{h_{n-1}} \frac{\alpha_n}{\alpha_{n-1}}. \end{aligned} \tag{5.38}$$

Using this notation the recurrence relation becomes

$$F_{n+1} = (\alpha_n x + \beta_n) F_n + \gamma_n F_{n-1} \tag{5.39}$$

and the coeffiecients $\alpha_n$, $\beta_n$ and $\gamma_n$ are simple to compute. For particular orthogonal polynomials the recurrence relations are straightforward, for example for the Hermite polynomials $H_n$

$$\frac{dH_n}{dx} = 2nH_{n-1} \tag{5.40}$$

and for the Legendre polynomials $P_n$

$$(1-x^2)\frac{dP_n}{dx} + nxP_n - nP_{n-1} = 0. \tag{5.41}$$

One can use the recurrence relations to find explicit expressions for the orthogonal polynomials $F_n$. We will consider the example of the Hermite polynomials $H_n$.

### 5.2.2 The Hermite Polynomials

To find the recurrence relation we will need to calculate some of the coefficients $a_n$, $a'_n$, as well as $h_n$. Let us develop an expression we saw earlier in the proof of the Rodrigues formula for orthogonal polynomials. Recall from equation (5.29) we have

$$\begin{aligned} h_n &\equiv \int_a^b a_n F_n x^n w dx \\ &= \frac{a_n}{K_n} \int_a^b x^n \frac{d^n}{dx^n}(ws^n) dx \\ &= (-1)^n \frac{a_n n!}{K_n} \int_a^b (ws^n) dx \end{aligned} \tag{5.42}$$

[1]The coefficient $\alpha_n$ is found by inspection; $\beta_n = C_n$ is found by writiing $F_{n+1} = a_{n+1}x^{n+1} + a'_{n+1}x^n + p_{<n}$ and $F_n = a_n x^n + a'_n x^{n-1} + p_{<(n-1)}$ and equating the coefficient of $x^n$ in equation (5.37); and for $\gamma_n$ note that $xF_{n-1} = \frac{a_{n-1}}{a_n} F_n + p_{\leq n-1}$ and use the orthogonality of the polynomials $F_k$ in equation (5.37).

where we have integrated by parts $n$ times and the boundary term has vanished using lemma 5.2.3. For the particular case of the Hermite polynomials we have $K_n = (-1)^n$, $w = e^{-x^2}$, $s = 1$ and $[a, b] = [-\infty, +\infty]$, i.e.

$$H_n \equiv (-1)^n e^{x^2} \frac{d^n}{dx^n}(e^{-x^2}). \tag{5.43}$$

Now the terms $h_n$ for the Hermite polynomial are

$$h_n = a_n n! \int_{-\infty}^{\infty} (e^{-x^2}) dx = a_n n! \sqrt{\pi}. \tag{5.44}$$

Recall that $a_n$ is the coefficient of the $x^n$ term in the Hermite polynomial and so we can determine it using the definition of the Hermite polynomials in equation (5.43). Each derivative of $e^{-x^2}$ brings down a factor of $(-2x)$ hence after $n$ derivatives we have

$$H_n = (-1)^n e^{x^2} (-2x)^n e^{-x^2} + p_{n-2} = 2^n x^n + p_{n-2}. \tag{5.45}$$

Now we may read off that $a_n = 2^n$. We have indicated above from our direct calculation that there are no terms of order $x^{n-1}$ in $H_n$ we may confirm this by noting that as $H_n(-x) = (-1)^n H_n(x)$ so $H_n$ must consist of all even or all odd terms, hence $a'_n = 0$. Returning to $h_n$ we have,

$$h_n = 2^n n! \sqrt{\pi}. \tag{5.46}$$

Therefore we have

$$\alpha_n = 2, \quad \beta_n = 0 \quad \text{and} \quad \gamma_n = -2n \tag{5.47}$$

and the recurrence relation for the Hermite polynomials is

$$H_{n+1} = 2xH_n - 2nH_{n-1}. \tag{5.48}$$

Hence we have for small $n$

$$\begin{aligned} H_0 &= 1 \\ H_1 &= 2x \\ H_2 &= 4x^2 - 2 \\ H_3 &= 8x^3 - 12x \\ H_4 &= 16x^4 - 48x^2 + 12. \end{aligned} \tag{5.49}$$

We may use the recurrence relation above to find a simpler recurrence relation for the Hermite polynomials. Consider the derivative of $H_n$,

$$\begin{aligned} \frac{dH_n}{dx} &= (-1)^n (2x) e^{x^2} \frac{d^n}{dx^n}(e^{-x^2}) + (-1)^n e^{x^2} \frac{d^{n+1}}{dx^{n+1}}(e^{-x^2}) \\ &= 2xH_n - H_{n+1} \\ &= 2xH_n - (2xH_n - 2nH_{n-1}) \\ &= 2nH_{n-1}. \end{aligned} \tag{5.50}$$

Using these two recurrence relations we can reconstruct the second order differential equation which the Hermite polynomials solve. Differentiating the second recurrence

relation and using the first we find:

$$\begin{aligned}\frac{d^2H_n}{dx^2} &= 2n\frac{dH_{n-1}}{dx} \\ &= \frac{d}{dx}(2xH_n - H_{n+1}) \\ &= 2H_n + 2x\frac{dH_n}{dx} - \frac{dH_{n+1}}{dx} \\ &= 2H_n + 2x\frac{dH_n}{dx} - 2(n+1)H_n \\ &= 2x\frac{dH_n}{dx} - 2nH_n\end{aligned} \tag{5.51}$$

and we arrive at the differential equation which the Hermite polynomial provide an infinite set of solutions:

$$\frac{d^2H_n}{dx^2} - 2x\frac{dH_n}{dx} + 2nH_n = 0. \tag{5.52}$$

It was this differential equation that originally motivated the discovery of the Hermite polynomials. They were first found by Laplace in 1810 and even Chebyshev discovered them in 1859 before Hermite rediscovered them independently in 1864.

**The Hermite Polynomials and the Quantum Harmonic Oscillator**

Recall that locally many potentials may be approximated by the quadratic potential of the simple harmonic oscillator. In one dimension the quantised Hamiltonian for the oscialltor is

$$\hat{\mathcal{H}} = -\frac{\hbar^2}{2m}\frac{d^2}{dx^2} + \frac{1}{2}m\omega^2x^2. \tag{5.53}$$

Schrödinger's equation for this Hamiltonian is

$$\left(-\frac{\hbar^2}{2m}\frac{d^2}{dx^2} + \frac{1}{2}m\omega^2x^2 - E\right)|\psi\rangle = 0 \tag{5.54}$$

where $E$ is the energy of the system. Let us rearrange the equation and make a simple change of variables:

$$\begin{aligned}\frac{\hbar^2}{m}\left(-\frac{1}{2}\left[\frac{d^2}{dx^2} - \frac{m^2\omega^2x^2}{\hbar^2}\right] - \frac{Em}{\hbar^2}\right)|\psi\rangle &= 0 \\ \Rightarrow \quad \left(-\frac{1}{2}\left[\frac{d^2}{dx^2} - \frac{x^2}{x_0^4}\right] - \frac{e}{x_0^2}\right)|\psi\rangle &= 0\end{aligned} \tag{5.55}$$

where we have substituted $x_0^2 = \frac{m\omega}{\hbar}$ and $E = \hbar\omega e$. Note that we are not presuming any knowledge of the discrete energy levels of the oscillator, that is we are not presuming that $e$ is a discrete variable we are simply making a convenient change of notation. The new constant $x_0$ is called the *characteristic size* of the system. Another change of variables is now in order, let $\xi = \frac{x}{x_0}$, so that $\frac{dx}{d\xi} = x_0$ and $\frac{d}{dx} = \frac{1}{x_0}\frac{d}{d\xi}$, hence we find we can simplify the equation to

$$\left(-\frac{1}{2}\left[\frac{d^2}{d\xi^2} - \xi^2\right] - e\right)|\psi\rangle = 0. \tag{5.56}$$

Now $|\psi\rangle$ must be square integrable so we expect it to vanish for large $|\xi|$. Hence at large $|\xi|$ we wish to solve

$$-\frac{1}{2}\left(\frac{d^2}{d\xi^2} - \xi^2\right)|\psi\rangle = 0. \tag{5.57}$$

This has solutions $|\psi\rangle_{\xi\to\infty} = e^{-\frac{\xi^2}{2}}$ which vanish as $\xi \to \infty$. We have found a particular solution at $\infty$ now suppose the general solution has the form

$$|\psi\rangle = H(\xi)e^{-\frac{\xi^2}{2}} \tag{5.58}$$

where $H(\xi)$ is an arbitrary function of $\xi$ then it must satisfy

$$\frac{d^2H}{d\xi^2} - 2\xi\frac{dH}{d\xi} + (2e-1)H = 0. \tag{5.59}$$

This is the same form as equation (5.52) and hence if $2e+1 = 2n$ this equation is solved if $H = H_n$ the Hermite polynomial. Consequently $e = n + \frac{1}{2}$, and takes discrete half-integer values. If $e \neq n + \frac{1}{2}$ one can check that the wavefunctions are no longer square-integrable. The wavefunctions take the form

$$|\psi\rangle = C_n H_n(\xi)e^{-\frac{1}{2}\xi^2}. \tag{5.60}$$

## 5.3 Green's Functions

Suppose we wish to solve a differential equation having the form of the Poisson equation

$$Df(x) = \rho(x) \tag{5.61}$$

then we may be aided in finding a formal solution by the Green's functions $G(x,y)$ for the differential operator $D$. These are defined by

$$DG(x,y) = \delta(x-y) \tag{5.62}$$

where $\delta$ is the Dirac delta function. We may use this to find a solution to the Poisson equation given as

$$f(x) = \int dy\rho(y)G(x,y). \tag{5.63}$$

This solves the Poisson equation as

$$\begin{aligned} Df(x) &= D(\int dy\rho(y)G(x,y)) \\ &= \int dy\rho(y)DG(x,y) \\ &= \int dy\rho(y)\delta(x-y) \\ &= \rho(x). \end{aligned} \tag{5.64}$$

However this may well be wishful thinking, as it is not always possible to identify a Green's function for every differential operator $D$. However for may equations of physical interest including the Klein-Gordon equation in quantum field theory it is possible to find the Green's function. The Green's function is effectively the inversion of the differential operator and in QFT the propagators are Green's functions. We will consider only two examples in this course, when $D = \frac{d}{dx}$ and when $D = \nabla^2$, the Laplacian in arbitrary dimension. We can immediately discuss the first example.

**Example:** $D = \frac{d}{dx}$**.**

Consider the equation

$$\frac{df}{dx} = \rho(x) \tag{5.65}$$

we must find a Green's function $G(x, y)$ such that

$$\frac{d}{dx}(G(x, y)) = \delta(x - y). \tag{5.66}$$

The function whose derivative is the Dirac delta function is the step function $\theta(x - y)$ defined by

$$\theta(x - y) \equiv \begin{cases} 0 & \text{if} \quad x < y \\ 1 & \text{if} \quad x > y. \end{cases} \tag{5.67}$$

Let us confirm in passing that the derivative of this function is indeed the Dirac delta function. Consider the function $T_\epsilon$ defined by

$$T_\epsilon(x - y) \equiv \begin{cases} 0 & \text{if} \quad x < y - \epsilon \\ \frac{1}{2\epsilon}(x - y + \epsilon) & \text{if} \quad y - \epsilon \leq x \leq y + \epsilon \\ 1 & \text{if} \quad x > y + \epsilon \end{cases} \tag{5.68}$$

where $y$ is a constant, then as $\epsilon \to 0$ then $T_\epsilon \to \theta(x - y)$. Now note that the derivative of this function integrates to one independent of $\epsilon$:

$$\int_{-\infty}^{\infty} \frac{dT_\epsilon}{dx} dx = \int_{y-\epsilon}^{y+\epsilon} \frac{1}{2\epsilon} dx = 1. \tag{5.69}$$

As we take the limit $\epsilon \to 0$ we have

$$\lim_{\epsilon \to 0} \left[ \int_{-\infty}^{\infty} \frac{dT_\epsilon}{dx} dx \right] = \int_{y-\epsilon}^{y+\epsilon} \frac{1}{2\epsilon} dx = 1. \tag{5.70}$$

### 5.3.1 The Dirac delta function and spherical coordinates.

Our aim is to find the Green's function for the Laplacian operator in $n$-dimensions. In order to do this we will need to see how the Dirac delta function transforms under a change of coordinates. First we note that in $n$-dimensions the Dirac delta function is written as a product of one-dimensional Dirac delta functions:

$$\delta(\mathbf{x} - \mathbf{y}) \equiv \delta(x_1 - y_1)\delta(x_2 - y_2) \ldots \delta(x_n - y_n) = \prod_{i=1}^{n} \delta(x_i - y_i). \tag{5.71}$$

An integral of a function $F(\mathbf{x})$ against the n-dimensional Dirac delta function gives

$$\int d^n x F(\mathbf{x}) \delta(\mathbf{x} - \mathbf{y}) = F(\mathbf{y}) \tag{5.72}$$

Suppose that we make a change of coordinates $\mathbf{x} \to \mathbf{q}$, under the change of coordinates the measure $d^n x \to |\mathcal{J}| d^n q$ where $\mathcal{J}_{ij} = \frac{\partial x_i}{\partial q_j}$ is the Jacobian of the transformation $\mathbf{x} \to \mathbf{q}$ and $|\mathcal{J}|$ is its determinant. Equation (5.72) is further transformed as $F(\mathbf{x}) \to G(\mathbf{q})$ and $\mathbf{y} \to \mathbf{p}$ giving

$$\int d^n q |\mathcal{J}| G(\mathbf{q}) \delta(\mathbf{q} - \mathbf{p}) = G(\mathbf{p}) \tag{5.73}$$

i.e. under a coordinate transformation $\delta(\mathbf{x}-\mathbf{y}) \to |\mathcal{J}|\delta(\mathbf{q}-\mathbf{p})$. The appearance of the determinant of the Jacobian appears to present a problem as it is not unreasonable that the determinant may vanish and hence the Dirac delta function would appear to no longer be defined under such a transformation. Such transformations are very common, consider the transformation from three dimensional Cartesian coordinates to spherical polar coordinates: at the origin the coordinates are independent of the angular coordinates $\theta$ and $\phi$ as the origin is defined as the point at which $r=0$. Consequently at the origin the function $G(\mathbf{0})$ is independent of $\theta$ and $\phi$ and the Jacobian of the transformation is singular and has determinant equal to zero. What we learn from this is that we have not understood the correct transformation of the Dirac delta function under a coordinate transformation. Let us suppose that the Jacobian is singular so that the function $G(\mathbf{p})$ appearing on the right-hand-side of equation (5.74) depends upon only $k<n$ components of the vector in $\mathbb{R}^n$, i.e. $G(\mathbf{p})=G(\mathbf{p}(q_1,q_2,\dots q_k))$ then we would re-write the above expression as

$$\int d^{(n-k)}q \int d^k q|\mathcal{J}|G(\mathbf{q})\delta(\mathbf{q}-\mathbf{p}) = G(\mathbf{p}) \tag{5.74}$$

and we read off that the Dirac delta function has transformed as

$$\delta(\mathbf{x}-\mathbf{y}) \to \frac{\delta(\mathbf{q}-\mathbf{p})}{\int d^{(n-k)}q|\mathcal{J}|}. \tag{5.75}$$

By way of example consider the transformation in two-dimensions from Cartesian coordinates $(x,y)$ to polar coordinates $(r,\theta)$ given by $x=r\cos\theta$, $y=r\sin\theta$. The Jacobian is

$$\mathcal{J} = \begin{pmatrix} \frac{\partial x}{\partial r} & \frac{\partial x}{\partial \theta} \\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial \theta} \end{pmatrix} = \begin{pmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{pmatrix} \tag{5.76}$$

and so $|\mathcal{J}|=r$. Hence at the origin $r=0$ and the Jacobian is singular. In this example we have $n=2$ and $k=1$

$$\int d\theta \int dr|\mathcal{J}|G(\mathbf{r})\delta(\mathbf{r}-\mathbf{0}) = 2\pi \int drrG(\mathbf{r})\delta(r) = G(\mathbf{0}) \tag{5.77}$$

and we see that the two-dimensional Dirac delta function is transformed as

$$\delta(\mathbf{x}) = \delta(x)\delta(y) \to \frac{\delta(r)}{2\pi r}. \tag{5.78}$$

Let us consider the transformation to generalised spherical coordinates valid for any $n \geq 2$ given by

$$\begin{aligned} x_1 &= r\cos\theta_1\cos\theta_2\dots\cos\theta_{n-1} \\ x_2 &= r\cos\theta_1\cos\theta_2\dots\cos\theta_{n-2}\sin\theta_{n-1} \\ &\vdots \\ x_k &= r\cos\theta_1\cos\theta_2\dots\cos\theta_{n-k}\sin\theta_{n-k+1} \\ &\vdots \\ x_n &= r\sin\theta_1 \end{aligned} \tag{5.79}$$

so that

$$|\mathcal{J}| = r^{n-1}(\cos\theta_1)^{n-2}(\cos\theta_2)^{n-3}\dots(\cos\theta_k)^{n-k-1}\dots\cos\theta_{n-2} \tag{5.80}$$

and

$$d^n x = |\mathcal{J}| dr d\theta_1 d\theta_2 \dots d\theta_{n-1} \equiv r^{n-1} dr d\Omega_n \tag{5.81}$$

where

$$d\Omega_n \equiv \prod_{i=1}^{n-1} (\cos\theta_i)^{n-i-1} d\theta_i. \tag{5.82}$$

Let

$$\frac{1}{2}\Omega_n \equiv \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \dots \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} d\Omega_n \tag{5.83}$$

where the factor of a half compensates for the range of the integration covering only half of the full (hyper)spherical surface. We may evaluate the integral $\Omega_n$ using a recursion relation:

$$\begin{aligned}
I_n &\equiv \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} (\cos\theta)^n d\theta \qquad (5.84)\\
&= \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} (\cos\theta)^{n-1} (\cos\theta) d\theta \\
&= \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} (\cos\theta)^{n-1} \frac{d}{d\theta}\Big(\sin\theta\Big) d\theta \\
&= -\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \frac{d}{d\theta}\Big((\cos\theta)^{n-1}\Big)(\sin\theta) d\theta + \underbrace{\Big[\sin\theta(\cos\theta)^{n-1}\Big]_{-\frac{\pi}{2}}^{\frac{\pi}{2}}}_{=0} \\
&= \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} (n-1)(\cos\theta)^{n-2}(\sin\theta)^2 d\theta \\
&= (n-1)\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \Big((\cos\theta)^{n-2} - (\cos\theta)^n\Big) d\theta \\
&= (n-1)I_{n-2} - (n-1)I_n \\
\Rightarrow I_n &= \frac{n-1}{n} I_{n-2} \qquad (5.85)
\end{aligned}$$

Applying the recursion relation $k$ times gives

$$I_n = \frac{(n-1)(n-3)\dots(n+1-2k)}{n(n-2)\dots(n-2k+2)} I_{n-2k}. \tag{5.86}$$

One can stop applying the recursion relation when the integral becomes simple to evaluate. Two simple integrals are $I_0 = \pi$ and $I_1 = 2$. Hence there are two cases we must consider, the first when $n$ is even and the second when $n$ is odd. When $n$ is even we may apply the recursion relation $k = \frac{n}{2}$ times to find

$$\begin{aligned}
I_n &= \frac{(n-1)(n-3)\dots(1)}{n(n-2)\dots(2)} I_0 \qquad (5.87)\\
&= \frac{2^{\frac{n}{2}}(\frac{n}{2}-\frac{1}{2})(\frac{n}{2}-\frac{3}{2})\dots\frac{1}{2}}{2^{\frac{n}{2}}\frac{n}{2}(\frac{n}{2}-1)\dots(1)} I_0 \\
&= \frac{\Gamma(\frac{n}{2}+\frac{1}{2})}{\Gamma(\frac{n}{2}+1)\Gamma(\frac{1}{2})}\pi \\
&= \frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n+2}{2})}\sqrt{\pi}
\end{aligned}$$

where in the third line we have used the gamma function which is defined by

$$\Gamma(z) \equiv \int_0^\infty t^{(z-1)} e^{-t} dt \qquad \text{where} \quad z \in \mathbb{C}. \tag{5.88}$$

The integral for the gamma function converges only for $z$ whose real part is positive. The gamma function satisfies the following relation

$$\Gamma(z+1) = z\Gamma(z). \tag{5.89}$$

Repeated use of this relation for any positive real integer $n$ gives

$$\Gamma(n+1) = n(n-1)(n-2)\ldots 3.2.\Gamma(1) = n! \qquad \text{as} \quad \Gamma(1) = 1. \tag{5.90}$$

As the gamma function is well-defined when its argument is any positive numbers, not just an integer, it may be considered an extension of the factorial function. For example consider the following gamma function which appeared in our earlier computation where $n$ is an even positive integer (hence the argument $\frac{n}{2} + \frac{1}{2}$ is half-integer)

$$\begin{aligned} \Gamma(\frac{n}{2}+\frac{1}{2}) &= (\frac{n-1}{2})\Gamma(\frac{n-1}{2}) \\ &= (\frac{n-1}{2})(\frac{n-3}{2})\Gamma(\frac{n-3}{2}) \\ &\vdots \\ &= (\frac{n-1}{2})(\frac{n-3}{2})(\frac{n-5}{2})\ldots\frac{5}{2}.\frac{3}{2}.\frac{1}{2}\Gamma(\frac{1}{2}) \\ &= (\frac{n-1}{2})(\frac{n-3}{2})(\frac{n-5}{2})\ldots\frac{5}{2}.\frac{3}{2}.\frac{1}{2}\sqrt{\pi} \end{aligned} \tag{5.91}$$

where we have used the observation that $\Gamma(\frac{1}{2}) = \sqrt{\pi}$ which is seen by direct computation using equation (5.88). The derivation of the integral $I_n$, where $n$ is an even integer, in equation (5.87) above should now be clear. It remains to find an expression for $I_n$ when $n$ is an odd integer in terms of gamma functions - in fact we will find exactly the same expression as for the case when $n$ is even. Commencing with equation (5.86) but now taking odd $n$ and $k = \frac{n-1}{2}$ (an integer) we have

$$\begin{aligned} I_n &= \frac{(n-1)(n-3)\ldots(2)}{n(n-2)\ldots(3)} I_1 \\ &= \frac{2^{\frac{n-1}{2}}(\frac{n}{2}-\frac{1}{2})(\frac{n}{2}-\frac{3}{2})\ldots\frac{1}{2}}{2^{\frac{n+1}{2}}\frac{n}{2}(\frac{n}{2}-1)\ldots(1)} I_1 \\ &= \frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n+2}{2})}\sqrt{\pi}. \end{aligned} \tag{5.92}$$

Now returning to the integral $\Omega_n$ defined in equation (5.83) we have

$$\begin{aligned} \Omega_n &= 2I_{n-2}.I_{n-3}\ldots I_1.I_0 \\ &= 2\pi^{(\frac{n-1}{2})}\frac{\Gamma(\frac{n-1}{2})\Gamma(\frac{n-2}{2})\Gamma(\frac{n-3}{2})\ldots\Gamma(1)\Gamma(\frac{1}{2})}{\Gamma(\frac{n}{2})\Gamma(\frac{n-1}{2})\Gamma(\frac{n-2}{2})\ldots\Gamma(\frac{3}{2})\Gamma(1)} \\ &= \frac{2\pi^{\frac{n}{2}}}{\Gamma(\frac{n}{2})}. \end{aligned} \tag{5.93}$$

We have found the surface area of an $(n-1)$-sphere of unit radius. Notice that when $n=2$ we have $\Omega_2 = 2\pi$ which is the circumference of an $S^1$ of unit radius, when $n=3$ we have $\Omega_3 = 4\pi$ which is the surface area of a unit $S^2$ et cetera. Finally we can draw together all our observations to find how the $n$-dimensional Dirac delta function $\delta(\mathbf{x})$ is transformed when we work in generalised spherical coordinates. From equations (5.75) and (5.81) we have

$$\begin{aligned}\delta(\mathbf{x}) \to & \frac{\delta(r)}{\int d\theta_1 d\theta_2 \dots d\theta_{(n-1)} |\mathcal{J}|} \\ &= \frac{\delta(r)}{r^{n-1}\Omega_n} \\ &= \frac{\delta(r)\Gamma(\frac{n}{2})}{2\pi^{\frac{n}{2}} r^{n-1}}.\end{aligned} \tag{5.94}$$

It is easy to confirm that this gives the expected expression when $n=2$.

### 5.3.2 The Green's function for the Laplacian operator

Recall that our aim is to find a function $G$ such that

$$\nabla^2 G = \frac{\delta(r)}{r^{(n-1)}\Omega_n}. \tag{5.95}$$

Consider a general function $F \equiv F(r)$ of the radial coordinate $r$. The Laplacian acts on $F$ to give

$$\begin{aligned}\nabla^2 F(r) &\equiv \sum_{i=1}^{n} \partial_i \partial_i (F(r)) \\ &= \sum_{i=1}^{n} \partial_i \left(\frac{\partial F}{\partial r}\frac{\partial r}{\partial x^i}\right) \\ &= \sum_{i=1}^{n} \partial_i \left(\frac{\partial F}{\partial r}\frac{x^i}{r}\right) \\ &= \sum_{i=1}^{n} \left(\frac{\partial^2 F}{\partial r^2}\left(\frac{x^i}{r}\right)^2 + \frac{\partial F}{\partial r}\frac{1}{r} - \frac{\partial F}{\partial r}\frac{(x^i)^2}{r^3}\right) \\ &= \frac{\partial^2 F}{\partial r^2} + n\frac{\partial F}{\partial r}\frac{1}{r} - \frac{\partial F}{\partial r}\frac{1}{r} \\ &= \frac{\partial^2 F}{\partial r^2} + \frac{(n-1)}{r}\frac{\partial F}{\partial r} \\ &= \frac{1}{r^{(n-1)}}\frac{\partial}{\partial r}\left(r^{(n-1)}\frac{\partial F}{\partial r}\right).\end{aligned} \tag{5.96}$$

As the Green's function is a function of only the radial coordinate we find we are searching for a function satisfying

$$\nabla^2 G(r) = \frac{1}{r^{(n-1)}}\frac{\partial}{\partial r}\left(r^{(n-1)}\frac{\partial G}{\partial r}\right) = \frac{\delta(r)}{r^{(n-1)}\Omega_n}. \tag{5.97}$$

Hence we have

$$\frac{\partial}{\partial r}\left(r^{(n-1)}\frac{\partial G}{\partial r}\right) = \frac{\delta(r)}{\Omega_n}. \tag{5.98}$$

When $r \neq 0$ we are left with

$$\frac{\partial}{\partial r}\left(r^{(n-1)}\frac{\partial G}{\partial r}\right) = 0 \tag{5.99}$$
$$\Rightarrow \frac{\partial G}{\partial r} = Cr^{(1-n)} \qquad \text{where } C \text{ is a constant}$$
$$\Rightarrow G = \frac{C}{(2-n)r^{(n-2)}} \qquad \text{for } n > 2.$$

We can fix the constant $C$ by using Stokes' theorem

$$\begin{aligned}\int_V (\nabla^2 G)dV &= \int_V \frac{\delta(r)}{r^{(n-1)}\Omega_n}dV = 1 \\ &= \int_S \nabla G dS \\ &= \int_S \frac{\partial G}{\partial r}dS \\ &= Cr^{(1-n)}\Omega_n r^{(n-1)}\end{aligned} \tag{5.100}$$

where $V$ is the volume integrated over, $S$ is the boundary (or surface) of the volume of integration and in the second line we have made use of Stokes' theorem to relate the volume integral to the surface integral. Hence we find that

$$C = \frac{1}{\Omega_n} \tag{5.101}$$

and the Green's function for the Laplacian in dimensions greater than two[2] is

$$G = \frac{1}{(2-n)r^{(n-2)}\Omega_n} = \frac{\Gamma(\frac{n}{2})}{2(2-n)\pi^{\frac{n}{2}}r^{(n-2)}} \qquad \text{for} \quad n > 2. \tag{5.102}$$

In two dimensions we still need to solve

$$\frac{\partial}{\partial r}\left(r\frac{\partial G}{\partial r}\right) = \frac{\delta(r)}{\Omega_2} = \frac{\delta(r)}{2\pi} \tag{5.103}$$

and we note that $G = C\ln(r)$ where $C$ is a constant gives a solution to the equation when $r \neq 0$:

$$\frac{\partial}{\partial r}\left(r\frac{\partial G}{\partial r}\right) = \frac{\partial}{\partial r}\left(Cr\frac{1}{r}\right) = 0. \tag{5.104}$$

We may evaluate the constant $C$ using Stokes' theorem (in fact this was the setting for the appearance of Greens theorem in the plane the precursor to Stokes more general theorem). We find that

$$C = \frac{1}{2\pi} \tag{5.105}$$

and in two dimensions Green's function for the Laplacian is

$$G = \frac{1}{2\pi}\ln(r). \tag{5.106}$$

Now we may take advantage of the Green's function to solve the Poisson equation $\nabla^2 f = -\rho(\mathbf{x})$. If we consider the case when the space has dimension greater than two[3]

[2] We assumed in the derivation that $n > 2$ - we will return to the two-dimensional Green's function momentarily.

[3] We could equally well work in two dimensions but we would use the two-dimensional Green's function for the Laplacian in that case.

immediately we may write

$$f(\mathbf{x}) = -\int d^n y G(\mathbf{x},\mathbf{y})\rho(\mathbf{y}) = -\frac{\Gamma(\frac{n}{2})}{2(2-n)\pi^{\frac{n}{2}}}\int d^n y \frac{\rho(\mathbf{y})}{|\mathbf{x}-\mathbf{y}|^{(n-2)}}. \tag{5.107}$$

In particular when $n = 3$ we have

$$\begin{aligned} f(\mathbf{x}) &= \frac{\Gamma(\frac{3}{2})}{2\pi^{\frac{3}{2}}}\int d^3 y \frac{\rho(\mathbf{y})}{|\mathbf{x}-\mathbf{y}|} \\ &= \frac{\frac{1}{2}\Gamma(\frac{1}{2})}{2\pi^{\frac{3}{2}}}\int d^3 y \frac{\rho(\mathbf{y})}{|\mathbf{x}-\mathbf{y}|} \\ &= \frac{1}{4\pi}\int d^3 y \frac{\rho(\mathbf{y})}{|\mathbf{x}-\mathbf{y}|}. \end{aligned} \tag{5.108}$$

This is the electrostatic potential due to a charge density $\rho(\mathbf{y})$. Another example of interest is the potential due to a point mass located at the origin in a scalar field theory of gravity where

$$\nabla^2\phi = \rho(r) = 4\pi M G_N \delta(r) \tag{5.109}$$

where $\phi$ is a scalar gravitational potential, $\rho(r)$ indicates a generic spherical mass distribution and $G_N$ is Newton's gravitational constant. At first sight this equation looks unfamiliar, in particular we may be concerned by the factor of $4\pi$. Let us check however that this choice of normalisation for the mass distribution gives precisely what one would expect from Newtonian gravity. Using the Green's function and specialising to three dimensions we have

$$\begin{aligned} \phi &= \int d\mathbf{r}'\rho(\mathbf{r}')G(\mathbf{r},\mathbf{r}') \\ &= \int d\mathbf{r}' 4\pi M G_N \delta(r')\left(\frac{-1}{4\pi|\mathbf{r}-\mathbf{r}'|}\right) \\ &= -\frac{G_N M}{r}. \end{aligned} \tag{5.110}$$

Hence the field strength associated to this potential is

$$F \equiv -\nabla(\phi) = -\frac{\partial}{\partial r}\left(\frac{-G_N M}{r}\right) = \frac{-G_N M}{r^2} \tag{5.111}$$

and so this mass distribution $\rho(\mathbf{r}) = 4\pi G_N M \delta(r)$ gives the expected Newtonian force for gravity.

## 5.4 Some Complex Calculus

### 5.4.1 Differentiation of Complex Functions

Let $f : \mathbb{C} \to \mathbb{C}$ act as $f(z) = w$ where $z = x + iy$ and $w = u + iv$, that is

$$f(x+iy) = u(x,y) + iv(x,y). \tag{5.112}$$

**Definition** The derivative of a complex function $f(z)$ at $z_0$ is

$$\left.\frac{df}{dz}\right|_{z_0} = \lim_{\Delta z \to 0}\left[\frac{f(z_0+\Delta z) - f(z_0)}{\Delta z}\right]$$

provided that the limit exists and is independent of $\Delta z$.

This is a rather restrictive condition on $f$, or so it would seem. It has an analogue in real calculus: consider the derivative of $|x|$ at $x = 0$, the value of the derivative depends upon the direction one approaches $x = 0$ whether from the positive or the negative directions. Such a derivative is not a normal function and typically one would say it does not exist. On the complex plane there are two independent (and canonical) directions along which one may approach $z_0$. We may choose the two independent directions to be parallel to the real and imaginary axes without loss of generality; for the choice $z = x + iy$ the two directions are along $x$ or along $iy$. For the derivative to exist these two choices of $\Delta z$ must give the same value:

$$\begin{aligned} \left.\frac{\partial f}{\partial z}\right|_{z_0} &= \frac{\partial u}{\partial x} + i\frac{\partial v}{\partial x} \\ &= \frac{\partial u}{i\partial y} + i\frac{\partial v}{i\partial y} = \frac{\partial v}{\partial y} - i\frac{\partial u}{\partial y} \end{aligned} \tag{5.113}$$

where in the first line we have chosen to take the limit $\Delta z = x$ and in the second line we have taken the limit $\Delta z = iy$. If the complex derivative of $f(z)$ exists these two expressions must be equal. Hence we have the existence condition for the complex derivative as a pair of differential equations

$$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y} \qquad \text{and} \qquad \frac{\partial v}{\partial x} = -\frac{\partial u}{\partial y}. \tag{5.114}$$

These are the Cauchy-Riemann equations and they guarantee the differentiability of $f = u + iv$ at $z_0$.

If we had written

$$x = \frac{1}{2}(z + z^*) \qquad \text{and} \qquad y = -\frac{i}{2}(z - z^*) \tag{5.115}$$

where $z^* = x - iy$ is the complex conjugate of $z$ and treated $z$ and $z^*$ as our two coordinates instead of $x$ and $y$ we can find a very simple but powerful condition on differentiable complex functions. First we note that

$$\begin{aligned} \frac{\partial}{\partial z} &= \frac{\partial x}{\partial z}\frac{\partial}{\partial x} + \frac{\partial y}{\partial z}\frac{\partial}{\partial y} = \frac{1}{2}\left(\frac{\partial}{\partial x} - i\frac{\partial}{\partial y}\right) \qquad \text{and} \\ \frac{\partial}{\partial z^*} &= \frac{\partial x}{\partial z^*}\frac{\partial}{\partial x} + \frac{\partial y}{\partial z^*}\frac{\partial}{\partial y} = \frac{1}{2}\left(\frac{\partial}{\partial x} + i\frac{\partial}{\partial y}\right). \end{aligned} \tag{5.116}$$

Hence for a differentiable complex function $f$

$$\begin{aligned} \frac{\partial f}{\partial z^*} &= \frac{1}{2}\left(\frac{\partial u}{\partial x} + i\frac{\partial u}{\partial y}\right) + i\frac{1}{2}\left(\frac{\partial v}{\partial x} + i\frac{\partial v}{\partial y}\right) \\ &= \frac{1}{2}\left(\frac{\partial u}{\partial x} - \frac{\partial v}{\partial y}\right) + \frac{i}{2}\left(\frac{\partial v}{\partial x} + \frac{\partial u}{\partial y}\right) \\ &= 0 \end{aligned} \tag{5.117}$$

where we have used the Cauchy-Riemann equations in the final line. If $\frac{\partial f}{\partial z^*} = 0$ $f(z)$ is called an holomorphic function, on the other hand if $\frac{\partial f}{\partial z} = 0$ $f(z^*)$ is called an antiholomorphic function. Consequently all holomorphic functions are differentiable functions.

**Definition** A function $f : \mathbb{C} \to \mathbb{C}$ is called analytic at $z_0$ if it is differentiable at $z_0$ and all points in a neighbourhood of $z_0$.

A couple of comments are worth making at this point:

- A point at which $f(z)$ is not analytic is called a singular point of $f$, or a singularity.
- The real and imaginary parts of an analytic function are harmonic as from differentiating the Cauchy-Riemann equations we have

$$\begin{aligned}\frac{\partial^2 u}{\partial x^2} &= \frac{\partial^2 v}{\partial x \partial y} \\ &= \frac{\partial}{\partial y}\left(\frac{\partial v}{\partial x}\right) \\ &= -\frac{\partial^2 u}{\partial y^2} \\ &\Rightarrow \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = \nabla^2 u = 0\end{aligned} \tag{5.118}$$

  and similarly $\nabla^2 v = 0$.

### 5.4.2 Integration of Complex Functions

Consider the integral

$$I \equiv \int_{z_1}^{z_2} f(z) dz. \tag{5.119}$$

As this is an integral over a two-dimensional space the path between the two points should be specified to give a single meaning to the integral. However, although it is not obvious at this point, if $f(z)$ is an analytic function then the integral $I$ is path-independent. As before let us write $f(z) = u + iv$ and taking $z = x + iy$ we have $dz = dx + idy$ so that

$$I = \int_{z_1}^{z_2} (u + iv)(dx + idy) = \int_{z_1}^{z_2} (udx + iudy + ivdx - vdy). \tag{5.120}$$

Let us define two vectors

$$\mathbf{A}_1 \equiv \begin{pmatrix} u \\ -v \end{pmatrix} \qquad \text{and} \qquad \mathbf{A}_2 \equiv \begin{pmatrix} v \\ u \end{pmatrix}$$

so that

$$I = \int_{z_1}^{z_2} (\mathbf{A}_1 \cdot d\mathbf{r} + i\mathbf{A}_2 \cdot d\mathbf{r}). \tag{5.121}$$

Previously in these lecture notes we considered similar line integrals in classical mechanics when we developed the path-dependent work function. To make that integral path-independent we wrote $\mathbf{F} = -\nabla V$ where $V$ was a scalar function. Similarly if we can write $\mathbf{A}_1 = -\nabla\phi_1$ and $\mathbf{A}_2 = -\nabla\phi_2$ then the integral $I$ will be path-independent. Now if $\mathbf{A} = -\nabla\phi$ then $\nabla \wedge \mathbf{A} = \mathbf{0}$, so let us check if this is the case for our vectors $\mathbf{A}_1$ and $\mathbf{A}_2$:

$$\begin{aligned}\nabla \wedge \mathbf{A}_1 &= \begin{pmatrix} \frac{\partial}{\partial x} \\ \frac{\partial}{\partial y} \\ \frac{\partial}{\partial z} \end{pmatrix} \wedge \begin{pmatrix} u \\ -v \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ -\frac{\partial v}{\partial x} - \frac{\partial u}{\partial y} \end{pmatrix} = \mathbf{0} \\ \nabla \wedge \mathbf{A}_2 &= \begin{pmatrix} \frac{\partial}{\partial x} \\ \frac{\partial}{\partial y} \\ \frac{\partial}{\partial z} \end{pmatrix} \wedge \begin{pmatrix} v \\ u \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \frac{\partial u}{\partial x} - \frac{\partial v}{\partial y} \end{pmatrix} = \mathbf{0}\end{aligned} \tag{5.122}$$

where we have observed the vectors are zero as $f(z)$ is an analytic function and so the Cauchy-Riemann equations are valid and their application gives the zero vector. To summarise if $f(z)$ is an analytic function its integral $I$ is well-defined as it is path-independent.

One trivial consequence of path-independence of the integral is

$$\oint_C f(z)dz = 0 \tag{5.123}$$

where $C$ indicates some closed contour, hence the integral has the same beginning and end points and hence the integral vanishes.

When integrating around a contour it is conventional to traverse the boundary in such a way as to keep the bounded region to the left. This is called integration in the positive sense, integration in the opposite direction acquires a minus sign.

**Theorem 5.4.1.** *(Cauchy integral formula) Let $f$ be analytic on and within a simple closed contour $C$ (integrated in the positive sense). Let $z_0$ be any interior point to $C$ then,*

$$f(z_0) = \frac{1}{2\pi i}\oint_C \frac{f(z)}{(z-z_0)}dz. \tag{5.124}$$

To prove this we will need the following lemma

**Lemma 5.4.2.** *(Darboux inequality) Let $f : \mathbb{C} \to \mathbb{C}$ be a continuous and bounded function on a path $\gamma$ then if $|f(\gamma)| \leq M$*

$$|\int_\gamma f(z)dz| \leq ML_\gamma \tag{5.125}$$

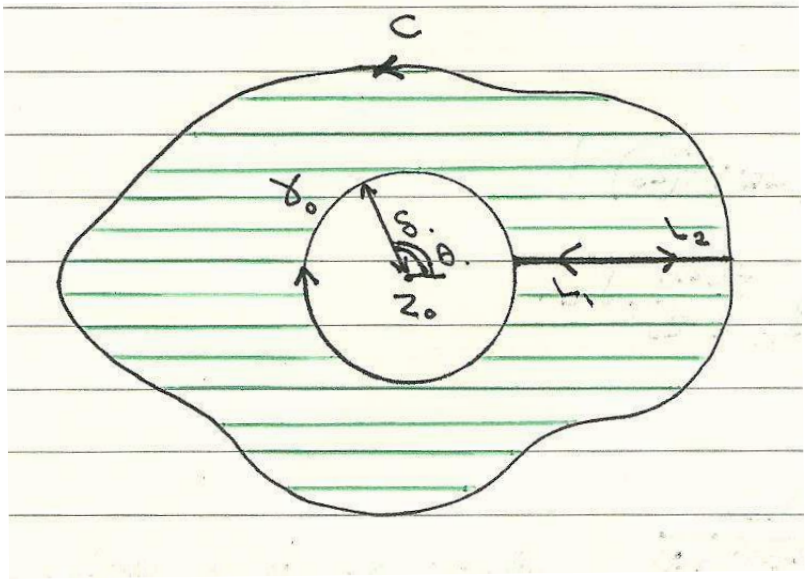
*where $L_\gamma$ is the length of $\gamma$.*

*Proof.* (Darboux inequality)

$$\begin{aligned}
|\int_\gamma f(z)dz| &= \left|\lim_{n\to\infty,\Delta z_i\to 0}\sum_{i=1}^{n} f(z_i)\Delta z_i\right| \qquad (5.126)\\
&= \lim_{n\to\infty,\Delta z_i\to 0}\left|\sum_{i=1}^{n} f(z_i)\Delta z_i\right| \\
&\leq \lim_{n\to\infty,\Delta z_i\to 0}\sum_{i=1}^{n}\left|f(z_i)\Delta z_i\right| \qquad \text{by the triangle inequality} \\
&= \lim_{n\to\infty,\Delta z_i\to 0}\sum_{i=1}^{n}\left|f(z_i)\right|\left|\Delta z_i\right| \\
&\leq M\lim_{n\to\infty,\Delta z_i\to 0}\sum_{i=1}^{n}\left|\Delta z_i\right| \\
&= ML_\gamma
\end{aligned}$$

□

*Proof.* (Cauchy integral formula) Consider integrating around a boundary enclosing a region with a neighbourhood of a point $z_0$ removed, e.g.




Let the contour around the enclosed region be $C' \equiv L_1 \cup \gamma_0 \cup L_2 \cup C$, with the direction of integration indicated by arrowheads on the diagram. Now $\frac{f(z)}{z-z_0}$ is analytic on the contour $C'$ so as $C'$ is a closed path then

$$\frac{1}{2\pi i}\oint_{C'}\left(\frac{f(z)}{z-z_0}\right)dz = 0 \tag{5.127}$$

where we have inserted the factor $\frac{1}{2\pi i}$ for convenience later. If we separate out the integral into the integrals over the four paths $C$, $L_1$, $L_2$ and $\gamma_0$ we have:

$$\begin{aligned} 0 &= \frac{1}{2\pi i}\left[\oint_C \frac{f(z)}{z-z_0}dz + \oint_{\gamma_0}\frac{f(z)}{z-z_0}dz + \int_{L_1}\frac{f(z)}{z-z_0}dz + \int_{L_2=-L_1}\frac{f(z)}{z-z_0}dz\right] \\ &= \frac{1}{2\pi i}\left[\oint_C \frac{f(z)}{z-z_0}dz + \oint_{\gamma_0}\frac{f(z)}{z-z_0}dz\right]. \end{aligned} \tag{5.128}$$

As $f(z)$ is continuous then there exists an $\epsilon$ such that $|f(z) - f(\epsilon)| < \epsilon$. Hence on $\gamma_0$ we have

$$\left|\frac{f(z)-f(z_0)}{z-z_0}\right| = \frac{|f(z)-f(z_0)|}{|z-z_0|} \leq \frac{\epsilon}{\delta} \tag{5.129}$$

From the lemma we have

$$\left|\oint_{\gamma_0}\frac{f(z)-f(z_0)}{z-z_0}\right| \leq \frac{\epsilon}{\delta}(2\pi\delta) = 2\pi\epsilon. \tag{5.130}$$

Taking the limit of $\delta \to 0$ then $\epsilon \to 0$ and rearranging the above inequality gives

$$\oint_{\gamma_0}\frac{f(z)}{z-z_0}dz = \oint_{\gamma_0}\frac{f(z_0)}{z-z_0}dz = f(z_0)\oint_{\gamma_0}\frac{dz}{z-z_0}. \tag{5.131}$$

On $\gamma_0$ $z = z_0 + \delta e^{i\theta}$ hence $dz = i\delta e^{i\theta}d\theta$ and $z - z_0 = \delta e^{i\theta}$, so that,

$$f(z_0)\oint_{\gamma_0}\frac{dz}{z-z_0} = f(z_0)\left(-\int_0^{2\pi}\frac{i\delta e^{i\theta}d\theta}{\delta e^{i\theta}}\right) = f(z_0)(-2\pi i) \tag{5.132}$$

where the minus sign has appeared as the path $\gamma_0$ is in the negative sense with respect to the interior region bounded by $\gamma_0$. We have now from equation (5.128)

$$0 = \frac{1}{2\pi i}\left[\oint_C \frac{f(z)}{z-z_0}dz - 2\pi i f(z_0)\right]. \tag{5.133}$$

After rearranging we have

$$f(z_0) = \frac{1}{2\pi i}\left[\oint_C \frac{f(z)}{z-z_0}dz\right] \tag{5.134}$$

as required. □

**Example.**

Evaluate

$$I_1 \equiv \oint_{C_1} \frac{z^2 dz}{(z^2+3)^2(z-i)}$$

where $C_1$ is a circle centred at the origin of radius $\frac{3}{2}$.

As $f(z) \equiv \frac{z^2}{(z^2+3)^2}$ is analytic on $C_1$ and in the region bounded by $C_1$, while $z = i$ is interior to $C_1$ we find

$$I_1 \equiv \oint_{C_1} \frac{f(z)dz}{z-i} = 2\pi i(f(i)) = 2\pi i(\frac{-1}{4}) = -\frac{i\pi}{2}.$$

The Cauchy integral formula gives a value to points which are bounded by a closed curve along which the function is analytic. The value of an analytic function on a boundary determines the function at all points inside the boundary. One can use the Cauchy integral formula to evaluate the derivatives of analytic functions and observe that they are all also analytic on the same domain of analyticity.

**Theorem 5.4.3.** *The derivatives of an analytic function $f(z)$ exist to all orders in the domain of analyticity of the function and are themselves analytic. The n'th derivative of $f(z)$ is*

$$\frac{d^n f}{dz^n} = \frac{n!}{2\pi i}\oint_C \frac{f(\xi)d\xi}{(\xi - z)^{(n+1)}}$$

*where $C$ is a closed curve bounding a region containing $z$.*

Note that the derivatives are analytic at all points $z \notin C$.

### 5.4.3 Laurent Series

Analytic functions may be written as a Taylor series. The expansion about a point $z_0$ converges on a circle $C_0$ of radius $r_0$ centred at $z_0$ and takes the expected form:

$$f(z_0 + \delta z) = f(z_0) + \delta z \left.\frac{df}{dz}\right|_{z_0} + \frac{1}{2!}(\delta z)^2 \left.\frac{d^2 f}{dz^2}\right|_{z_0} + \ldots \tag{5.135}$$

and converges when $\delta z \equiv |z - z_0| < r_0$.

*Proof.*

$$f(z) = \frac{1}{2\pi i}\oint_{C_0} \frac{f(\xi)d\xi}{\xi - z} \tag{5.136}$$

Now we may re-write

$$\begin{aligned}\frac{1}{\xi - z} &= \frac{1}{\xi - z_0 + z_0 - z} \\ &= \frac{1}{(\xi - z_0)(1 - \frac{z - z_0}{\xi - z_0})} \\ &= \frac{1}{(\xi - z_0)} \sum_{n=0}^{\infty} \left( \frac{z - z_0}{\xi - z_0} \right)^n\end{aligned} \tag{5.137}$$

where we have used the observation that

$$S \equiv \sum_{n=0}^{\infty} x^n \qquad \Rightarrow \quad S = \frac{1}{1-x} \qquad \text{when} \quad |x| < 1 \tag{5.138}$$

for the case when $x = \frac{z - z_0}{\xi - z_0}$ it is guaranteed that $|x| < 1$ as $z$ lies within the circle of radius $r_0$ and $\xi$ lies on the boundary of the circle, hence $|z - z_0| < |\xi - z_0| = r_0$. Therefore we have

$$f(z) = \frac{1}{2\pi i} \oint_{C_0} \frac{f(\xi) d\xi}{(\xi - z_0)} \sum_{n=0}^{\infty} \left( \frac{z - z_0}{\xi - z_0} \right)^n. \tag{5.139}$$

We recall that

$$\frac{d^n f}{dz^n} = \frac{n!}{2\pi i} \oint_C \frac{f(\xi) d\xi}{(\xi - z)^{(n+1)}} \tag{5.140}$$

whose value at $z = z_0$ we may substitute, after some rearranging of terms, into equation (5.139):

$$\begin{aligned}f(z) &= \frac{1}{2\pi i} \sum_{n=0}^{\infty} (z - z_0)^n \oint_{C_0} \frac{f(\xi) d\xi}{(\xi - z_0)^{n+1}} \\ &= \sum_{n=0}^{\infty} \frac{(z - z_0)^n}{n!} \frac{d^n f}{dz^n}\bigg|_{z=z_0} \\ f(z_0 + \delta z) &= \sum_{n=0}^{\infty} \frac{(\delta z)^n}{n!} \frac{d^n f}{dz^n}\bigg|_{z=z_0}\end{aligned} \tag{5.141}$$

where $\delta z = z - z_0$. □

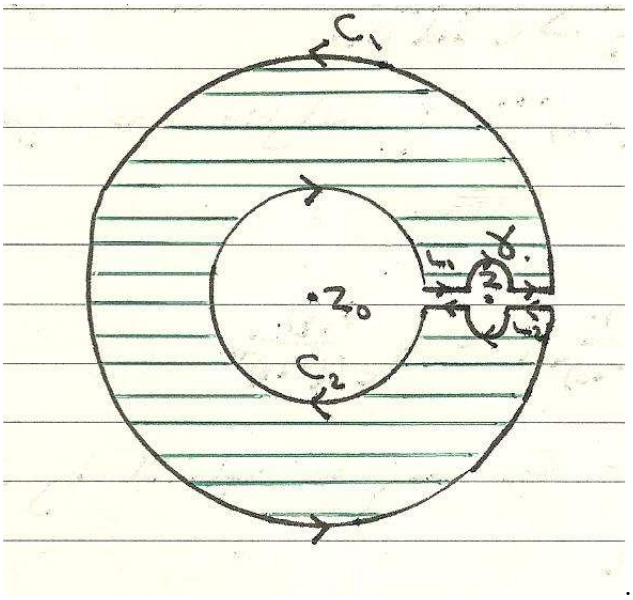
If $f(z)$ is not analytic at all points inside the circle we can no longer use the Taylor series instead we can expand the function as a Laurent series. This is the extension of the Taylor series to include negative powers of $\delta z$. When we cut out a hole around the singular point, and the Taylor expansion around the boundary of the hole will give negative powers of the expansion parameter. The Laurent series will converge on an annulus formed by puncturing a circle.

**Theorem 5.4.4.** *(Laurent series) Let $C_1$ and $C_2$ be circles of radii $r_1$ and $r_2$ both centred at $z_0$ with $r_1 > r_2$. Let $f : \mathbb{C} \to \mathbb{C}$ be analytic on $C_1$ and $C_2$ and throughout the annulus $S$ between the two circles. Then at each point $z \in S$ $f(z)$ is*

$$f(z) = \sum_{n=-\infty}^{\infty} a_n (z - z_0)^n \qquad \textit{where} \quad a_n = \frac{1}{2\pi i} \oint_C \frac{f(\xi) d\xi}{(\xi - z_0)^{(n+1)}} \tag{5.142}$$

*where $C$ is any closed contour within $S$ that surrounds $z_0$.*

*Proof.* Let $\gamma$ be a small closed contour in $S$ enclosing a point $z \in S$ as shown below




.

Let us denote the boundary of the shaded region by $C' \equiv C_1 \cup \pm L_1 \cup \gamma \cup C_2 \cup \pm L_2$. We have

$$\begin{aligned} 0 &= \oint_{C'} \frac{f(\xi)d\xi}{\xi - z} \\ &= \oint_{C_1} \frac{f(\xi)d\xi}{\xi - z} - \oint_{\gamma} \frac{f(\xi)d\xi}{\xi - z} - \oint_{C_2} \frac{f(\xi)d\xi}{\xi - z} \\ &= \oint_{C_1} \frac{f(\xi)d\xi}{\xi - z} - 2\pi i f(z) - \oint_{C_2} \frac{f(\xi)d\xi}{\xi - z} \end{aligned} \tag{5.143}$$

that is,

$$2\pi i f(z) = \oint_{C_1} \frac{f(\xi)d\xi}{\xi - z} - \oint_{C_2} \frac{f(\xi)d\xi}{\xi - z}. \tag{5.144}$$

Now for the first contour integral $\xi \in C_1$ and $z \in S$ so that $|z - z_0| < |\xi - z_0|$ (as was the case for the Taylor series expansion) so as we saw earlier

$$\frac{1}{\xi - z} = \frac{1}{(\xi - z_0)} \sum_{n=0}^{\infty} \left( \frac{z - z_0}{\xi - z_0} \right)^n. \tag{5.145}$$

While for the second contour integral $\xi \in C_2$ so that $|z - z_0| > |\xi - z_0|$ and now the relevant small parameter is $\left| \frac{\xi - z_0}{z - z_0} \right| < 1$ and so

$$\begin{aligned} \frac{1}{\xi - z} &= \frac{1}{\xi - z_0 + z_0 - z} \\ &= \frac{1}{(z - z_0)(-1 + \frac{\xi - z_0}{z - z_0})} \\ &= -\frac{1}{(z - z_0)} \sum_{n=0}^{\infty} \left( \frac{\xi - z_0}{z - z_0} \right)^n. \end{aligned} \tag{5.146}$$

Hence

$$2\pi i f(z) = \sum_{n=0}^{\infty} (z - z_0)^n \oint_{C_1} \frac{f(\xi)d\xi}{(\xi - z_0)^{(n+1)}} + \sum_{n=0}^{\infty} \frac{1}{(z - z_0)^{(n+1)}} \oint_{C_2} (\xi - z_0)^n f(\xi)d\xi. \tag{5.147}$$

Next consider an arbitrary closed contour $C$ in $S$ that surrounds $z_0$. Let us denote by $S_1$ the region within $S$ having outer boundary $C$ and inner boundary $C_1$. For $\xi \in S_1$, $\frac{f(\xi)}{(\xi - z_0)^{(n+1)}}$ is analytic as $\xi \neq z_0$. Hence

$$\begin{aligned} 0 &= \oint_{C \cup C_1} \frac{f(\xi)d\xi}{(\xi - z_0)^{(n+1)}} \\ &\Rightarrow \oint_C \frac{f(\xi)d\xi}{(\xi - z_0)^{(n+1)}} = \oint_{C_1} \frac{f(\xi)d\xi}{(\xi - z_0)^{(n+1)}}. \end{aligned} \tag{5.148}$$

Similarly the integral around $C_2$ may also be replaced with one over $C$:

$$\begin{aligned} 0 &= \oint_{C \cup C_2} f(\xi)(\xi - z_0)^n d\xi \\ &\Rightarrow \oint_C f(\xi)(\xi - z_0)^n d\xi = \oint_{C_2} f(\xi)(\xi - z_0)^n d\xi. \end{aligned} \tag{5.149}$$

Substituting these two integrals over $C$ into equation (5.147) gives

$$\begin{aligned} 2\pi i f(z) &= \sum_{n=0}^{\infty} (z - z_0)^n \oint_C \frac{f(\xi)d\xi}{(\xi - z_0)^{(n+1)}} + \sum_{n=0}^{\infty} \frac{1}{(z - z_0)^{(n+1)}} \oint_C f(\xi)(\xi - z_0)^n d\xi \\ &= \sum_{n=0}^{\infty} (z - z_0)^n \oint_C \frac{f(\xi)d\xi}{(\xi - z_0)^{(n+1)}} + \sum_{m=-\infty}^{-1} (z - z_0)^m \oint_C \frac{f(\xi)d\xi}{(\xi - z_0)^{(m+1)}} \\ &= \sum_{n=-\infty}^{\infty} (z - z_0)^n \oint_C \frac{f(\xi)d\xi}{(\xi - z_0)^{(n+1)}} \end{aligned} \tag{5.150}$$

as required. □

**Example 1**

Expand $e^z$ around $z = 0$.

We may use the Taylor series to find

$$e^z = \sum_{n=0}^{\infty} \frac{z^n}{n!} \frac{d^n(e^z)}{dz^n}\bigg|_{z=0} = \sum_{n=0}^{\infty} \frac{z^n}{n!}. \tag{5.151}$$

**Example 2**

Expand $f(z) = \frac{2+3z}{z^2+z^3}$ about $z = 0$.

As the function is not analytic at $z = 0$ we should expand this using the Laurent series. However it is quicker, in practise, to find the power expansion by careful rear-

rangement of the expression:

$$\begin{aligned} f(z) &= \frac{1}{z^2}\left(\frac{2+3z}{1+z}\right) \\ &= \frac{1}{z^2}\left(\frac{3+3z-1}{1+z}\right) \\ &= \frac{1}{z^2}\left(3-\sum_{n=0}^{\infty}(-1)^n z^n\right) \\ &= \frac{1}{z^2}\left(3-1+z-z^2+z^3-\dots\right) \\ &= \frac{2}{z^2}+\frac{1}{z}-1+z-\dots. \end{aligned} \tag{5.152}$$

where the only non-trivial move we have made was to recall that $\sum_{n=0}^{\infty}(-1)^n z^n = \frac{1}{1+z}$.

**Example 3**

Expand $f(z) = \frac{z}{(z-1)(z-2)}$ about $z = 0$.

We observe that the function is not analytic at $z = 1$ and $z = 2$ and split the expansion into three regions $|z| < 1$, $1 < |z| < 2$ and $2 < |z|$. In the first region $|z| < 1$ the function is analytic so we may use the Taylor expansion (only positive powers of $z$):

$$\begin{aligned} f(z) &= \frac{-1}{(z-1)}+\frac{2}{(z-2)} \\ &= \frac{1}{(1-z)}-\frac{1}{(1-\frac{z}{2})} \\ &= \sum_{n=0}^{\infty} z^n - \sum_{n=0}^{\infty}\left(\frac{z}{2}\right)^n \qquad \text{as} \quad |\frac{z}{2}| < |z| < 1 \\ &= \sum_{n=0}^{\infty}(1-2^{-n})z^n. \end{aligned} \tag{5.153}$$

For the region $1 < |z| < 2$ we (formally) find the Laurent series (both positive and negative powers of $z$)

$$\begin{aligned} f(z) &= -\frac{1}{z}\left(\frac{1}{1-\frac{1}{z}}\right)-\frac{1}{(1-\frac{z}{2})} \\ &= -\frac{1}{z}\sum_{n=0}^{\infty}\left(\frac{1}{z}\right)^n - \sum_{n=0}^{\infty}\left(\frac{z}{2}\right)^n \qquad \text{as} \quad |\frac{1}{z}| < 1 \text{ and } |\frac{z}{2}| < 1 \\ &= -\sum_{n=0}^{\infty}\left(\frac{1}{z}\right)^{(n+1)} - \sum_{n=0}^{\infty}\left(\frac{z}{2}\right)^n \\ &= -\sum_{n=-\infty}^{-1} z^n - \sum_{n=0}^{\infty}\left(\frac{z}{2}\right)^n. \end{aligned} \tag{5.154}$$

In the region $|z| > 2$ we find only negative powers of $z$:

$$\begin{aligned} f(z) &= -\frac{1}{z}\left(\frac{1}{1-\frac{1}{z}}\right) + \frac{2}{z}\left(\frac{1}{1-\frac{2}{z}}\right) \\ &= -\frac{1}{z}\sum_{n=0}^{\infty}\left(\frac{1}{z}\right)^n + \frac{2}{z}\sum_{n=0}^{\infty}\left(\frac{2}{z}\right)^n \qquad \text{as} \quad |\frac{1}{z}| < 1 \text{ and } |\frac{2}{z}| < 1 \\ &= \sum_{n=-\infty}^{-1}(-1+\frac{1}{2^n})z^n. \end{aligned} \tag{5.155}$$

Most usefully a Laurent series may be used to investigate closed contour integrals within which the integrand may have non-analytic points. The coefficient $a_{-1}$ of the first negative power of the expansion coefficient in the Laurent series is

$$a_{-1} = \frac{1}{2\pi i}\oint_C f(\xi)d\xi. \tag{5.156}$$

Therefore one way to integrate a non-analytic function around $C$ is to read off $a_{-1}$ from the Laurent series, i.e. for an expansion about $z_0$ $a_{-1}$ is the coefficient of $\frac{1}{z-z_0}$.

**Example**

Evaluate

$$\oint_C \frac{dz}{z^2(z-2)} \tag{5.157}$$

where $C$ is a circle of unit radius centred at the origin.

The function $f(z) = \frac{1}{z^2(z-2)}$ is analytic on the annulus defined by $0 < |z| < 2$. Within the contour of integration the origin is a non-analytic point, so we have a Laurent series expansion and we may evaluate the integral by finding the coefficient of $\frac{1}{z}$. If there had been no singular point within $C$ the integral would of course have been zero: we might understand this by noting that the analytic function is expanded with a Taylor series having only positive powers and hence $a_{-1} = 0$ - which is, of course, consistent with the simple observation we made earlier that the integral would be zero as it is path-independent. Hence we expand to find

$$\begin{aligned} \frac{1}{z^2(z-2)} &= -\frac{1}{2z^2}\left(\frac{1}{1-\frac{z}{2}}\right) \\ &= -\frac{1}{2z^2}\sum_{n=0}^{\infty}\left(\frac{z}{2}\right)^n \\ &= -\frac{1}{2z^2} - \frac{1}{4z} - \frac{1}{8} - \frac{z}{16} - \dots \end{aligned} \tag{5.158}$$

We read off $a_{-1} = -\frac{1}{4}$ and hence we have

$$\begin{aligned} -\frac{1}{4} &= \frac{1}{2\pi i}\oint_C \frac{dz}{z^2(z-2)} \\ \therefore \oint_C \frac{dz}{z^2(z-2)} &= -\frac{i\pi}{2}. \end{aligned} \tag{5.159}$$

For the expansion

$$f(z) = \sum_{-\infty}^{\infty} a_n(z-z_0)^n \tag{5.160}$$

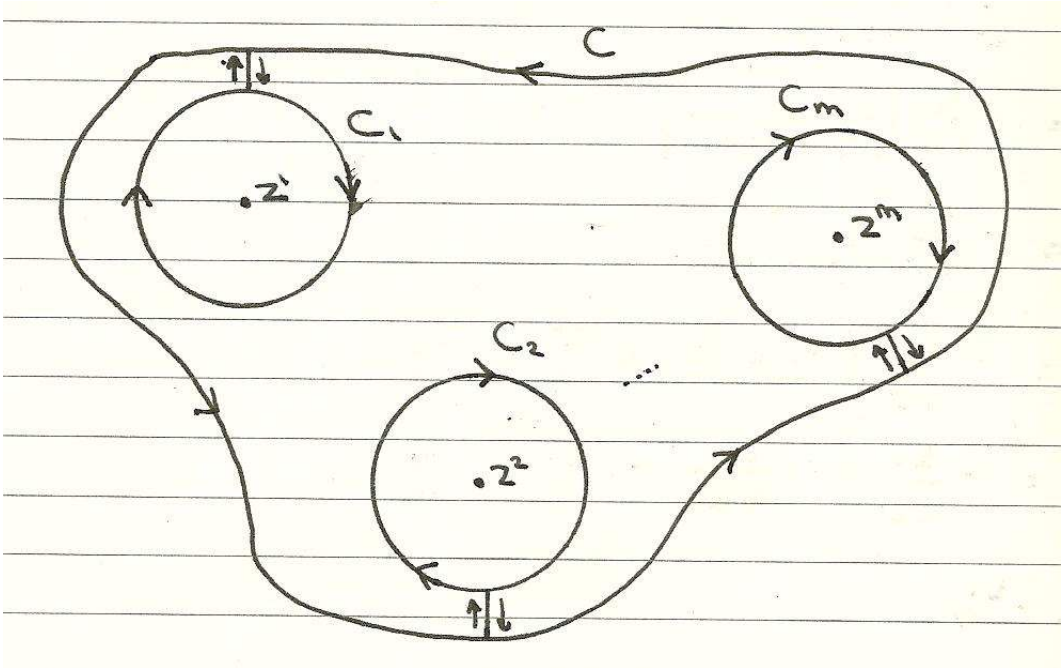
the coefficient $a_{-1}$ is called the residue of $f(z)$ at the isolated $z = z_0$ and denoted $Res[f(z_0)]$:

$$2\pi i Res[f(z_0)] = \oint_C f(z)dz. \tag{5.161}$$

**Theorem 5.4.5.** *(The residue theorem) Let $C$ be a positively oriented simple closed contour bounding a region where $f(z)$ is analytic except at a finite number of singular points $z_1$, $z_2$, $z_3$, ... $z_m$ then*

$$\oint_C f(z)dx = 2\pi i \sum_{k=1}^{m} Res[f(z_k)]$$

*Proof.* Let us denote the curves surrounding the singular points by $C_1$, $C_2$, ... $C_m$ as illustrated:




Now let $C' \equiv C \cup C_1 \cup C_2 \cup \ldots \cup C_m$ then

$$\begin{aligned} 0 &= \oint_{C'} f(z)dz \\ &= \sum_{k=1}^{m} \left[ -\oint_{C_k} f(z)dz \right] + \oint_C f(z)dz \\ &= -\sum_{k=1}^{m} 2\pi i Res[f(z_k)] + \oint_C f(z)dz \\ \therefore \oint_C f(z)dz &= \sum_{k=1}^{m} 2\pi i Res[f(z_k)]. \end{aligned} \tag{5.162}$$

□

**Example**

Evaluate

$$\oint_C \frac{(2z-3)dz}{z(z-1)}$$

where $C$ is the circle of points defined by $|z| = 2$.

There exist two singularities within $C$ at $z_1 \equiv 0$ and $z_2 \equiv 1$. To find the integral we must sum $2\pi i Res[f(z_1)]$ and $2\pi i Res[f(z_2)]$. To find $Res[f(z_1)]$ we expand around $z = 0$:

$$\begin{aligned} \frac{2z-3}{z(z-1)} &= \frac{1}{z}\left(\frac{3z-3-z}{(z-1)}\right) \\ &= \frac{1}{z}\left(3 - \frac{z}{z-1}\right) \\ &= \frac{3}{z} + \frac{1}{1-z} \\ &= \frac{3}{z} + \sum_{n=0}^{\infty} z^n \qquad \text{as} \quad |z| < 1 \end{aligned} \tag{5.163}$$

So $Res[f(z_1)] = 3$.
Expanding around $z = 1$ (so that $(z-1)$ is the expansion parameter) gives

$$\begin{aligned} \frac{2z-3}{z(z-1)} &= \frac{3}{z} - \frac{1}{z-1} \\ &= \frac{3}{z-1+1} - \frac{1}{z-1} \\ &= 3\sum_{n=0}^{\infty}(-1)^n(z-1)^n - \frac{1}{z-1} \end{aligned} \tag{5.164}$$

hence the coefficient of $\frac{1}{z-1}$ tells us that $Res[f(z_2)] = -1$.

$$\begin{aligned} \therefore \oint_C f(z)dz &= 2\pi i Res[f(z_1)] + 2\pi i Res[f(z_2)] \\ &= 4\pi i. \end{aligned} \tag{5.165}$$