Symmetry and Relativity

University of Oxford

Third Year, Part B2

Caroline Terquem
Department of Physics

caroline.terquem@physics.ox.ac.uk

Michaelmas Term 2026

Contents

1 Symmetries **7**

1.1 Newton's law of motion and continuous symmetries . . . 9

1.1.1 Translation in space . . . 9

1.1.2 Rotation . . . 9

1.1.3 Translation in time . . . 10

1.2 Lagrangian and Hamiltonian formalisms . . . 11

1.2.1 Generalised coordinates . . . 11

1.2.2 D'Alembert's principle . . . 12

1.2.3 Euler-Lagrange equations . . . 14

1.2.4 The principle of least action . . . 17

1.2.5 Lagrangian for an electromagnetic force . . . 20

1.2.6 Field theory for continuous systems . . . 21

1.3 Conservation laws and Noether's theorem . . . 25

1.3.1 Cyclic coordinates . . . 25

1.3.2 Noether's theorem . . . 27

1.3.3 Applications . . . 30

2 Groups and the Lorentz transformation **33**

2.1 Introduction to groups . . . 33

2.1.1 Definition . . . 33

2.1.2 Examples . . . 34

2.1.3 Groups of matrices . . . 35

2.2 Representations . . . 36

2.2.1 Definition of a representation . . . 36

2.2.2 Examples . . . 37

2.3 Lie algebra and generators . . . 39

2.3.1 The SO(2) group . . . 39

2.3.2 The SO(3) group . . . 41

2.3.3 Commutators . . . 43

2.4 Galilean and Lorentz transformations in mechanics . . . 44

2.4.1 The Galilean group . . . 44

2.4.2 Relativistic mechanics: invariance of the spacetime interval . . . 49

2.4.3 The Lorentz group SO(1, 3) . . . 51

3 Fundamental concepts in special relativity **57**
3.1 Deriving the Lorentz transformation for a boost . . . 58
3.2 Spacetime diagram . . . 60
3.3 Relativistic effects on measurements . . . 60
3.3.1 Proper length and length contraction . . . 60
3.3.2 Proper time and time dilation . . . 61
3.4 The invariance of the spacetime interval . . . 62
3.4.1 Classification of worldlines . . . 62
3.4.2 Causality in relativity . . . 64
3.5 Relativistic velocity . . . 64
3.5.1 Velocity composition . . . 64
3.5.2 Velocity transformation . . . 65
3.6 Four-vectors . . . 66
3.6.1 Overview and definition . . . 66
3.6.2 Minkowski metric: length and scalar product . . . 67
3.6.3 Index notation, covariant and contravariant four-vectors . . . 69
3.6.4 Covariant and contravariant derivatives . . . 73
3.6.5 Transformations of four-vectors in index notation . . . 76

4 Dynamics and conservation laws **79**
4.1 Four-velocity . . . 79
4.1.1 Derivation of the four-velocity . . . 79
4.1.2 Velocity transformation using four-vectors . . . 80
4.2 Four-momentum . . . 81
4.2.1 Lagrangian in special relativity . . . 81
4.2.2 Relativistic momentum and energy . . . 83
4.2.3 Four-momentum and energy-momentum relation . . . 85
4.2.4 Four-momentum of massless particles . . . 85
4.2.5 Wave four-vector . . . 86
4.3 Four-force . . . 87
4.4 Four-acceleration and instantaneous rest frame . . . 89
4.4.1 Four-acceleration . . . 89
4.4.2 Instantaneous rest frame . . . 89
4.4.3 Proper acceleration and rapidity . . . 91
4.5 Particle collisions . . . 92
4.5.1 Total four-momentum . . . 93
4.5.2 Centre of momentum frame . . . 94
4.5.3 Elastic collision: proton-proton scattering . . . 95
4.5.4 Emission and absorption . . . 96
4.5.5 Proton-proton collision and pion production: threshold energy . . . 98
4.5.6 Annihilation processes . . . 99
4.5.7 Compton scattering . . . 99

5 Special relativity and electromagnetism **101**
5.1 Tensors . . . 101
5.1.1 Preliminary . . . 101
5.1.2 Definition . . . 103
5.1.3 Operations on tensors . . . 104
5.1.4 Symmetric and antisymmetric tensors . . . 107
5.2 Invariance of the electric charge . . . 107
5.3 Four-vectors for the current and potential . . . 108
5.3.1 Definition of the four-current . . . 108
5.3.2 Conservation law . . . 109
5.3.3 Lorenz gauge . . . 109
5.3.4 Definition of the four-potential . . . 111
5.4 Electromagnetic tensor . . . 113
5.4.1 Definition of the electromagnetic tensor . . . 113
5.4.2 Maxwell's equations in terms of the electromagnetic tensor . . . 114
5.5 Lorentz transformation of the electromagnetic field . . . 115
5.6 The Lorentz force . . . 117
5.7 Lagrangian formulation of electromagnetism . . . 118
5.7.1 Covariant relativistic Lagrangian for a particle . . . 118
5.7.2 Euler-Lagrange equations . . . 120
5.7.3 Lagrangian for the electromagnetic field . . . 122
5.8 Energy and momentum conservation for fields in vacuum . . . 123
5.8.1 Noether's theorem for the electromagnetic field . . . 123
5.8.2 Canonical energy-momentum tensor . . . 124
5.8.3 Symmetric energy-momentum tensor . . . 126
5.8.4 Physical interpretation of the energy-momentum tensor . . . 127
5.9 Exchange of energy and momentum with charges . . . 129
5.9.1 Energy-momentum exchange equation . . . 129
5.9.2 Energy and momentum conservation . . . 130

6 Radiation from moving charges **133**
6.1 Solution of the wave equation . . . 133
6.1.1 Field due to a uniformly moving charge . . . 134
6.1.2 Retarded potentials . . . 135
6.1.3 Liénard-Wiechert potential . . . 138
6.1.4 Field due to an accelerated charge . . . 140
6.2 Radiated power . . . 142
6.2.1 Radiation from a non-relativistic accelerated point charge . . . 143
6.2.2 Radiation from a relativistic accelerated point charge . . . 148

A Radiation from moving charges **151**

These notes borrow from the following books (with the chapters of the lecture notes for which the books are most relevant indicated in parentheses):

F. Laloe, *Introduction to Continuous Symmetries*, Oxford University Press, 2023, Wiley VCH (chapters 1 and 2)

H. Goldstein, *Classical Mechanics*, 2nd edition, Addison Wesley, 1980 (chapter 1, 3 and 4)

A. Zee, *Group Theory in a Nutshell for Physicists*, Princeton University Press, 2016 (chapter 2)

L. Susskind and A. Friedman, *Special Relativity and Classical Field Theory, The Theoretical Minimum*, Penguin Books, 2017 (chapters 3, 4 and 5)

D. Morin, *Introduction to Classical Mechanics*, Cambridge University Press, 2008 (chapters 3 and 4)

A. M. Steane, *Relativity Made Relatively Easy*, Oxford University Press, 2012 (chapters 3 and 4)

J. D. Jackson, *Classical Electrodynamics*, 3rd edition, Wiley, 1999 (chapters 5 and 6)

These notes are intended to supplement the course, but they should not be seen as a substitute for textbooks. It is highly recommended to regularly consult textbooks, as they offer much more detailed explanations, along with numerous examples and problems.

Chapter 1

Symmetries

Symmetries underlie all fundamental physical laws. They play a crucial role in shaping the theory of quantum mechanics, and lead to the theory of relativity once the invariance of the speed of light is assumed.

The most straightforward way to define a symmetry is to use Feynman's rewording of Weyl[1]: *a thing is symmetrical if there is something we can do to it so that after we have done it, it looks the same as it did before*[2].

In classical physics, a symmetry is a transformation that leaves the equations of motion unchanged, meaning it transforms one possible motion into another possible motion, governed by the same laws. Symmetries therefore restrict the solutions of the equations of motion, and can be used to put constraints on these solutions. For example, consider the rotation around the axis of angular momentum of a planet orbiting a star: rotating the entire orbit (which may be elliptical) results in another valid orbit with the same energy and eccentricity.

In quantum mechanics, the consequences of a symmetry are richer, because states can be superposed. For example, if a system is rotationally invariant and $|\psi\rangle$ is an energy eigenstate, the rotated state $R|\psi\rangle$ is also an energy eigenstate with the same energy. Any superposition of these states, such as $|\psi\rangle + R|\psi\rangle$, is then again an eigenstate with that energy. Symmetries can therefore lead to degenerate energy levels, as in the $(2l+1)$-fold degeneracy of states with angular momentum l in a central potential. In classical mechanics, by contrast, two trajectories cannot in general be combined to produce another valid trajectory, except for linear systems such as the harmonic oscillator.

Another important symmetry in quantum mechanics involves the exchange of identical particles. Bosons are characterised by the fact that exchanging two identical bosons leaves the wavefunction unchanged, whereas fermions exhibit the opposite behaviour: exchanging two identical fermions causes the wavefunction to change sign.

[1] Hermann Weyl, *Symmetry*, 1952, Princeton University Press, Princeton, NJ

[2] Richard P. Feynman, Robert B. Leighton, Matthew Sands, 1963, *The Feynman Lectures on Physics – Volume 1*, Chap. 52 *Symmetry in Physical Laws*, Addison-Wesley

These symmetries are examples of *global symmetries*, meaning the transformations are the same at every point in space and time. There is another type of invariance which is called *gauge symmetry*: it is related to the fact that fundamental laws can be described in different ways. Different descriptions are related by a transformation whose parameter is a function of position and time, and can therefore be chosen independently at each point. The transformations associated with gauge symmetries do not change the physical state of the system; rather, they give a different but precisely equivalent mathematical description of the same physical situation. The introductory review *The role of symmetry in fundamental physics* by David J. Gross[3] is recommended for more details.

A symmetry is either *continuous* or *discrete*.

Continuous symmetries involve transformations that can be described using one or more continuous parameters, resulting in an infinite number of possible transformations. Examples include rotations around a point or an axis, where the continuous parameter is the rotation angle, and translations, where the continuous parameter is the distance by which objects or fields are shifted. Lorentz symmetries, which include both rotations and boosts (changes in velocity) are also continuous symmetries. According to Noether's theorem, every continuous symmetry corresponds to a conserved quantity.

Discrete symmetries involve transformations that cannot be described by any continuous parameter, resulting in a finite or countably infinite number of possible transformations. Examples include parity symmetry, such as reflection through the origin ($\mathbf{r} \to -\mathbf{r}$), time reversal ($t \to -t$) and charge conjugation (transformation that changes particles into their antiparticles). Crystal symmetries are also discrete because they are defined by specific mathematical operations that map the crystal onto itself. For instance, if the periodicity of a crystal lattice is $\mathbf{R}$, only translations that shift the crystal by an integer multiple of $\mathbf{R}$ will leave the crystal unchanged. Unlike continuous symmetries, discrete symmetries are not associated with conservation laws through Noether's theorem, although in quantum mechanics they lead to conserved quantum numbers, such as parity, which give rise to selection rules.

In what follows, we will focus on continuous symmetries.

[3] Proceedings of the National Academy of Sciences of the USA, 1996, p. 14256–14259

1.1 Newton's law of motion and continuous symmetries

In this section, we illustrate how continuous symmetries imply conservation laws starting from Newton's law of motion. We will show that invariance under translation, rotation, and time translation corresponds respectively to conservation of momentum, angular momentum and energy.

Consider an isolated system of two masses m_1 and m_2 with position vectors $\mathbf{r}_1(t)$ and $\mathbf{r}_2(t)$ which exert the forces $\mathbf{f}_1$ and $\mathbf{f}_2$ on each other. The equations of motion are:

$$\begin{aligned} m_1\ddot{\mathbf{r}}_1 &= \mathbf{f}_1\left(\mathbf{r}_1(t), \mathbf{r}_2(t), t\right), \\ m_2\ddot{\mathbf{r}}_2 &= \mathbf{f}_2\left(\mathbf{r}_1(t), \mathbf{r}_2(t), t\right), \end{aligned} \tag{1.1}$$

where the dots represent time derivatives. We assume that the forces are conservative and denote by U the interaction potential, such that $\mathbf{f}_i = -\boldsymbol{\nabla}_{\mathbf{r}_i} U$, where $i = 1, 2$ and the subscript denotes the coordinates with respect to which the derivatives are taken.

1.1.1 Translation in space

For an isolated system, there is no preferred position in space: translating the entire system by a constant vector $\mathbf{c}$ leaves its potential energy, and hence the forces, unchanged. Therefore, $U(\mathbf{r}_1 + \mathbf{c}, \mathbf{r}_2 + \mathbf{c}, t) = U(\mathbf{r}_1, \mathbf{r}_2, t)$. By setting $\mathbf{c} = -\mathbf{r}_2(t_0)$, where t_0 is an arbitrary time, we then get, for all values of t:

$$U\left(\mathbf{r}_1(t) - \mathbf{r}_2(t_0), \mathbf{r}_2(t) - \mathbf{r}_2(t_0), t\right) = U\left(\mathbf{r}_1(t), \mathbf{r}_2(t), t\right).$$

This equality holds in particular at time $t = t_0$, yielding

$$U\left(\mathbf{r}_1(t_0) - \mathbf{r}_2(t_0), \mathbf{0}, t_0\right) = U\left(\mathbf{r}_1(t_0), \mathbf{r}_2(t_0), t_0\right).$$

Since t_0 is arbitrary, this shows that U depends solely on the *relative* position $\mathbf{r} = \mathbf{r}_1 - \mathbf{r}_2$ of the particles. Consequently, $\mathbf{f}_1 = -\mathbf{f}_2$ (weak form of the law of action and reaction), and the forces themselves depend only on $\mathbf{r}$. The equations of motion then yield:

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(m_1\dot{\mathbf{r}}_1 + m_2\dot{\mathbf{r}}_2\right) = \mathbf{0}. \tag{1.2}$$

Invariance under translation therefore implies conservation of momentum.

1.1.2 Rotation

We now assume that the system is invariant under any rotation about any axis through the origin. From section 1.1.1, the forces depend only on the relative position $\mathbf{r} = \mathbf{r}_1 - \mathbf{r}_2$, with $\mathbf{f}_2 = -\mathbf{f}_1$, and we write $\mathbf{f}_1 = \mathbf{f}(\mathbf{r})$. Rotational invariance means that rotating the configuration of the system rotates the forces accordingly. Therefore, $\mathbf{f}(R\mathbf{r}) = R\,\mathbf{f}(\mathbf{r})$ for all rotations R.

Consider first a rotation R about an axis aligned with $\mathbf{r}$ itself. Such a rotation leaves $\mathbf{r}$ unchanged, $R\mathbf{r} = \mathbf{r}$, so the condition above gives $\mathbf{f}(\mathbf{r}) = R\,\mathbf{f}(\mathbf{r})$: the force is unchanged by any rotation about $\mathbf{r}$. This is only possible if $\mathbf{f}(\mathbf{r})$ has no component perpendicular to $\mathbf{r}$, since such a component would be rotated. The force is therefore parallel to $\mathbf{r}$, and we can write $\mathbf{f}(\mathbf{r}) = f(\mathbf{r})\,\hat{\mathbf{r}}$, where a hat denotes a unit vector.
Consider now a rotation R that takes $\mathbf{r}$ to another vector $\mathbf{r}' = R\mathbf{r}$ with the same modulus but a different direction. The condition $\mathbf{f}(R\mathbf{r}) = R\,\mathbf{f}(\mathbf{r})$ becomes $f(\mathbf{r}')\,\hat{\mathbf{r}}' = f(\mathbf{r})\,R\hat{\mathbf{r}} = f(\mathbf{r})\,\hat{\mathbf{r}}'$, so that $f(\mathbf{r}') = f(\mathbf{r})$. The magnitude of the force therefore depends only on the modulus r of $\mathbf{r}$, and not on its direction:

$$\mathbf{f}_1 = -\mathbf{f}_2 = f(r)\,\hat{\mathbf{r}}.$$

Correspondingly, the potential U depends only on r. The internal forces are thus central, i.e. they are equal and opposite, and directed along the line joining the two particles. This is the strong form of the law of action and reaction. It follows that:

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(m_1\mathbf{r}_1\times\dot{\mathbf{r}}_1 + m_2\mathbf{r}_2\times\dot{\mathbf{r}}_2\right) = \mathbf{r}_1\times\mathbf{f}_1 + \mathbf{r}_2\times\mathbf{f}_2 = (\mathbf{r}_1 - \mathbf{r}_2)\times\mathbf{f}_1 = \mathbf{0}. \tag{1.3}$$

Thus, rotational invariance implies the conservation of angular momentum. Note that invariance under rotations about a single fixed axis is a weaker condition, as it only implies conservation of the component of the angular momentum along that axis.

1.1.3 Translation in time

Here we assume that the system is invariant under a time translation that shifts t backwards by some arbitrary interval τ. This implies $\mathbf{f}_i\left(\mathbf{r}_1, \mathbf{r}_2, t-\tau\right) = \mathbf{f}_i\left(\mathbf{r}_1, \mathbf{r}_2, t\right)$. Therefore, the change in U associated with a time translation can only be a function of t, and possibly τ, which we denote as g: $U\left(\mathbf{r}_1, \mathbf{r}_2, t-\tau\right) = U\left(\mathbf{r}_1, \mathbf{r}_2, t\right) + g(t,\tau)$. A function of time alone does not contribute to the forces and can be removed by redefining U, so we take $g = 0$. The equality above is satisfied in particular at time $t = \tau$, so $U\left(\mathbf{r}_1, \mathbf{r}_2, 0\right) = U\left(\mathbf{r}_1, \mathbf{r}_2, \tau\right)$. Since τ is arbitrary, this implies that U *does not explicitly depend on time* (it only depends on time through $\mathbf{r}_1$ and $\mathbf{r}_2$). It follows that:

$$\frac{\mathrm{dU}}{\mathrm{d}t} = \dot{\mathbf{r}}_1\cdot\boldsymbol{\nabla}_{\mathbf{r}_1}U + \dot{\mathbf{r}}_2\cdot\boldsymbol{\nabla}_{\mathbf{r}_2}U = -\dot{\mathbf{r}}_1\cdot\mathbf{f}_1 - \dot{\mathbf{r}}_2\cdot\mathbf{f}_2 = -\frac{\mathrm{dK}}{\mathrm{d}t}, \tag{1.4}$$

where $K = \dfrac{1}{2}\left(m_1\dot{\mathbf{r}}_1^2 + m_2\dot{\mathbf{r}}_2^2\right)$ is the kinetic energy.

Invariance under time translation therefore implies conservation of the total energy $K+U$.

1.2 Lagrangian and Hamiltonian formalisms

The examples above illustrate how continuous symmetries lead to conservation laws. However, they also show that Newton's law of motion is not the most effective approach for deriving general principles about these laws. For this purpose, we use the Lagrangian and Hamiltonian formalisms, which were introduced in the first-year Mechanics course[4]. In that course, the Euler-Lagrange equations were derived from the principle of least action. However, this does not explain why the Lagrangian is equal to the kinetic energy minus the potential energy for conservative forces. Noether's theorem was also introduced, but only for the limited case of transformations of the generalised coordinates, with time held constant. Here, we will derive the Euler-Lagrange equations from d'Alembert's principle, which offers the advantage of enabling us to calculate the Lagrangian. Additionally, we will derive Noether's theorem for the more general case where transformations of both the generalised coordinates and time are considered.

1.2.1 Generalised coordinates

We consider a system composed of N particles, each with mass m_i and position vector $\mathbf{r}_i(t)$. Constraints may exist, implying that the $\mathbf{r}_i$ are not all independent.

One type of constraint, known as *holonomic*, can be expressed as algebraic equations involving the coordinates of the particles and, possibly, time. For example, a particle restricted to move along a circular path of radius R must satisfy $x^2 + y^2 = R^2$. Another example is the rolling without slipping condition for a cylinder of radius R along a table in one dimension. The relevant coordinates are x, the distance along the table, and θ, the angle of rotation of the cylinder. The rolling without slipping condition is $\dot{x} = R\dot{\theta}$, which integrates into $x - R\theta = \text{const}$. In such cases, we can introduce a set of *generalised coordinates* $q_k(t)$, with k ranging from 1 to some integer less than $3N$, by eliminating the constraints. These coordinates, which do not necessarily have the units of length, are generally not simply related to the position vectors, but they can be convenient to use even in the absence of constraints. We refer to $\dot{q}_k = \mathrm{d}q_k/\mathrm{d}t$ as generalised velocities, although they do not necessarily have the units of conventional velocity.

Other types of constraints are referred to as *nonholonomic*. An example is the rolling of a disc with radius R on a two-dimensional table without slipping, where the plane of the disc stays vertical, but the disc can freely rotate about a vertical axis. The relevant coordinates are x and y, which represent the position of the disc's centre on the table, θ, the rotation angle of the disc about its own axis, and α, the angle between the plane of the disc and the x-axis, which gives the direction of motion. The rolling without slipping condition gives the velocity-dependent constraints $\dot{x} = R\dot{\theta}\cos\alpha$ and $\dot{y} = R\dot{\theta}\sin\alpha$. These conditions cannot be integrated into a relation involving only the coordinates, and thus cannot be eliminated by reducing the number of generalised coordinates.

[4]See also Prof. David Marshall's S7/B7 *Classical Mechanics* Lecture Notes

Constraints may also take the form of inequalities, restricting the system to a certain region of configuration space. For example, a particle confined to move inside a sphere of radius R must satisfy $x^2 + y^2 + z^2 < R^2$. Such constraints likewise do not reduce the number of generalised coordinates.

Although the rolling without slipping condition in two dimensions does not reduce the number of generalised coordinates, it imposes relations between the allowed generalised velocities. For example, if $x(t)$ and $\theta(t)$ are known, $\dot{y}$ is determined, up to a sign, by $\dot{y}^2 = R^2\dot{\theta}^2 - \dot{x}^2$ and y follows by integration. Thus, the coordinates x, y, θ and α may be used to specify the configuration of the disc, even though their velocities cannot vary independently.

Constraints are often enforced by *forces of constraint*, such as the friction force in the case of rolling without slipping, or forces that keep particles confined within a certain region of space.

1.2.2 D'Alembert's principle

Forces of constraint restrict motion to certain directions or surfaces, ensuring that the configuration of the system remains consistent with those constraints. For example, a normal force prevents penetration through a surface, tension in a rod prevents elongation or compression and, in rolling motion without slipping, friction prevents sliding at the contact point. If slipping were allowed, friction could do work, introducing non-ideal constraints beyond this analysis.

We now introduce the concept of *virtual displacements*, $\delta\mathbf{r}_i$, which are infinitesimal, hypothetical changes in the positions $\mathbf{r}_i$ of the particles of the system. These displacements are 'frozen in time', meaning they represent possible instantaneous changes in configuration that are consistent with the constraints of the system, without considering the actual motion or dynamics.

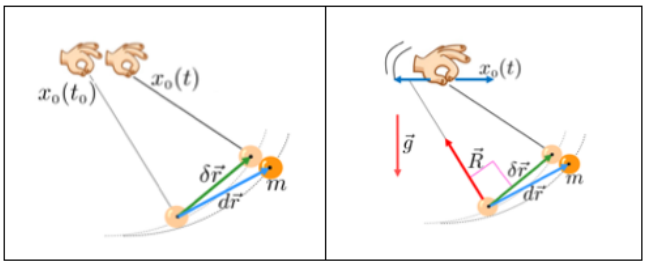
For example, for a bead sliding on a rigid ring, a virtual displacement lies along the circumference of the ring, respecting the constraint that the bead cannot leave it. The normal force from the ring is perpendicular to this displacement and therefore performs no work. This principle extends to more complex systems, where the geometry of the constraints determines the possible directions of virtual displacements.

In general, however, constraint forces need not be perpendicular to the motion or to the virtual displacement of each particle individually. For example, in an Atwood machine, the tension in the string does virtual work on each mass separately, but the two contributions cancel because the virtual displacements are related by the constraint. The hallmark of an *ideal constraint* is that the total virtual work of all constraint forces on the system vanishes:

$$\sum_i \mathbf{F}_{c,i} \cdot \delta\mathbf{r}_i = 0. \tag{1.5}$$

This is the defining property of ideal constraints used in d'Alembert's principle: although constraint forces may exchange energy between parts of the system, they do no net virtual work on the system as a whole.

The figure below[5] illustrates the concept of a virtual displacement for a pendulum whose pivot moves in time, along the trajectory $x_0(t)$. The rod is rigid, so a virtual displacement $\delta\mathbf{r}$ at time t_0 moves the mass along the circular arc defined by the position of the pivot at that time. In contrast, the actual displacement $\mathrm{d}\mathbf{r}$ occurs while the pivot itself moves, so the trajectory of the mass is not exactly circular.



The virtual displacement $\delta\mathbf{r}$ is orthogonal to the instantaneous tension $\mathbf{R}$, i.e. $\mathbf{R}\cdot\delta\mathbf{r}=0$. However, because the pivot moves during the actual displacement, $\mathrm{d}\mathbf{r}$ need not be orthogonal to the instantaneous tension $\mathbf{R}$, i.e. $\mathbf{R}\cdot d\mathbf{r}\neq 0$. The constraint is still ideal, because it does no virtual work.

Virtual displacements describe all possible instantaneous changes consistent with the constraints, independent of the dynamics of the system. This makes them particularly convenient in the formulation of d'Alembert's principle, which can then be expressed without prior knowledge of the actual motion.

Each particle i is subject to both forces of constraint and additional forces, not related to the constraints, denoted by $\mathbf{F}_{nc,i}$. From Newton's law of motion: $\mathbf{F}_{c,i}+\mathbf{F}_{nc,i}-m_i\ddot{\mathbf{r}}_i=\mathbf{0}$. Taking the dot product with $\delta\mathbf{r}_i$, summing over all particles i and applying equation (1.5) leads to the conclusion that, for any set of virtual displacements $\delta\mathbf{r}_i$:

$$\boxed{\sum_{i=1}^{N}(\mathbf{F}_{nc,i}-m_i\ddot{\mathbf{r}}_i)\cdot\delta\mathbf{r}_i=0.} \tag{1.6}$$

This is d'Alembert's principle, established in 1743. It recasts Newton's second law as a condition of equilibrium: for every virtual displacement allowed by the constraints, the total virtual work of the applied non-constraint forces and the inertial forces $-m_i\ddot{\mathbf{r}}_i$ is zero. A dynamical system thus behaves as if it were momentarily in equilibrium when inertial effects are treated as additional 'fictitious forces'. This bridges statics and dynamics and, most importantly, eliminates the unknown constraint forces, paving the way for the formulation of Lagrange's equations and the broader framework of analytical mechanics.

Practically, d'Alembert's principle provides a unified framework for analysing constrained motion, from idealised systems to complex mechanisms in engineering.

[5] From X. Jaén, J. Salud, C. Serra, J. Calaf, M. Khoury, 2023, *Fundamental Mechanics, Newtonian Mechanics for Engineering*, Iniciativa Digital Politècnica, Universitat Politécnica de Catalunya, Barcelona

1.2.3 Euler-Lagrange equations

Building on d'Alembert's principle, Lagrange introduced in 1788 a more elegant formulation, now known as the Euler-Lagrange equations, that expresses the dynamics of a constrained system in terms of generalised coordinates $q_k(t)$. These coordinates are chosen so that the constraints are automatically satisfied, reducing the problem to independent variables only. For instance, in a pendulum, the constraint of fixed length is handled by using the angle θ as the generalised coordinate, eliminating the need to enforce the string length constraint explicitly.

The position vectors $\mathbf{r}_i(t)$ can be written as a function of $q_k(t)$ and may also depend explicitly on time. Virtual displacements can then be expressed as:

$$\delta\mathbf{r}_i = \sum_k \frac{\partial \mathbf{r}_i}{\partial q_k}\delta q_k + \frac{\partial \mathbf{r}_i}{\partial t}\delta t = \sum_k \frac{\partial \mathbf{r}_i}{\partial q_k}\delta q_k, \tag{1.7}$$

where we set $\delta t = 0$ because virtual displacements are defined at fixed time.

Substituting this expression into the first term of equation (1.6), we obtain:

$$\sum_i \mathbf{F}_{nc,i}\cdot\delta\mathbf{r}_i = \sum_{i,k}\mathbf{F}_{nc,i}\cdot\frac{\partial\mathbf{r}_i}{\partial q_k}\delta q_k = \sum_k \mathcal{F}_k\delta q_k, \tag{1.8}$$

where the *generalised force* $\mathcal{F}_k$ is defined as:

$$\mathcal{F}_k = \sum_i \mathbf{F}_{nc,i}\cdot\frac{\partial\mathbf{r}_i}{\partial q_k} \tag{1.9}$$

While $\mathcal{F}_k$ may not have the units of a force, $\mathcal{F}_k\delta q_k$ does have the units of work.

We now consider the term involving acceleration in equation (1.6):

$$\sum_i m_i\ddot{\mathbf{r}}_i\cdot\delta\mathbf{r}_i = \sum_{i,k} m_i\ddot{\mathbf{r}}_i\cdot\frac{\partial\mathbf{r}_i}{\partial q_k}\delta q_k = \sum_{i,k}\left[\frac{\mathrm{d}}{\mathrm{d}t}\left(m_i\dot{\mathbf{r}}_i\cdot\frac{\partial\mathbf{r}_i}{\partial q_k}\right) - m_i\dot{\mathbf{r}}_i\cdot\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial\mathbf{r}_i}{\partial q_k}\right)\right]\delta q_k \tag{1.10}$$

We have:

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial\mathbf{r}_i}{\partial q_k}\right) = \sum_l \frac{\partial}{\partial q_l}\left(\frac{\partial\mathbf{r}_i}{\partial q_k}\right)\dot{q}_l + \frac{\partial}{\partial t}\left(\frac{\partial\mathbf{r}_i}{\partial q_k}\right) = \frac{\partial}{\partial q_k}\underbrace{\left(\sum_l\frac{\partial\mathbf{r}_i}{\partial q_l}\dot{q}_l + \frac{\partial\mathbf{r}_i}{\partial t}\right)}_{\dot{\mathbf{r}}_i} \tag{1.11}$$

In deriving the last equality, we used $\partial\dot{q}_l/\partial q_k = 0$. When we perform a virtual displacement, we take a configuration of the system characterised by the generalised coordinates q_k and velocities $\dot{q}_l = \mathrm{d}q_l/\mathrm{d}t$, and change q_k by δq_k instantaneously. Since there is no time evolution, the generalised velocities $\dot{q}_l$ are not affected by the changes in q_k. We are therefore effectively considering that $\dot{q}_l$ is independent of q_k during the displacement, and thus $\partial\dot{q}_l/\partial q_k = 0$, even if along a real trajectory $\dot{q}_l$ generally depends on the q_k.

With $\mathbf{r}_i$ a function of q_k and t, the velocity is given by:

$$\dot{\mathbf{r}}_i \equiv \frac{\mathrm{d}\mathbf{r}_i}{\mathrm{d}t} = \sum_k \frac{\partial \mathbf{r}_i}{\partial q_k}\dot{q}_k + \frac{\partial \mathbf{r}_i}{\partial t}, \tag{1.12}$$

where we retain the time derivative since we are calculating the velocity, not a change due to a virtual displacement. The velocity $\dot{\mathbf{r}}_i$ is then a linear function of the generalised velocities $\dot{q}_k$, which leads to the relation:

$$\frac{\partial \mathbf{r}_i}{\partial q_k} = \frac{\partial \dot{\mathbf{r}}_i}{\partial \dot{q}_k} \tag{1.13}$$

Substituting this expression and equation (1.11) into equation (1.10) gives:

$$\sum_i m_i \ddot{\mathbf{r}}_i \cdot \delta\mathbf{r}_i = \sum_{i,k} \left[\frac{\mathrm{d}}{\mathrm{d}t}\left(m_i \dot{\mathbf{r}}_i \cdot \frac{\partial \dot{\mathbf{r}}_i}{\partial \dot{q}_k} \right) - m_i \dot{\mathbf{r}}_i \cdot \frac{\partial \dot{\mathbf{r}}_i}{\partial q_k} \right] \delta q_k \tag{1.14}$$

Defining the kinetic energy as:

$$T = \sum_i \frac{1}{2} m_i \dot{\mathbf{r}}_i^2, \tag{1.15}$$

this equation becomes:

$$\sum_i m_i \ddot{\mathbf{r}}_i \cdot \delta\mathbf{r}_i = \sum_k \left[\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial T}{\partial \dot{q}_k} \right) - \frac{\partial T}{\partial q_k} \right] \delta q_k \tag{1.16}$$

Substituting equations (1.8) and (1.16) into d'Alembert's principle (1.6) then yields:

$$\sum_k \left[\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial T}{\partial \dot{q}_k} \right) - \frac{\partial T}{\partial q_k} - \mathcal{F}_k \right] \delta q_k = 0 \tag{1.17}$$

Equation (1.17) holds for any system, but if the constraints are holonomic (so the generalised coordinates are independent), each coefficient of δq_k must vanish individually, yielding:

$$\boxed{\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial T}{\partial \dot{q}_k} \right) - \frac{\partial T}{\partial q_k} = \mathcal{F}_k, \quad \text{for all } k} \tag{1.18}$$

These equations apply whether the forces $\mathbf{F}_{nc,i}$ can be derived from a potential or not.

If the forces derive from a scalar potential $V(q_k, t)$, we have $\mathbf{F}_{nc,i} = -\boldsymbol{\nabla}_{\mathbf{r}_i} V$, where the subscript indicates that the derivatives are taken with respect to the coordinates of particle i. Then, equation (1.9) can be rewritten as:

$$\mathcal{F}_k = -\sum_i \boldsymbol{\nabla}_{\mathbf{r}_i} V \cdot \frac{\partial \mathbf{r}_i}{\partial q_k} = -\frac{\partial V}{\partial q_k} \tag{1.19}$$

Since V depends only on the coordinates and possibly on time, but not on the velocities, it does not involve the $\dot{q}_k$. Therefore, we can write $\partial T/\partial \dot{q}_k = \partial\left(T - V\right)/\partial \dot{q}_k$.

Substituting this expression, along with equation (1.19), into equation (1.18) yields the Euler-Lagrange equations:

$$\boxed{\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial L}{\partial \dot{q}_k}\right) - \frac{\partial L}{\partial q_k} = 0, \quad \text{for all } k,} \tag{1.20}$$

where the *Lagrangian* is defined as:

$$\boxed{L\left[q_k(t), \dot{q}_k(t), t\right] = T - V.} \tag{1.21}$$

For velocity-dependent forces, such as the magnetic part of the electromagnetic force, which cannot be derived from a scalar potential $V(q_k, t)$, a generalised potential $\mathcal{V}$ can be introduced, which depends on both the q_k and their time derivatives $\dot{q}_k$. This potential is constructed so that equation (1.20) with $L = T - \mathcal{V}$ remains valid. This will be discussed later in this chapter.

Now, consider the simple case where there are no constraints, using Cartesian coordinates $\mathbf{r}_i = (x_i, y_i, z_i)$, with $q_{3i-2} = x_i$, $q_{3i-1} = y_i$ and $q_{3i} = z_i$ for $i = 1 \ldots N$. The Lagrangian for this system is:

$$L = \sum_i \frac{1}{2} m_i \left(\dot{x}_i^2 + \dot{y}_i^2 + \dot{z}_i^2\right) - V\left(x_i, y_i, z_i\right). \tag{1.22}$$

Applying the Euler-Lagrange equations (1.20), we obtain:

$$m_i \ddot{\mathbf{r}}_i = -\boldsymbol{\nabla}_{\mathbf{r}_i} V, \quad \text{for } i = 1 \ldots N, \tag{1.23}$$

which is simply Newton's law of motion. This equivalence is expected, since d'Alembert's principle, and hence the Euler-Lagrange equations, originate directly from Newton's laws. The significance, however, lies in the generality of the Lagrangian formulation, which reproduces Newtonian mechanics in simple Cartesian cases but extends seamlessly to systems with constraints or curvilinear coordinates, where Newton's laws would be cumbersome to apply. Furthermore, these equations arise from a principle far more general than Newton's law, which we explore in the next subsection.

Note on independent variables:

In the partial derivative $\partial L / \partial \dot{q}_k$, only $\dot{q}_k$ varies while q_i, $\dot{q}_{j \neq k}$ and t are held fixed. The subsequent total derivative, however, follows a trajectory of the system and includes all time-dependent variations:

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial L}{\partial \dot{q}_k}\right) = \sum_i \left(\frac{\partial^2 L}{\partial q_i \, \partial \dot{q}_k}\dot{q}_i + \frac{\partial^2 L}{\partial \dot{q}_i \, \partial \dot{q}_k}\ddot{q}_i\right) + \frac{\partial^2 L}{\partial t \, \partial \dot{q}_k}.$$

In deriving the Euler-Lagrange equations, we regard the q_k and $\dot{q}_k$ as independent arguments of the Lagrangian $L(q_i, \dot{q}_i, t)$ when taking partial derivatives. This does not contradict the relation $\dot{q}_k = \mathrm{d}q_k/\mathrm{d}t$, which holds along any trajectory. This relation defines $\dot{q}_k$ in terms of the trajectory $q_k(t)$, but does not determine $\dot{q}_k$ from the instantaneous value of q_k. Different trajectories can pass through the same configuration with different velocities, so specifying the configuration alone is not sufficient to specify the instantaneous state of motion, which is why L depends on both q_k and $\dot{q}_k$.
A particular relation between q_k and $\dot{q}_k$ that follows from the equations of motion must not be substituted into the Lagrangian before deriving those equations, since such a relation is an output of the dynamics rather than an input to the variational problem. Although q_k and $\dot{q}_k$ are treated as independent arguments when taking partial derivatives of L, their variations along a varied trajectory are related, as we will make clear in section 1.2.4.

1.2.4 The principle of least action

The Euler-Lagrange equations can also be derived from Hamilton's principle, commonly referred to as the principle of *least action*[6]. This principle states that a system evolves or moves from point A (characterised by a set of values for q_i) to point B (characterised by another set of values for q_i) by following the trajectory that makes the action stationary (typically a minimum). The action, denoted by S, is defined as:

$$S = \int_{t_A}^{t_B} L\left[q_i\left(t\right), \dot{q}_i\left(t\right), t\right] \mathrm{d}t, \tag{1.24}$$

where L is the Lagrangian and t_A and t_B are the times when the system is at points A and B, respectively. Hamilton's principle applies to systems where non-constraint forces are either conservative or derivable from a generalised potential that may depend on the velocities.

We assume that the constraints are holonomic, so the generalised coordinates q_i are independent. If the path $q_i(t)$ makes S stationary, introducing an arbitrary infinitesimal variation $\delta q_i(t)$ at all times between t_A and t_B, while keeping q_i fixed at these endpoints, will not change S. In other words, $\delta S = 0$.

It is important to note that these variations in $q_i(t)$ span the entire time interval between t_A and t_B, and affect the entire trajectory of the system. This differs from the virtual displacements discussed earlier, which were instantaneous changes at a fixed time. Here, a variation $\delta q_i(t)$ in $q_i(t)$ induces a corresponding variation $\delta\dot{q}_i$ in $\dot{q}_i(t)$. By definition, the generalised velocity along the displaced trajectory $q_i + \delta q_i$ is given by $\dot{q}_i + \delta\dot{q}_i = \mathrm{d}\left(q_i + \delta q_i\right)/\mathrm{d}t$, which implies that $\delta\dot{q}_i = \mathrm{d}\left(\delta q_i\right)/\mathrm{d}t$.

[6]For an in-depth discussion of this principle, see Richard P. Feynman, Robert B. Leighton, Matthew Sands, 1963, *The Feynman Lectures on Physics – Volume 2*, Chap. 19 *The Principle of Least Action*, Addison-Wesley.

Note also that these variations are not constrained by the system's dynamics. Since the exact trajectory is unknown, we examine all nearby paths to determine which path makes the action S stationary.

The variation in the action is given by:

$$\delta S = \int_{t_A}^{t_B} L\left(q_i + \delta q_i, \dot{q}_i + \delta \dot{q}_i, t\right) \mathrm{d}t - \int_{t_A}^{t_B} L\left(q_i, \dot{q}_i, t\right) \mathrm{d}t. \tag{1.25}$$

Expanding this to first order using a Taylor expansion, we obtain:

$$\delta S = \int_{t_A}^{t_B} \sum_i \left(\frac{\partial L}{\partial q_i} \delta q_i + \frac{\partial L}{\partial \dot{q}_i} \delta \dot{q}_i \right) \mathrm{d}t, \tag{1.26}$$

where the derivatives are evaluated at $(q_i, \dot{q}_i, t)$. Importantly, the partial derivatives $\partial L/\partial q_i$ and $\partial L/\partial \dot{q}_i$ are taken with L regarded as a function of independent arguments q_i, $\dot{q}_i$ and t: $\partial L/\partial q_i$ accounts only for the explicit dependence of L on q_i, without considering the induced change in $\dot{q}_i$ (and similarly for $\partial L/\partial \dot{q}_i$). This is standard practice in variational calculus. The trajectory itself, however, is varied only through $\delta q_i(t)$, which then determines $\delta \dot{q}_i = \mathrm{d}(\delta q_i)/\mathrm{d}t$: the variations δq_i and $\delta \dot{q}_i$ are not independent.

The next step is to use this relationship between δq_i and $\delta \dot{q}_i$ by integrating by parts the term in equation (1.26) involving $\delta \dot{q}_i$:

$$\int_{t_A}^{t_B} \frac{\partial L}{\partial \dot{q}_i} \delta \dot{q}_i \mathrm{d}t = \int_{t_A}^{t_B} \frac{\partial L}{\partial \dot{q}_i} \frac{\mathrm{d}\left(\delta q_i\right)}{\mathrm{d}t} \mathrm{d}t = \left[\delta q_i \frac{\partial L}{\partial \dot{q}_i} \right]_{t_A}^{t_B} - \int_{t_A}^{t_B} \frac{\mathrm{d}}{\mathrm{d}t} \left(\frac{\partial L}{\partial \dot{q}_i} \right) \delta q_i \mathrm{d}t. \tag{1.27}$$

The boundary term in the square brackets is zero since $\delta q_i = 0$ at both $t = t_A$ and t_B. Substituting this into equation (1.26), we obtain:

$$\delta S = \int_{t_A}^{t_B} \sum_i \left[\frac{\partial L}{\partial q_i} - \frac{\mathrm{d}}{\mathrm{d}t} \left(\frac{\partial L}{\partial \dot{q}_i} \right) \right] \delta q_i(t) \mathrm{d}t, \tag{1.28}$$

where we now indicate explicitly the time dependence of δq_i. Since the q_i are independent variables, the variations δq_i are also independent. Therefore, for $\delta S = 0$ to hold for any arbitrary variations δq_i, each term within the integrand must vanish separately[7], resulting in the Euler-Lagrange equations (1.20).

We have shown that Hamilton's principle leads to the Euler-Lagrange equations, and it is evident that the reverse is also true. These equations are quite general. In their standard form, as given by equations (1.20), they are valid when the constraints are holonomic and the non-constraint forces can be derived from a potential or generalised potential.

[7] This follows from the *fundamental lemma of the calculus of variations*, which states that if

$$\int_a^b f(x)\eta(x)\mathrm{d}x = 0$$

for all arbitrary functions $\eta(x)$ with continuous first and second derivatives on $[a, b]$, then $f(x) = 0$ on $[a, b]$.

However, the applicability of the Euler-Lagrange equations extends beyond such idealised cases. Even when the forces acting on a system are not derivable from a potential, such as friction or other dissipative forces, the Euler-Lagrange equations in the form of equations (1.18) still hold. In these situations, the generalised forces $\mathcal{F}_k$ can be included explicitly to account for non-conservative effects, allowing the equations to remain valid without requiring a potential function.
Moreover, Hamilton's principle, and consequently the Euler-Lagrange equations, can be extended to handle certain types of nonholonomic constraints, though this is generally not very practical.

Deriving the Euler-Lagrange equations from d'Alembert's principle showed that, in classical mechanics, the Lagrangian takes the form $T - V$. However, as shown in this section, these equations can also be derived directly from Hamilton's principle without specifying a particular form of the Lagrangian. Hamilton's principle is more fundamental and provides a unified framework for deriving the equations of motion for a wide variety of physical systems, encompassing both classical and quantum mechanics, as well as relativity. It also extends to fields in addition to particles.

The Euler-Lagrange equations therefore remain valid in the context of relativity, provided the action is Lorentz invariant, so that a trajectory that makes it stationary in one inertial frame makes it stationary in all inertial frames (see section 4.2.1).

The Lagrangian defined by the principle of least action is not uniquely determined. Indeed, if L is a function that renders the action stationary, then:

$$\tilde{L} = L + \frac{\mathrm{d}f}{\mathrm{d}t}, \tag{1.29}$$

where f is a function of q_i and t, will also make the action stationary.
To see this, define:

$$\tilde{S} = \int_{t_A}^{t_B} \tilde{L}\left[q_i(t), \dot{q}_i(t), t\right] \mathrm{d}t = \int_{t_A}^{t_B} L\left[q_i(t), \dot{q}_i(t), t\right] \mathrm{d}t + \int_{t_A}^{t_B} \frac{\mathrm{d}f}{\mathrm{d}t} \mathrm{d}t, \tag{1.30}$$

which yields:

$$\tilde{S} = S + f\left(t_B\right) - f\left(t_A\right). \tag{1.31}$$

Variations $\delta q_i(t)$ that hold q_i fixed at t_A and t_B yield $\delta\tilde{S} = \delta S$, since the additional boundary term contributes nothing to the variation. Therefore, a path $q_i(t)$ for which $\delta S = 0$ also satisfies $\delta\tilde{S} = 0$, implying that both L and $\tilde{L}$ are valid Lagrangians.

1.2.5 Lagrangian for an electromagnetic force

To apply the Euler-Lagrange equations in the presence of an electromagnetic field, we need to construct a Lagrangian $L = T - \mathcal{V}$ where the generalised potential $\mathcal{V}$ accounts for the electromagnetic force. For a charge q moving with velocity $\dot{\mathbf{r}}$ in an electric field $\mathbf{E}$ and a magnetic field $\mathbf{B}$, the force is given by:

$$\mathbf{f} = q\left(\mathbf{E} + \dot{\mathbf{r}} \times \mathbf{B}\right). \tag{1.32}$$

Denoting ϕ as the electric scalar potential and $\mathbf{A}$ as the magnetic vector potential, we have:

$$\mathbf{E} = -\boldsymbol{\nabla}\phi - \frac{\partial \mathbf{A}}{\partial t} \quad \text{and} \quad \mathbf{B} = \boldsymbol{\nabla} \times \mathbf{A}. \tag{1.33}$$

This leads to the expression for the force:

$$\mathbf{f} = q\left[-\boldsymbol{\nabla}\phi - \frac{\partial \mathbf{A}}{\partial t} + \dot{\mathbf{r}} \times \left(\boldsymbol{\nabla} \times \mathbf{A}\right)\right]. \tag{1.34}$$

Using the vector identity $\dot{\mathbf{r}} \times \left(\boldsymbol{\nabla} \times \mathbf{A}\right) = \boldsymbol{\nabla}\left(\dot{\mathbf{r}} \cdot \mathbf{A}\right) - \left(\dot{\mathbf{r}} \cdot \boldsymbol{\nabla}\right)\mathbf{A}$, we observe that:

$$\frac{\partial \mathbf{A}}{\partial t} + \left(\dot{\mathbf{r}} \cdot \boldsymbol{\nabla}\right)\mathbf{A} = \frac{\partial \mathbf{A}}{\partial t} + \frac{\mathrm{d}x_i}{\mathrm{d}t}\frac{\partial \mathbf{A}}{\partial x_i} = \frac{\mathrm{d}\mathbf{A}}{\mathrm{d}t}.$$

This is the total time derivative of $\mathbf{A}$ along the particle's trajectory. Equation (1.34) can then be rewritten as:

$$\mathbf{f} = q\left[-\boldsymbol{\nabla}\left(\phi - \dot{\mathbf{r}} \cdot \mathbf{A}\right) - \frac{\mathrm{d}\mathbf{A}}{\mathrm{d}t}\right].$$

Since the electromagnetic field, and therefore ϕ and $\mathbf{A}$, are independent of the particle's velocity, the x-component of the force can be expressed as:

$$f_x = q\left\{-\frac{\partial}{\partial x}\left(\phi - \dot{\mathbf{r}} \cdot \mathbf{A}\right) + \frac{\mathrm{d}}{\mathrm{d}t}\left[\frac{\partial}{\partial \dot{x}}\left(\phi - \dot{x}A_x\right)\right]\right\},$$

where $\partial\phi/\partial\dot{x} = \partial A_x/\partial\dot{x} = 0$, and similarly for the y and z-components. Using Cartesian coordinates, with $L = T - \mathcal{V}$ and $T = m\left(\dot{x}^2 + \dot{y}^2 + \dot{z}^2\right)/2$, the Euler-Lagrange equations (1.20) yield:

$$m\ddot{x} = -\frac{\partial \mathcal{V}}{\partial x} + \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial \mathcal{V}}{\partial \dot{x}}\right).$$

Since this is equal to f_x, and similarly for f_y and f_z, we find the generalised potential:

$$\mathcal{V} = q\left(\phi - \dot{\mathbf{r}} \cdot \mathbf{A}\right). \tag{1.35}$$

Thus, the Lagrangian for a single charged particle in an electromagnetic field is given by:

$$\boxed{L = \frac{1}{2}m\dot{\mathbf{r}}^2 - q\left(\phi - \dot{\mathbf{r}} \cdot \mathbf{A}\right).} \tag{1.36}$$

1.2.6 Field theory for continuous systems

So far, we have focused on the Lagrangian for particles, where motion is described by a discrete set of coordinates, $q_i(t)$. We now extend this formalism to continuous systems, where a mass element at position x undergoes a displacement $\phi(x,t)$ at time t. Instead of discrete coordinates $q_i(t)$, we work with a *field* $\phi(x,t)$. Examples of such systems include a stretched string or a fluid through which a sound wave propagates. Our objective is to derive an action that leads to the Euler-Lagrange equations which, in these cases, correspond to wave equations. This approach also applies to systems governed entirely by fields, without the presence of particles.

Lagrangian formulation for continuous fields

Consider a one-dimensional crystal composed of N atoms, each with the same mass m, arranged along the x-axis. At equilibrium, the i-th atom is at position x_i, and we denote its displacement from equilibrium by $\phi_i(t)$. We assume that each atom interacts only with its nearest neighbours, so the total interaction potential, U_{total}, is the sum of the interaction potentials between successive atoms, which depend on the distance between them:

$$U_{\text{total}} = \sum_{i=1}^{N-1} U\left(x_{i+1} + \phi_{i+1} - x_i - \phi_i\right).$$

We allow for an external potential V that depends on the displacement of the atoms.
The Lagrangian for this system is then given by:

$$L = \sum_{i=1}^{N} \left[\frac{1}{2} m \dot{\phi}_i^2 - V\left(\phi_i\right)\right] - U_{\text{total}}.$$

Let a denote the equilibrium separation between atoms, so that $x_i = ia$. For small displacements, we expand the interaction potential in a Taylor series:

$$U\left(x_{i+1} + \phi_{i+1} - x_i - \phi_i\right) = U\left(a + \phi_{i+1} - \phi_i\right) \simeq \frac{1}{2} k \left(\phi_{i+1} - \phi_i\right)^2,$$

where we have used the fact that U has a minimum at a, and we set this minimum to zero. Here, k is the second derivative of the potential evaluated at a, and is constant along the chain. The Lagrangian then becomes:

$$L = \sum_{i=1}^{N} \left[\frac{1}{2} m \dot{\phi}_i^2 - V\left(\phi_i\right)\right] - \sum_{i=1}^{N-1} \frac{1}{2} k \left(\phi_{i+1} - \phi_i\right)^2. \tag{1.37}$$

This expression describes interactions between nearest neighbours mediated by elastic forces, effectively, springs with spring constant k.

Now, considering the case where a is small, we replace the discrete displacements $\phi_i(t)$ with a continuous field $\phi\left(x,t\right)$. To first order in a, we approximate:

$$\phi_{i+1} - \phi_i = \phi\left(x_i + a\right) - \phi\left(x_i\right) \simeq a \frac{\partial \phi}{\partial x}\left(x_i\right).$$

The terms in the last sum of equation (1.37) then become:

$$\frac{1}{2}ka^2\left(\frac{\partial\phi}{\partial x}\right)^2,$$

where the derivative is evaluated at the position of the i–th atom. This can be interpreted as the potential energy of a string segment of length a, stretched by the relative displacement $a\partial\phi/\partial x$, suggesting that $ka = T$ serves as the effective tension in the string.

In the continuous limit $N \to +\infty$, where $a = \mathrm{d}x$, we replace m with ρa, where ρ is the mass per unit length. The discrete sums in equation (1.37) then become integrals over the length of the chain:

$$L = \int_0^l \left[\frac{1}{2}\rho \mathrm{d}x\left(\frac{\partial\phi}{\partial t}\right)^2 - \frac{1}{2}T\mathrm{d}x\left(\frac{\partial\phi}{\partial x}\right)^2 - \mathrm{d}x\mathcal{V}(\phi)\right], \tag{1.38}$$

where $l = Na$ is the total length of the crystal, and $\mathcal{V}$ is the *potential density* defined by:

$$\sum_{i=1}^{N} V(\phi_i) = \int_0^l \mathrm{d}x\mathcal{V}(\phi).$$

This derivation can be naturally extended to three dimensions, leading to the Lagrangian:

$$\boxed{L = \int \mathrm{d}^3x\ \mathcal{L}\left(\phi, \boldsymbol{\nabla}\phi, \frac{\partial\phi}{\partial t}\right),} \tag{1.39}$$

where the integral is over the volume of the system and $\mathcal{L}$ is the *Lagrangian density*. It is given by:

$$\mathcal{L} = \frac{1}{2}\rho\left(\frac{\partial\phi}{\partial t}\right)^2 - \frac{1}{2}T(\boldsymbol{\nabla}\phi)^2 - \mathcal{V}(\phi), \tag{1.40}$$

where ρ is now the mass per unit volume, T is a constant characterising the elastic restoring force (with units of energy per unit volume, i.e. of pressure) and $\mathcal{V}$ is the potential energy per unit volume.

Euler-Lagrange equations for fields

We now impose the condition that the action S is stationary, with S given by:

$$S = \int_{t_A}^{t_B} \mathrm{d}t \int \mathrm{d}^3x\ \mathcal{L}\big(\phi, \boldsymbol{\nabla}\phi, \partial_t\phi\big), \tag{1.41}$$

where $\partial_t = \partial/\partial t$. Following a similar approach as in the particle case, but noting that the Lagrangian now depends on ϕ, $\boldsymbol{\nabla}\phi$ and $\partial_t\phi$, we consider variations of the field ϕ. If $\phi(\mathbf{r}, t)$ is the field that makes S stationary, then any small variation $\delta\phi$, applied throughout the volume and at all times between t_A and t_B, while keeping ϕ fixed at the boundaries and at the times t_A and t_B, will leave S unchanged.

This means $\delta S = 0$, with:

$$\delta S = \int_{t_A}^{t_B} \mathrm{d}t \int \mathrm{d}^3x \left[\mathcal{L}\Big(\phi + \delta\phi, \boldsymbol{\nabla}\phi + \delta\left(\boldsymbol{\nabla}\phi\right), \partial_t\phi + \delta\left(\partial_t\phi\right)\Big) - \mathcal{L}\Big(\phi, \boldsymbol{\nabla}\phi, \partial_t\phi\Big) \right]. \quad (1.42)$$

To first order, this simplifies to:

$$\delta S = \int_{t_A}^{t_B} \mathrm{d}t \int \mathrm{d}^3x \left(\frac{\partial\mathcal{L}}{\partial\phi}\delta\phi + \frac{\partial\mathcal{L}}{\partial(\partial_t\phi)}\delta(\partial_t\phi) + \frac{\partial\mathcal{L}}{\partial(\partial_i\phi)}\delta(\partial_i\phi) \right), \quad (1.43)$$

where $i = 1 \ldots 3$ and $\partial_{1,2,3} \equiv \partial_{x,y,z}$. Here, $\delta(\partial_t\phi)$ and $\delta(\partial_i\phi)$ represent the changes in $\partial_t\phi$ and $\partial_i\phi$ resulting from the variations in ϕ. Using the relations $\delta(\partial_t\phi) = \partial_t\left(\delta\phi\right)$ and $\delta(\partial_i\phi) = \partial_i\left(\delta\phi\right)$ and applying integration by parts, we obtain:

$$\int_{t_A}^{t_B} \mathrm{d}t \, \frac{\partial\mathcal{L}}{\partial(\partial_t\phi)}\,\delta(\partial_t\phi) \;=\; -\int_{t_A}^{t_B} \mathrm{d}t \, \partial_t\left(\frac{\partial\mathcal{L}}{\partial(\partial_t\phi)}\right)\delta\phi, \quad (1.44)$$

$$\int \mathrm{d}x_i \, \frac{\partial\mathcal{L}}{\partial(\partial_i\phi)}\,\delta(\partial_i\phi) \;=\; -\int \mathrm{d}x_i \, \partial_i\left(\frac{\partial\mathcal{L}}{\partial(\partial_i\phi)}\right)\delta\phi. \quad (1.45)$$

The boundary terms vanish because $\delta\phi$ is zero at times t_A and t_B as well as on the surface enclosing the system.

Note on partial derivatives: The term $\partial_t\left(\ldots\right)$ on the right-hand side of equation (1.44) involves taking the derivative with respect to time while holding x, y and z constant, but it accounts for the time dependence of the field and its derivatives in $\mathcal{L}$ *via* the chain rule. Thus, it is not a total time derivative, since it does not account for explicit time dependence in the coordinates, but it is also not strictly a partial derivative, as it includes implicit time dependence through the field. If we were to 'unpack' ϕ and make $\mathcal{L}$ an explicit function of x, y, z and t, instead of treating it as a function of the field and its derivatives, then integration by parts above would yield the conventional partial time derivative. The same reasoning applies to ∂_i. Throughout this course, we will see many instances of such 'partial' derivatives, and the context will typically clarify which derivatives are being used.

Substituting equations (1.44) and (1.45) into equation (1.43) yields:

$$\delta S = \int_{t_A}^{t_B} \mathrm{d}t \int \mathrm{d}^3x \left[\frac{\partial\mathcal{L}}{\partial\phi} - \partial_t\left(\frac{\partial\mathcal{L}}{\partial(\partial_t\phi)}\right) - \partial_i\left(\frac{\partial\mathcal{L}}{\partial(\partial_i\phi)}\right) \right]\delta\phi.$$

Since $\delta S = 0$ must hold for any arbitrary variation $\delta\phi$, the expression in brackets within the integrand must vanish, yielding the Euler-Lagrange equation for the field:

$$\boxed{\partial_t\left(\frac{\partial\mathcal{L}}{\partial(\partial_t\phi)}\right) + \partial_i\left(\frac{\partial\mathcal{L}}{\partial(\partial_i\phi)}\right) - \frac{\partial\mathcal{L}}{\partial\phi} = 0.} \quad (1.46)$$

In this case, only one equation arises because there is a single field being varied. In more general cases, the Lagrangian density may depend on multiple fields or on a vector field with three spatial components, each varied independently, leading to a set of Euler-Lagrange equations.

For the crystal discussed earlier, the Lagrangian density is given by equation (1.40), which yields:

$$\frac{\partial \mathcal{L}}{\partial(\partial_t \phi)} = \rho \partial_t \phi, \quad \frac{\partial \mathcal{L}}{\partial(\partial_i \phi)} = -T \partial_i \phi.$$

Ignoring the external potential, the Euler-Lagrange equation becomes:

$$\rho \frac{\partial^2 \phi}{\partial t^2} - T \boldsymbol{\nabla}^2 \phi = 0,$$

which is a wave equation describing the propagation of a displacement ϕ within the crystal, with a wave speed $\sqrt{T/\rho}$.

Lagrangian for the electromagnetic field

Previously, starting from the Lagrangian for a single particle, we derived a Lagrangian for a system of particles whose motion is described by a field $\phi(\mathbf{r}, t)$, with the principle of least action leading to a wave equation for this field. This approach can be generalised to apply to fields in general, whether or not particles are present.
Consider, for example, the electromagnetic field $(\mathbf{E}, \mathbf{B})$ in the absence of charges and currents. Our aim is to find a Lagrangian density such that applying the principle of least action yields Maxwell's equations in vacuum. In Problem Set 1, we will verify that the Lagrangian density is given by:

$$\mathcal{L} = \frac{\epsilon_0}{2} \mathbf{E}^2 - \frac{1}{2\mu_0} \mathbf{B}^2. \tag{1.47}$$

By expressing the electric and magnetic fields in terms of the scalar potential ϕ and the vector potential $\mathbf{A}$, we can rewrite $\mathcal{L}$ as a function of ϕ, $\mathbf{A}$ and their derivatives.

In the presence of charges and currents, an interaction term must be added to this Lagrangian density. It can be obtained from the Lagrangian for a charged particle derived in section 1.2.5. From equation (1.36), the part of the Lagrangian of a particle of charge q that couples it to the electromagnetic field is $L_{\text{int}} = -q\phi(\mathbf{r}, t) + q\dot{\mathbf{r}} \cdot \mathbf{A}(\mathbf{r}, t)$, where ϕ and $\mathbf{A}$ are evaluated at the position $\mathbf{r}$ of the particle.
For a system of N particles with charges q_k, positions $\mathbf{r}_k$ and velocities $\dot{\mathbf{r}}_k$, the interaction terms simply add:

$$L_{\text{int}} = \sum_{k=1}^{N} \left[-q_k \phi(\mathbf{r}_k, t) + q_k \dot{\mathbf{r}}_k \cdot \mathbf{A}(\mathbf{r}_k, t) \right]. \tag{1.48}$$

As for the one-dimensional crystal, we now go to the continuum limit. Consider a continuous distribution of charge with charge density $\rho(\mathbf{r}, t)$, moving with velocity $\mathbf{v}(\mathbf{r}, t)$. A small volume element d^3x located at $\mathbf{r}$ carries a charge $\rho \mathrm{d}^3x$ moving with velocity $\mathbf{v}$, and therefore contributes $-\rho\phi\,\mathrm{d}^3x + \rho\mathbf{v}\cdot\mathbf{A}\,\mathrm{d}^3x$ to the interaction Lagrangian. Introducing the current density $\mathbf{j} = \rho\,\mathbf{v}$, the sum over particles becomes an integral over space:

$$L_{\text{int}} = \int d^3x \left(-\rho\,\phi + \mathbf{j}\cdot\mathbf{A}\right), \tag{1.49}$$

so that the interaction Lagrangian density is:

$$\mathcal{L}_{\text{int}} = -\rho\,\phi + \mathbf{j}\cdot\mathbf{A}. \tag{1.50}$$

The Lagrangian density of the electromagnetic field in the presence of charges and currents is therefore:

$$\boxed{\mathcal{L} = \frac{\epsilon_0}{2}E^2 - \frac{1}{2\mu_0}B^2 - \rho\,\phi + \mathbf{j}\cdot\mathbf{A}.} \tag{1.51}$$

If ρ and $\mathbf{j}$ are treated as prescribed, i.e. unaffected by variations of ϕ and $\mathbf{A}$, applying the principle of least action to this Lagrangian density yields Maxwell's equations with sources. This is verified in Problem Set 1, where it is also shown that consistency of these equations with gauge invariance requires the charge and current densities to satisfy the continuity equation $\partial_t\rho + \boldsymbol{\nabla}\cdot\mathbf{j} = 0$, i.e. conservation of charge.

1.3 Conservation laws and Noether's theorem

We now return to the topic of symmetries.

1.3.1 Cyclic coordinates

Consider a point mass m subject to a conservative force, with no constraints. Using Cartesian coordinates $q_1 = x$, $q_2 = y$ and $q_3 = z$, the Lagrangian is given by:

$$L = \frac{1}{2}m\left(\dot{x}^2 + \dot{y}^2 + \dot{z}^2\right) - V\left(x, y, z\right). \tag{1.52}$$

If the system is invariant under translation along the x-axis, then the potential V cannot depend on x. From equation (1.20), this implies that $\partial L/\partial\dot{x}$, which is equal to the x-component of the momentum $p_x = m\dot{x}$, is constant.

Now, consider motion in a plane using polar coordinates $q_1 = r$ and $q_2 = \theta$. The Lagrangian becomes:

$$L = \frac{1}{2}m\left(\dot{r}^2 + r^2\dot{\theta}^2\right) - V\left(r, \theta\right). \tag{1.53}$$

If the system is invariant under rotation about the vertical axis through the origin, then V cannot depend on θ. From equation (1.20), this implies that $\partial L/\partial\dot{\theta}$, which is equal to the z-component of the angular momentum $J_z = mr^2\dot{\theta}$, is constant.

These two examples show that the Euler-Lagrange equations allow us to easily recover results that were more laboriously derived using Newton's law of motion in sections 1.1.1 and 1.1.2.

More generally, if the Lagrangian of a system does not depend on the coordinate q_k, then the system is invariant, or symmetric, under a transformation that changes q_k. Such a coordinate is referred to as a *cyclic coordinate*. In this case, equation (1.20) implies that:

$$p_k \equiv \frac{\partial L}{\partial \dot{q}_k}, \tag{1.54}$$

is a constant. We refer to p_k as the *conjugate momentum* or *canonical momentum*, although it does not necessarily have the units of momentum.

Conserved momentum when a magnetic field is present

For a particle subject to an electromagnetic force, the Lagrangian is given by equation (1.36). The conjugate momentum is therefore $\mathbf{p} = m\dot{\mathbf{r}} + q\mathbf{A}$, which differs from the mechanical momentum $\mathbf{p}_{\text{mec}} = m\dot{\mathbf{r}}$. If ϕ and $\mathbf{A}$ do not depend on, say, x, then x is a cyclic coordinate and $p_x = m\dot{x} + qA_x$ is conserved, whereas $m\dot{x}$ is in general not, since the electromagnetic force changes it. The additional term $q\mathbf{A}$ reflects the fact that the electromagnetic field itself can carry momentum, which can be exchanged with the particle.

The role of the field momentum becomes clearer when two charges interact with each other. Consider two like charges moving at right angles to each other. The electric forces between them are repulsive, equal and opposite, but the magnetic forces they exert on each other are not equal and opposite[8].

In section 1.1.1, we showed that the total mechanical momentum of two particles subject only to conservative interaction forces was conserved, relying on the law of action and reaction. When magnetic forces are present, this is no longer the case. For each particle, Newton's second law still applies and gives $\mathbf{f} = \mathrm{d}\mathbf{p}_{\text{mec}}/\mathrm{d}t$, but since the forces are not equal and opposite, the sum of the mechanical momenta is not conserved: the momentum carried by the field must be included. Note that the Lagrangian (1.36) describes a particle moving in prescribed external fields, and does not apply directly to this situation, where each particle moves in the field produced by the other. A Lagrangian describing the mutual interaction of the charges can be constructed (to lowest order in v/c, this is the Darwin Lagrangian), and the conserved quantity is then the sum of the canonical momenta of the particles, which includes the momentum carried by the field.

[8] For a detailed discussion, see David Griffiths, *Introduction to Electrodynamics*, section 8.2.

To restore the concept of action and reaction, we need to reinterpret it in terms of momentum exchange rather than direct forces or consider the forces as being mediated by the field. When a charged particle moves, it generates a changing magnetic field, which can exert forces on other charged particles. Conversely, changes in the motion of other charged particles affect the magnetic field, which in turn affects the original particle. This entire interaction is mediated by the field. For the system as a whole (particles + fields), total momentum is conserved. However, a particle does not directly exert a force on another particle. Instead, particles transfer momentum to the fields, altering the field configuration and momentum, and the fields transfer momentum back to the particles, exerting forces on them. This process is more accurately described within the framework of special relativity, which will be discussed later in these notes.

1.3.2 Noether's theorem

In some cases, symmetries correspond not just to a particular generalised coordinate remaining constant, but to a combination of generalised coordinates and velocities being conserved. Noether's theorem states that, *if a system exhibits a continuous symmetry, there is an associated conserved quantity.* In this section, we will prove Noether's theorem and derive the corresponding constant of the motion.

A continuous transformation of a system is considered a symmetry if it leaves the equations of motion unchanged. For this to hold, the action before the transformation, S, and the action after the transformation, S', must yield the same equations of motion, i.e. must be stationary for the same trajectories. This is the case, in particular, if S and S' differ only by a boundary term, as discussed in section 1.2.4. In this course, we will primarily focus on transformations that leave the action unchanged, meaning $S' = S$. Under these conditions, we will now investigate which quantities must remain constant for S to be left unchanged when such a transformation is performed.

A finite continuous transformation can be viewed as the result of a sequence of infinitesimal changes. We denote by ϵ the parameter that characterises an infinitesimal continuous transformation, and define the transformations of time and the generalised coordinates as follows:

$$t \quad \to \quad t' = t + \epsilon\tau\left(q_k, t\right), \tag{1.55}$$

$$q_i(t) \quad \to \quad q_i'(t') = q_i(t) + \epsilon\eta_i\left(q_k, t\right), \tag{1.56}$$

where τ and η_i may depend on all the q_k and time t. The dependence of t' on the coordinates q_k is particularly important for applications in relativity. We will also need the change in q_i at fixed time, which we denote $\bar{\eta}_i$:

$$q_i'(t) = q_i(t) + \epsilon\bar{\eta}_i\left(q_k, t\right). \tag{1.57}$$

From equation (1.56), we have $q_i'(t') \equiv q_i'(t+\epsilon\tau) = q_i(t) + \epsilon\eta_i\left(q_k, t\right)$. From equation (1.57), we have $q_i'(t+\epsilon\tau) = q_i(t+\epsilon\tau) + \epsilon\bar{\eta}_i\left(q_k, t+\epsilon\tau\right)$. Comparing these two expressions, we obtain, to first order in ϵ:

$$\bar{\eta}_i\left(q_k, t\right) = \eta_i\left(q_k, t\right) - \tau\dot{q}_i(t). \tag{1.58}$$

For a time translation, ϵ represents the time interval by which t is shifted, with $\tau = 1$ and $\eta_i = 0$ when the spatial coordinates remain unchanged. In this case, $q_i'(t') = q_i(t)$ and $q_i'(t) = q_i(t) - \epsilon\dot{q}_i$, which can also be expressed as $q_i'(t) = q_i(t-\epsilon)$. This means that a time translation by ϵ replaces each coordinate $q_i(t)$ with $q_i(t-\epsilon)$.
For rotations, ϵ corresponds to the angle of rotation θ, while for spatial translations, it represents the distance by which the positions are shifted. In both these cases, $\tau = 0$.

The action between arbitrary times t_A and t_B is given by:

$$S = \int_{t_A}^{t_B} L\Big(q_i(t), \dot{q}_i(t), t\Big)\mathrm{d}t. \tag{1.59}$$

After the transformation is applied, this becomes:

$$S' = \int_{t_A+\delta t_A}^{t_B+\delta t_B} L\Big(q_i'(t'), \dot{q}_i'(t'), t'\Big)\mathrm{d}t', \tag{1.60}$$

where $\delta t_A = \epsilon\tau\big(q_k\left(t_A\right), t_A\big)$ and similarly for δt_B.

We will now examine the condition implied by the equality $S = S'$. In the expression for S', t' is a dummy variable and can be relabelled t. Additionally, we separate the integral into three contributions:

$$S' = -\int_{t_A}^{t_A+\delta t_A} L\Big(q_i'(t), \dot{q}_i'(t), t\Big)\mathrm{d}t + \int_{t_A}^{t_B} L\Big(q_i'(t), \dot{q}_i'(t), t\Big)\mathrm{d}t + \int_{t_B}^{t_B+\delta t_B} L\Big(q_i'(t), \dot{q}_i'(t), t\Big)\mathrm{d}t. \tag{1.61}$$

Since δt_A is an infinitesimal quantity, the first integral can be approximated by evaluating the integrand at t_A and multiplying by δt_A. The same approximation applies to the third integral.
This yields:

$$S' - S = L\Big(q_i'(t_B), \dot{q}_i'(t_B), t_B\Big)\delta t_B - L\Big(q_i'(t_A), \dot{q}_i'(t_A), t_A\Big)\delta t_A + \int_{t_A}^{t_B} \delta L\ \mathrm{d}t, \tag{1.62}$$

where:

$$\begin{aligned} \delta L &= L\Big(q_i'(t), \dot{q}_i'(t), t\Big) - L\Big(q_i(t), \dot{q}_i(t), t\Big) \\ &= \sum_i \left[\Big(q_i'(t) - q_i(t)\Big)\frac{\partial L}{\partial q_i} + \Big(\dot{q}_i'(t) - \dot{q}_i(t)\Big)\frac{\partial L}{\partial \dot{q}_i}\right]. \end{aligned} \tag{1.63}$$

Here, the coordinates $q_i(t)$ undergoing the transformation define a trajectory that satisfies the Euler-Lagrange equations (1.20). Using these and equation (1.57), we can rewrite δL as:

$$\delta L = \sum_i \left[\frac{\mathrm{d}}{\mathrm{d}t} \left(\frac{\partial L}{\partial \dot{q}_i} \right) \epsilon \bar{\eta}_i + \frac{\partial L}{\partial \dot{q}_i} \epsilon \dot{\bar{\eta}}_i \right] = \sum_i \frac{\mathrm{d}}{\mathrm{d}t} \left(\frac{\partial L}{\partial \dot{q}_i} \epsilon \bar{\eta}_i \right), \tag{1.64}$$

where the derivatives of L are evaluated at $\big(q_i(t), \dot{q}_i(t), t\big)$.
For the first two terms on the right-hand side of equation (1.62), we can drop the primes to first order in ϵ. This is because these expressions are multiplied by the infinitesimal quantity $\delta t = \epsilon\tau$, and the difference between the primed and unprimed quantities is also first order in ϵ. We then combine these terms as follows:

$$\begin{aligned} L\Big(q_i'(t_B), \dot{q}_i'(t_B), t_B\Big)\delta t_B - L\Big(q_i'(t_A), \dot{q}_i'(t_A), t_A\Big)\delta t_A &= L\Big(q_i(t), \dot{q}_i(t), t\Big)\delta t \,\Big|_{t_A}^{t_B} \\ &= \int_{t_A}^{t_B} \frac{\mathrm{d}}{\mathrm{d}t} \Big(L\Big(q_i(t), \dot{q}_i(t), t\Big)\delta t \Big) \mathrm{d}t. \end{aligned}$$

Using this expression and equation (1.64), equation (1.62) becomes:

$$S' - S = \int_{t_A}^{t_B} \frac{\mathrm{d}}{\mathrm{d}t} \left(\sum_i \frac{\partial L}{\partial \dot{q}_i} \epsilon \bar{\eta}_i + L\Big(q_i(t), \dot{q}_i(t), t\Big)\delta t \right) \mathrm{d}t. \tag{1.65}$$

For the condition $S' - S = 0$ to be satisfied for arbitrary times t_A and t_B, the integrand must be zero. This implies that the function inside the time derivative must remain constant over time:

$$\sum_i \frac{\partial L}{\partial \dot{q}_i} \bar{\eta}_i + L\tau = \text{const.} \tag{1.66}$$

Using equation (1.58), this can also be written as:

$$\boxed{\left(\sum_i \frac{\partial L}{\partial \dot{q}_i} \dot{q}_i - L \right) \tau - \sum_i \frac{\partial L}{\partial \dot{q}_i} \eta_i = \text{const},} \tag{1.67}$$

which is a form of Noether's theorem. Since this theorem is derived without assuming a specific form for the Lagrangian, **it remains applicable in relativity**, provided the coordinate transformations preserve the action.

Noether's theorem applies not only to mechanics but also to field theory, where the symmetries are related to fields rather than just coordinates. This concept will be further explored in Problem Set 1.

1.3.3 Applications

Translation in space

We first revisit some of the conservation laws obtained in section 1.3.1, now as illustrations of Noether's theorem.

Consider a point mass m moving in a potential V. In Cartesian coordinates, with $q_1 = x$, $q_2 = y$ and $q_3 = z$, the Lagrangian is given by equation (1.52). A translation along the x-axis corresponds to $\tau = 0$, $\eta_1 = 1$ and $\eta_2 = \eta_3 = 0$. If the potential is invariant under this transformation, the action remains unchanged and Noether's theorem (1.67) applies. This gives $\partial L/\partial \dot{x} = m\dot{x} = \text{const}$, confirming that the momentum along the x-axis is a constant of the motion.

Now, consider N point masses with position vectors $\mathbf{r}_i = (x_i, y_i, z_i)$. We define $q_{3i-2} = x_i$, $q_{3i-1} = y_i$ and $q_{3i} = z_i$ for $i = 1 \ldots N$. We assume that the system is invariant under translation by any vector $\mathbf{c}$. This corresponds to $\tau = 0$, $\eta_{3i-2} = c_x$, $\eta_{3i-1} = c_y$ and $\eta_{3i} = c_z$ for all i. Noether's theorem then yields:

$$\sum_i \frac{\partial L}{\partial \dot{q}_i}\eta_i = c_x \sum_i m_i \dot{x}_i + c_y \sum_i m_i \dot{y}_i + c_z \sum_i m_i \dot{z}_i = \text{const.}$$

Since this must be satisfied for any values of the independent quantities c_x, c_y and c_z, each sum must be constant:

$$\sum_i m_i \dot{x}_i = P_x, \quad \sum_i m_i \dot{y}_i = P_y, \quad \sum_i m_i \dot{z}_i = P_z,$$

where $\mathbf{P} = (P_x, P_y, P_z)$ is the total momentum. Therefore, invariance under any translation implies that the total momentum is conserved.

Rotation

In polar coordinates, with $q_1 = r$, $q_2 = \theta$, the Lagrangian is given by equation (1.53). Consider a rotation around the vertical axis passing through the origin, which corresponds to $\tau = 0$, $\eta_1 = 0$ and $\eta_2 = 1$. If the potential is invariant under this rotation, the action remains unchanged. Noether's theorem then yields $\partial L/\partial \dot{\theta} = mr^2\dot{\theta} = \text{const}$, confirming that the angular momentum about the z-axis is conserved.

Translation in time

For a time translation, $\tau = 1$ and $\eta_i = 0$ for all i. If the system is invariant under time translations, then Noether's theorem yields:

$$\boxed{H \equiv \sum_i \frac{\partial L}{\partial \dot{q}_i}\dot{q}_i - L = \text{const.}} \tag{1.68}$$

This quantity is known as the *Hamiltonian*.

Consider a Lagrangian of the form $L = T - V$, where V does not depend on the velocities, in the particular case where the position vectors $\mathbf{r}_k$ depend on the generalised coordinates q_i but not explicitly on time. We then have:

$$\dot{\mathbf{r}}_k = \sum_i \frac{\partial \mathbf{r}_k}{\partial q_i} \dot{q}_i,$$

so that the kinetic energy T is quadratic in the generalised velocities. Therefore:

$$\sum_i \frac{\partial T}{\partial \dot{q}_i} \dot{q}_i = 2T.$$

Since V does not depend on the velocities, this yields:

$$\boxed{H = T + V.} \tag{1.69}$$

The Hamiltonian represents the total energy, and Noether's theorem then states that this energy is conserved when the system is invariant under time translations.

Time-translation invariance corresponds to the Lagrangian having no explicit time dependence, $\partial L/\partial t = 0$, as can also be seen directly by writing the rate of change of H:

$$\frac{\mathrm{d}H}{\mathrm{d}t} = \sum_i \left[\frac{\partial L}{\partial \dot{q}_i} \ddot{q}_i + \frac{\mathrm{d}}{\mathrm{d}t} \left(\frac{\partial L}{\partial \dot{q}_i} \right) \dot{q}_i \right] - \frac{\mathrm{d}L}{\mathrm{d}t}. \tag{1.70}$$

Expanding the total derivative of L gives:

$$\frac{\mathrm{d}L}{\mathrm{d}t} = \sum_i \left(\frac{\partial L}{\partial q_i} \dot{q}_i + \frac{\partial L}{\partial \dot{q}_i} \ddot{q}_i \right) + \frac{\partial L}{\partial t}. \tag{1.71}$$

Substituting this expression into equation (1.70) yields:

$$\frac{\mathrm{d}H}{\mathrm{d}t} = \sum_i \left[\frac{\mathrm{d}}{\mathrm{d}t} \left(\frac{\partial L}{\partial \dot{q}_i} \right) - \frac{\partial L}{\partial q_i} \right] \dot{q}_i - \frac{\partial L}{\partial t}. \tag{1.72}$$

The terms in the brackets vanish because L satisfies the Euler-Lagrange equations (1.20). This yields:

$$\boxed{\frac{\mathrm{d}H}{\mathrm{d}t} = -\frac{\partial L}{\partial t}.} \tag{1.73}$$

Therefore, if the Lagrangian has no explicit time dependence ($\partial L/\partial t = 0$), the Hamiltonian H is conserved.

For a particle in an electromagnetic field, with the Lagrangian given by equation (1.36), the Hamiltonian is:

$$H = (m\dot{\mathbf{r}} + q\mathbf{A}) \cdot \dot{\mathbf{r}} - L = \frac{1}{2}m\dot{\mathbf{r}}^2 + q\phi. \tag{1.74}$$

This shows that the sum of the kinetic energy $\frac{1}{2}m\dot{\mathbf{r}}^2$ and the potential energy $q\phi$ is constant if the potentials are not explicitly time-dependent. The terms involving the vector potential $\mathbf{A}$ cancel from the Hamiltonian, consistent with the fact that the magnetic force does no work on the particle.

Just as the canonical momentum contains the term $q\mathbf{A}$, the Hamiltonian contains the term $q\phi$. Both describe the interaction between the particle and the field: $q\phi$ contributes to the particle's energy and $q\mathbf{A}$ to its canonical momentum. Neither the Hamiltonian nor the canonical momentum includes the energy or momentum of the external field itself, whose energy density is $\frac{1}{2}\epsilon_0 E^2 + \frac{1}{2\mu_0}B^2$. This is because the Lagrangian (1.36) describes a particle moving in prescribed fields, and does not account for the dynamics of the fields. A complete description of the system requires a Lagrangian that includes the particle, the field and their interaction, as discussed in section 1.2.6. We will also see in the context of relativity that $q\phi$ and $q\mathbf{A}$ combine into a single four-vector, in the same way as energy and momentum combine into the four-momentum.

Chapter 2

Groups and the Lorentz transformation

If two transformations preserve the invariance of a geometrical figure, then applying one transformation followed by the other will also preserve the invariance of the figure. Similarly, two transformations that leave the laws of physics unchanged can be performed successively, resulting in another transformation that maintains the same invariance. Furthermore, every transformation has an inverse that undoes its effect. More generally, a set of transformations associated with symmetries forms a *group*, which is the focus of this chapter[1].

2.1 Introduction to groups

2.1.1 Definition

A group G is a set equipped with an internal operation denoted by $\cdot$ and commonly referred to as multiplication (though it may not be a multiplication in the usual sense). This operation satisfies the following properties:

(a) Closure: $\forall\, g_1, g_2 \in G,\ \ g_1 \cdot g_2 \in G$,

(b) Associativity: $\forall\, g_1, g_2, g_3 \in G,\ \ g_1 \cdot (g_2 \cdot g_3) = (g_1 \cdot g_2) \cdot g_3$,

(c) Identity element: $\exists\, e \in G \,/\, \forall\, g \in G,\ e \cdot g = g \cdot e = g$. This element e is called the *identity*.

(d) Inverse element: $\forall\, g \in G,\ \exists\, h \in G \,/\, g \cdot h = h \cdot g = e$. This element h is called the *inverse* of g and is usually denoted by g^{-1}.

[1]For further reading, refer to the chapter *Groups and Representations* in Prof. Andre Lukas' second-year Mathematical Methods Lecture Notes

The identity element and the inverse of each element are unique. Indeed, if e and e' are both identity elements, then $e = e \cdot e' = e'$. Similarly, if h and h' are both inverses of g, then $h = h \cdot (g \cdot h') = (h \cdot g) \cdot h' = h'$. Using associativity together with properties (c) and (d), we also have, $\forall\, g_1, g_2 \in G$, $\left(g_2^{-1} \cdot g_1^{-1}\right) \cdot (g_1 \cdot g_2) = e$ and $(g_1 \cdot g_2) \cdot \left(g_2^{-1} \cdot g_1^{-1}\right) = e$, which means that $g_2^{-1} \cdot g_1^{-1} = (g_1 \cdot g_2)^{-1}$.

The multiplication is not necessarily commutative. However, if it is, meaning that, $\forall\, g_1, g_2 \in G$, $g_1 \cdot g_2 = g_2 \cdot g_1$, then the group is said to be *commutative* or *abelian*.

A group may have either an infinite number of elements (whether countable or not), or a finite number N of elements. In the latter case, N is referred to as the *order* of the group.

2.1.2 Examples

Groups are of central importance in mathematics and are not always related to transformations. Here are a few examples:

- Group $(\mathbb{Z}, +)$ of all integers (positive, negative and zero) under the operation of addition. This group is *countably infinite*, meaning that while the number of elements is infinite, there exists a bijection between $\mathbb{Z}$ and the set of natural numbers $\mathbb{N} = \{1, 2, 3, \ldots\}$.

- Group $(\mathbb{R}^*, \cdot)$ of non-zero real numbers under the operation of multiplication. This group is *uncountably infinite*.

- *Cyclic* group C_N of the complex numbers $\mathrm{e}^{2\mathrm{i}\pi p/N}$ with $p = 0 \ldots N-1$ under the operation of multiplication. This group is abelian and *finite*, with order N. The element $g = \mathrm{e}^{2\mathrm{i}\pi/N}$ is a *generator* of the group, as all the elements $\left(e = g^0, g, g^2, g^3, \ldots, g^{N-1}\right)$ can be derived from it. More generally, any element $\mathrm{e}^{2\mathrm{i}\pi \mathrm{p}/\mathrm{N}}$ with p coprime to N is also a generator.[2]

- Permutation group S_N of N objects. This group is non-abelian if $N \geq 3$ and is finite, of order $N!$

In this chapter, we will focus on groups whose elements are transformations, which are particularly significant in physics. Among these are the groups of translations $T_{(n)}$ and the groups of rotations $R_{(n)}$ in n-dimensions. The internal operation for these groups is the *composition* of transformations, meaning performing one transformation after another.

[2] This notion of a generator, as an element whose powers produce the whole group, should not be confused with the generators of a Lie group introduced in section 2.3, which are in general not elements of the group.

2.1.3 Groups of matrices

Groups of matrices under multiplication are particularly important in physics because they provide a formal framework for representing transformations, as will be discussed in the next section. Two groups that will be especially useful in this course are $\mathrm{O}(n)$ and $\mathrm{SO}(n)$.

The $\mathbf{O}(n)$ group

Known as the *orthogonal* group, it consists of all $n \times n$ orthogonal matrices. These matrices M are real and satisfy $M^{-1} = M^\top$. We verify that they form a group:

(a) Closure: if M_1 and M_2 are orthogonal, then $(M_1 M_2)^{-1} = M_2^{-1} M_1^{-1} = M_2^\top M_1^\top = (M_1 M_2)^\top$, demonstrating that $M_1 M_2$ is also orthogonal.

(b) Associativity: matrix multiplication is associative in general, so it is associative for orthogonal matrices in particular.

(c) Identity element: the identity matrix I is orthogonal and therefore an element of the group.

(d) Inverse element: if M is an orthogonal matrix, then $M^{-1} = M^\top$. This implies that $\left(M^{-1}\right)^{-1} = \left(M^\top\right)^{-1} = \left(M^{-1}\right)^\top$, using the fact that the transpose and inverse operations can be interchanged for any invertible matrix. This demonstrates that M^{-1} is itself an orthogonal matrix.

Since $\det\left(M^\top\right) = \det(M)$ and an orthogonal matrix M satisfies $M^\top M = I$, we have $\det\left(M^\top M\right) = \det\left(M^\top\right)\det(M) = \det(M)^2 = 1$. Therefore, $\det(M) = \pm 1$.

We now focus on the $n = 2$ case, and express a matrix in $\mathrm{O}(2)$ as:

$$M = \begin{pmatrix} a & b \\ c & d \end{pmatrix}, \quad \text{so that} \quad M^{-1} = \begin{pmatrix} \epsilon d & -\epsilon b \\ -\epsilon c & \epsilon a \end{pmatrix} \tag{2.1}$$

with $\epsilon \equiv \det(M) = ad - bc = \pm 1$. The condition $M^{-1} = M^\top$ implies that $b = -\epsilon c$ and $d = \epsilon a$. Therefore $\det(M) = \epsilon\left(a^2 + c^2\right)$. Since this is equal to ϵ, we have $a^2 + c^2 = 1$, allowing us to define θ such that $a = \cos\theta$ and $c = \epsilon \sin\theta$. This gives:

$$M = \begin{pmatrix} \cos\theta & -\sin\theta \\ \epsilon \sin\theta & \epsilon \cos\theta \end{pmatrix}. \tag{2.2}$$

If $\epsilon = +1$, this matrix represents a counterclockwise rotation by an angle θ around the origin, applied to a vector in Cartesian coordinates. If $\epsilon = -1$, we can write $M = PM'$ with:

$$P = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \text{ and } M' = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}. \tag{2.3}$$

In this case, M represents a rotation by an angle θ around the origin, followed by a reflection about the x-axis.

The SO(n) group

Known as the *special orthogonal* group, it is the subgroup of O(n) which contains all $n \times n$ orthogonal matrices with determinant equal to 1. Once again, we verify that they form a group:

(a) Closure: if $M_1, M_2 \in \mathrm{SO}(n)$, they are in particular orthogonal, and we have already shown that $M_1 M_2$ is then also orthogonal. Furthermore, since their determinants are both equal to 1, $\det(M_1 M_2) = \det(M_1)\det(M_2) = 1$.

(b) Associativity: as in the O(n) group, matrix multiplication is associative.

(c) Identity element: the identity matrix I is orthogonal and has a determinant of 1.

(d) Inverse element: if M is orthogonal, we have already shown that M^{-1} is also orthogonal. Additionally, $\det(M)\det\left(M^{-1}\right) = \det\left(MM^{-1}\right) = \det(I) = 1$, implying that $\det\left(M^{-1}\right) = 1/\det(M)$. Therefore, if the determinant of M is equal to 1, so is the determinant of M^{-1}.

In the case $n = 2$, SO(2) is the subgroup of O(2) for which $\epsilon = +1$. As noted above, and as will be shown in section 2.2.2, the matrices in SO(2) represent rotations in Cartesian coordinates.

2.2 Representations

We mentioned earlier that the matrices in O(2) and SO(2) represent certain transformations in Cartesian coordinates. This leads us to the topic of representations.

2.2.1 Definition of a representation

In simple terms, a representation connects each element of a group to a specific operation that applies to vectors in a space. These operations are consistent with the group's structure and reversible. More formally, a representation of a group G is a way of expressing the elements of the group as linear, invertible operators acting on a vector space. If this vector space has a finite dimension n, then once a basis is chosen, any linear operator can be represented by an $n \times n$ matrix. Consequently, the representation of G can be viewed as a correspondence between each element g of the group and an $n \times n$ matrix, denoted by $M(g)$. Since the operator is invertible, the matrix $M(g)$ must be non-singular, meaning $\det(M(g)) \neq 0$ so that $[M(g)]^{-1}$ exists. Denoting by $\cdot$ the internal operation of G, the representation must also satisfy the following property:

$$\boxed{\forall\, g, g' \;\in\; G,\; M\left(g \cdot g'\right) = M\left(g\right) M\left(g'\right).} \tag{2.4}$$

This ensures that the representation preserves the group structure, meaning the matrices $M(g)$ defined in this way form a group under matrix multiplication.

We can demonstrate this as follows:

(a) Closure: if $M(g)$ and $M(g')$ are the matrices associated with the elements $g \in G$ and $g' \in G$, then the relation above shows that the product $M(g)\,M(g')$ is associated with the element $g \cdot g' \in G$.

(b) Associativity: matrix multiplication is associative in general, so it is associative for these matrices in particular.

(c) Identity element: if e is the identity element of G, then equation (2.4) implies that $M(e)\,M(g) = M(e \cdot g) = M(g)$. Similarly, $M(g)\,M(e) = M(g \cdot e) = M(g)$. Therefore, $M(e)$ is the identity element of the group formed by the matrices $M(g)$. Additionally, since these matrices are invertible, $M(e)\,M(g) = M(g)$ implies that $M(e) = M(g)\,[M(g)]^{-1} = I$. Thus, the matrix associated with e is the identity matrix.

(d) Inverse element: equation (2.4) gives $M(g)\,M\left(g^{-1}\right) = M\left(g \cdot g^{-1}\right) = M(e) = I$. Similarly, $M\left(g^{-1}\right)M(g) = I$. Therefore, the inverse of $M(g)$ is $[M(g)]^{-1} = M\left(g^{-1}\right)$, which is the matrix associated with the element $g^{-1} \in G$.

Note that the correspondence $g \mapsto M(g)$ need not be one-to-one: several elements of G may be associated with the same matrix. For example, setting $M(g) = I$ for every $g \in G$ trivially satisfies equation (2.4). This is called the *trivial representation*. A representation in which distinct elements of G are associated with distinct matrices is said to be *faithful*.

2.2.2 Examples

Group of real numbers under addition

Consider the group $(\mathbb{R}, +)$ of real numbers under the operation of addition. Any real number a can be represented by the matrix:

$$D(a) = \begin{pmatrix} 1 & 0 \\ a & 1 \end{pmatrix}.$$

We verify that $D(a)D(b) = D(a+b)$, ensuring that equation (2.4) is satisfied. The identity element of $(\mathbb{R}, +)$ is 0 with $D(0) = I$. Therefore, the matrices $D(a)$ are a representation of the group $(\mathbb{R}, +)$. We will revisit these matrices when discussing the group of Galilean transformations.

Group of rotations

There are different but equivalent ways to define what we mean by a rotation. We can either describe a rotation as a transformation acting on vectors, or by the properties it preserves, such as the norm and the angle between vectors.

We first define a rotation by its explicit transformation of a vector in two-dimensional space:

$$R(\theta) = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}. \tag{2.5}$$

As shown in section 2.1.3, any matrix in SO(2) can be written in this form for some angle θ, and therefore represents a rotation in two dimensions. Conversely, the group $R_{(2)}$ of rotations in two dimensions can be represented by SO(2): each element of $R_{(2)}$ corresponds to a unique matrix $R(\theta)$ in SO(2), with θ defined modulo 2π.

Performing two successive rotations by angles θ and θ' is equivalent to a single rotation by $\theta + \theta'$. Therefore, equation (2.4) requires that $R(\theta + \theta') = R(\theta) R(\theta')$. It is straightforward to verify that this relation holds, confirming that the matrices $R(\theta)$ form a representation of the rotation group. This discussion generalises naturally to any dimension n, where rotations are represented by the matrices of SO(n). For $n \geq 3$, however, a rotation is no longer described by a single angle, and rotations about different axes do not commute, as we will see in section 2.3.

Alternatively, we can define a rotation by what it *preserves*, namely, the norm and the angle between vectors. We now demonstrate that matrices in SO(2) have these properties, and therefore correctly represent rotations.

Consider a vector $\mathbf{r} = \begin{pmatrix} x \\ y \end{pmatrix}$ in Cartesian coordinates. Rotating $\mathbf{r}$ by an angle θ around the origin gives:

$$\mathbf{r}' = \begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = R(\theta)\,\mathbf{r}.$$

The squared norm of $\mathbf{r}$ is $\|\mathbf{r}\|^2 = \mathbf{r}^\top \mathbf{r}$, where $\mathbf{r}^\top = (x, y)$ is the transpose of $\mathbf{r}$. The squared norm of $\mathbf{r}'$ is $\|\mathbf{r}'\|^2 = \mathbf{r}'^\top \mathbf{r}' = (R\mathbf{r})^\top R\mathbf{r} = \mathbf{r}^\top R^\top R\mathbf{r}$, where the dependence on θ is understood. Since R is orthogonal, we have $R^\top R = I$, so $\|\mathbf{r}'\| = \|\mathbf{r}\|$. Therefore, the norm is conserved.

The angle α between two vectors $\mathbf{r}_1$ and $\mathbf{r}_2$ is defined by their scalar product: $\cos\alpha = \mathbf{r}_1^\top \mathbf{r}_2 / (\|\mathbf{r}_1\| \|\mathbf{r}_2\|)$. Since R preserves the norm, it will also preserve the angle if it conserves the scalar product $\mathbf{r}_1^\top \mathbf{r}_2$. We have $\mathbf{r}_1'^\top \mathbf{r}_2' = (R\mathbf{r}_1)^\top R\mathbf{r}_2 = \mathbf{r}_1^\top R^\top R\mathbf{r}_2 = \mathbf{r}_1^\top \mathbf{r}_2$. Therefore, the angle between $\mathbf{r}_1'$ and $\mathbf{r}_2'$ is the same as the angle between $\mathbf{r}_1$ and $\mathbf{r}_2$.

We have shown that the matrices of SO(2) leave both the norm and scalar product invariant, confirming that they provide a representation of rotations in two dimensions. Conversely, if we require a matrice R representing rotations to conserve the scalar product (and therefore the norm), it must satisfy $R^\top R = I$ and therefore belong to the O(2) group. Indeed, if $\mathbf{r}_1^\top R^\top R\,\mathbf{r}_2 = \mathbf{r}_1^\top \mathbf{r}_2$ for all $\mathbf{r}_1$ and $\mathbf{r}_2$, choosing them among the basis vectors $(1, 0)^\top$ and $(0, 1)^\top$ shows that each coefficient of $R^\top R$ equals the corresponding coefficient of I. Moreover, as shown in section 2.1.3, only the matrices in O(2) with determinant $+1$ correspond to rotations; thus, the group representing rotations in two dimensions is SO(2).

In summary, rotations can be defined either directly through the transformation $R(\theta)$, or by specifying that they leave the norm and scalar product invariant, which implies that their representation must be orthogonal matrices. To distinguish rotations from reflections, we must additionally require that these matrices have a determinant of $+1$. These two viewpoints are equivalent and both establish SO(2) as the representation of the rotation group in two dimensions.

2.3 Lie algebra and generators

The discussion of the group of rotations leads to an important question: if we define a transformation by specifying that it leaves certain quantities invariant, can we derive the representations of this transformation? For rotations in two dimensions, we have shown that the invariance of the scalar product requires $R^\top R = I$. In section 2.1.3, we demonstrated that this condition imposes specific relationships between the coefficients of the matrix. However, is there a more general method to derive the matrix given the condition $R^\top R = I$, which would be especially useful for dimensions greater than two?

This section addresses these questions by introducing the concept of Lie algebra, which allows us to view a transformation as a succession of infinitesimal transformations. This relies on the fact that the transformations form a group: since the composition of two transformations is itself a transformation of the same kind (closure), a finite transformation can be built by composing infinitesimal ones. This approach is applicable only to *continuous* groups of transformations, where it is possible to find elements that are infinitesimally close to the identity. These groups are known as *Lie groups*.

2.3.1 The SO(2) group

Consider a rotation by an infinitesimal angle around the origin. Without loss of generality, this can be represented by a matrix $R = I + X$ where X is first-order in the infinitesimal angle. We define rotations as transformations that leave the scalar product (and therefore the norm) invariant. As shown above, this condition implies $R^\top R = I$. To first order in the infinitesimal angle, this yields $I + X^\top + X = I$, and therefore $X^\top = -X$, meaning that X is antisymmetric. This implies that X is proportional to the following matrix:

$$J = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \tag{2.6}$$

and since X is first-order in the infinitesimal angle, the factor, which we denote as $\delta\theta$, is proportional to this angle (we will show *a posteriori* that $\delta\theta$ *is* indeed the infinitesimal angle). Therefore, to first order in $\delta\theta$:

$$R = I + \delta\theta J = \begin{pmatrix} 1 & -\delta\theta \\ \delta\theta & 1 \end{pmatrix}. \tag{2.7}$$

Applying this transformation to a vector $\mathbf{r} = \begin{pmatrix} x \\ y \end{pmatrix}$ in Cartesian coordinates results in a new vector $\mathbf{r}'$ where $x' = x - \delta\theta y$ and $y' = y + \delta\theta x$. This recovers the geometric definition of a rotation, confirming that $\delta\theta$ is indeed the rotation angle. From now on, we denote the matrix as $R(\delta\theta)$.

Lie algebra is based on the idea that a rotation by an angle θ is equivalent to performing N rotations by infinitesimal angles $\delta\theta = \theta/N$. In the limit as $N \to \infty$, these two approaches are equivalent. The matrix representing the rotation by an angle θ is then given by[3]:

$$R(\theta) = \lim_{N\to+\infty} [R(\delta\theta)]^N = \lim_{N\to+\infty} \left(I + \frac{\theta}{N} J\right)^N = \mathrm{e}^{\theta J}, \tag{2.8}$$

where the exponential is defined by:

$$\mathrm{e}^{\theta J} = \sum_{k=0}^{+\infty} \frac{(\theta J)^k}{k!}. \tag{2.9}$$

The matrix J is called the *generator* of the rotation group. More generally, the vector space spanned by the generators of a group, equipped with the commutator introduced in section 2.3.3, is known as the *Lie algebra* of the group. In this specific case, the generator is an element of the group, though this is not generally true for all groups.

Since $J^2 = -I$, we have $J^{2n} = (-1)^n I$ and $J^{2n+1} = JJ^{2n} = (-1)^n J$. Therefore, we can separate the contributions from the odd and even powers in the exponential series (2.9), yielding:

$$R(\theta) = \underbrace{\sum_{k=0}^{+\infty} \frac{\theta^{2k}}{(2k)!}(-1)^k}_{\cos\theta} I + \underbrace{\sum_{k=0}^{+\infty} \frac{\theta^{2k+1}}{(2k+1)!}(-1)^k}_{\sin\theta} J. \tag{2.10}$$

Inserting the values of I and J, this gives:

$$R(\theta) = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}. \tag{2.11}$$

[3]We use the relation $\mathrm{e}^A = \lim_{N\to+\infty}(I + A/N)^N$, valid for any square matrix A. To prove this, we use the binomial theorem which gives, since I and A commute:

$$\left(I + \frac{A}{N}\right)^N = \sum_{k=0}^{N} \binom{N}{k} \frac{A^k}{N^k} = \sum_{k=0}^{N} \frac{N(N-1)\cdots(N-k+1)}{N^k} \frac{A^k}{k!}.$$

Each coefficient $N(N-1)\cdots(N-k+1)/N^k$ lies between 0 and 1 and tends to 1 as $N \to +\infty$, so that $\lim_{N\to+\infty}(I + A/N)^N = \sum_{k=0}^{+\infty} A^k/k! = \mathrm{e}^A$.

Thus, using only the requirement that certain quantities remain invariant under the transformation, we recover the matrix obtained earlier through geometric arguments. Although we only imposed the condition that the matrix be orthogonal, without explicitly requiring a determinant constraint, we still obtain matrices with a determinant of 1, which belong to SO(2). This is because, by assuming that for infinitesimal transformations R was infinitesimally close to I, we excluded reflections and thus ruled out matrices that satisfy the orthogonality condition but have a determinant of -1.

2.3.2 The SO(3) group

The Lie algebra of SO(2) is one-dimensional, meaning there is only one generator. This is because only one parameter is needed to describe a rotation around a fixed point in a plane. In general, the number of generators corresponds to the number of independent parameters required to describe all the transformations within the group. We now examine rotations around an arbitrary axis in three dimensions.

As before, we define a rotation in three dimensions as a transformation that preserves the scalar product, implying it is represented by a matrix R that satisfies $R^\top R = I$. For an infinitesimal rotation by an angle $\delta\theta$ around a unit vector $\hat{\mathbf{n}}$, R can be written as $R(\delta\theta, \hat{\mathbf{n}}) = I + X$, with $X^\top = -X$. In two dimensions, any antisymmetric matrix is a multiple of J. In three dimensions, X can be expressed as a linear combination of the following three matrices:

$$J_x = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}, \quad J_y = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{pmatrix}, \quad J_z = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}. \tag{2.12}$$

Therefore, $X = \delta\theta_x J_x + \delta\theta_y J_y + \delta\theta_z J_z$, where the matrix $I + \delta\theta_x J_x$ corresponds to a rotation by an angle $\delta\theta_x$ around the x-axis, and similarly for the y- and z-terms.

To relate $\delta\theta_x$, $\delta\theta_y$ and $\delta\theta_z$ to $\delta\theta$ and $\hat{\mathbf{n}}$, we write the rotation matrix $R(\delta\theta, \hat{\mathbf{n}}) = I + X$ as:

$$R(\delta\theta, \hat{\mathbf{n}}) = \begin{pmatrix} 1 & -\delta\theta_z & \delta\theta_y \\ \delta\theta_z & 1 & -\delta\theta_x \\ -\delta\theta_y & \delta\theta_x & 1 \end{pmatrix}. \tag{2.13}$$

We then show that, for this matrix to represent a rotation by an angle $\delta\theta$ around the unit vector $\hat{\mathbf{n}}$, its components must satisfy $\delta\theta_i = n_i \delta\theta$ for $i = x$, y or z. To see this, we note first that $\hat{\mathbf{n}}$ remains invariant under the rotation, meaning $R\hat{\mathbf{n}} = \hat{\mathbf{n}}$ or, equivalently, $X\hat{\mathbf{n}} = 0$. Since $X\mathbf{v} = (\delta\theta_x, \delta\theta_y, \delta\theta_z) \times \mathbf{v}$ for any vector $\mathbf{v}$, as can be checked from equation (2.13), the condition $X\hat{\mathbf{n}} = 0$ implies that the vector $(\delta\theta_x, \delta\theta_y, \delta\theta_z)$ must be parallel to $\hat{\mathbf{n}}$, with a proportionality coefficient we denote as λ. A vector $\mathbf{u}$ therefore transforms as $R\mathbf{u} = \mathbf{u} + \lambda\, \hat{\mathbf{n}} \times \mathbf{u}$. For a vector $\mathbf{u}$ perpendicular to $\hat{\mathbf{n}}$, with norm $u = \|\mathbf{u}\|$, the norm of $R\mathbf{u} - \mathbf{u}$ should be $u\,\delta\theta$ for an infinitesimal rotation by an angle $\delta\theta$, which gives $|\lambda| = \delta\theta$. Choosing the rotation to be counterclockwise about $\hat{\mathbf{n}}$ (right-hand rule), as for J in two dimensions, fixes the sign and we obtain $\lambda = \delta\theta$.

The matrices J_x, J_y and J_z serve as the generators of the SO(3) group and form its Lie algebra. Since there are three independent parameters required to describe a rotation in three dimensions, there are three corresponding generators. Note that the determinant of each of these matrices is 0, so they do not belong to SO(3) themselves.

Following the reasoning from the previous section, we can decompose a three-dimensional rotation by an angle θ around the vector $\hat{\mathbf{n}}$ into a sequence of infinitesimal rotations by angles $\delta\theta = \theta/N$ around $\hat{\mathbf{n}}$, with $N \to +\infty$. Each of these infinitesimal rotations is associated with the matrix $R(\delta\theta, \hat{\mathbf{n}})$, which can be written as:

$$R(\delta\theta, \hat{\mathbf{n}}) = I + \delta\theta n_x J_x + \delta\theta n_y J_y + \delta\theta n_z J_z = I + \frac{\theta}{N} n_x J_x + \frac{\theta}{N} n_y J_y + \frac{\theta}{N} n_z J_z.$$

The matrix representing the rotation by an angle θ around $\hat{\mathbf{n}}$ is then:

$$R\left(\theta, \hat{\mathbf{n}}\right) = \lim_{N\to+\infty} \left[R\left(\delta\theta, \hat{\mathbf{n}}\right)\right]^N = \lim_{N\to+\infty} \left(I + \frac{\theta}{N} n_x J_x + \frac{\theta}{N} n_y J_y + \frac{\theta}{N} n_z J_z\right)^N. \tag{2.14}$$

This leads to:

$$R\left(\theta, \hat{\mathbf{n}}\right) = \mathrm{e}^{\theta n_x J_x + \theta n_y J_y + \theta n_z J_z} = \mathrm{e}^{\theta \hat{\mathbf{n}}\cdot\mathbf{J}}, \tag{2.15}$$

where $\mathbf{J}$ is defined through $\hat{\mathbf{n}} \cdot \mathbf{J} = n_x J_x + n_y J_y + n_z J_z$.

Note that this does not represent a combination of finite rotations around the x-, y- and z-axes by angles θn_x, θn_y and θn_z. While it is possible to express a rotation by an angle θ around $\hat{\mathbf{n}}$ as a succession of three finite rotations around the x-, y- and z-axes, the angles around each of these axes (known as Tait-Bryan angles, a variant of the Euler angles) are generally different from θn_x, θn_y and θn_z as defined above.

To determine the coefficients of the matrix $R\left(\theta, \hat{\mathbf{n}}\right)$, we expand the exponential as follows:

$$\mathrm{e}^{\theta \hat{\mathbf{n}}\cdot\mathbf{J}} = \sum_{k=0}^{+\infty} \frac{(\theta \hat{\mathbf{n}} \cdot \mathbf{J})^k}{k!}. \tag{2.16}$$

Using the expressions for the matrices J_i given in equations (2.12), we obtain:

$$\hat{\mathbf{n}} \cdot \mathbf{J} = \begin{pmatrix} 0 & -n_z & n_y \\ n_z & 0 & -n_x \\ -n_y & n_x & 0 \end{pmatrix}. \tag{2.17}$$

Given that $n_x^2 + n_y^2 + n_z^2 = 1$, it follows that $(\hat{\mathbf{n}} \cdot \mathbf{J})^3 = -\hat{\mathbf{n}} \cdot \mathbf{J}$ (note that this does not imply that the square of the matrix equals minus the identity, as the matrix is not invertible). Therefore, for $k \geq 1$, we have $(\hat{\mathbf{n}} \cdot \mathbf{J})^{2k} = (-1)^{k+1} (\hat{\mathbf{n}} \cdot \mathbf{J})^2$ and $(\hat{\mathbf{n}} \cdot \mathbf{J})^{2k-1} = (-1)^{k-1} \hat{\mathbf{n}} \cdot \mathbf{J}$.

By separating the contributions from the odd and even powers in the exponential series (2.16), we obtain:

$$R(\theta, \hat{\mathbf{n}}) = I - \sum_{k=1}^{+\infty} \frac{\theta^{2k}}{(2k)!}(-1)^k (\hat{\mathbf{n}} \cdot \mathbf{J})^2 + \sum_{k=1}^{+\infty} \frac{\theta^{2k-1}}{(2k-1)!}(-1)^{k-1} \hat{\mathbf{n}} \cdot \mathbf{J}, \tag{2.18}$$

which simplifies to:

$$\boxed{R(\theta, \hat{\mathbf{n}}) = I + (1 - \cos\theta)(\hat{\mathbf{n}} \cdot \mathbf{J})^2 + \sin\theta\, \hat{\mathbf{n}} \cdot \mathbf{J}.} \tag{2.19}$$

This is known as *Rodrigues' rotation formula*.

2.3.3 Commutators

The group of rotations is non-commutative, meaning that performing one rotation followed by another is generally not equivalent to performing them in the reverse order. Consider two infinitesimal transformations $M = I + X$ and $M' = I + X'$ (not necessarily rotations). These transformations commute if $MM' = M'M$, which implies $XX' = X'X$. We define the commutator as the matrix:

$$[X, X'] = XX' - X'X. \tag{2.20}$$

Thus, the transformations commute if $[X, X'] = 0$.

For the rotation group SO(3), the commutation relations are $[J_x, J_y] = J_z$, $[J_y, J_z] = J_x$ and $[J_z, J_x] = J_y$, demonstrating that rotations do not commute.

Combined with equation (2.15), this implies that, for two matrices A and B, $\mathrm{e}^A\mathrm{e}^B$ is generally not equal to e^{A+B}. The equality holds if A and B commute, but not in general otherwise. This can be illustrated by considering a rotation about the x-axis by an angle θ_x and a rotation around the y-axis by an angle θ_y, represented respectively by $R_x = \mathrm{e}^{\theta_x J_x}$ and $R_y = \mathrm{e}^{\theta_y J_y}$. Performing the x-rotation followed by the y-rotation gives $R_y R_x = \mathrm{e}^{\theta_y J_y}\mathrm{e}^{\theta_x J_x}$. If this were equal to $\mathrm{e}^{\theta_x J_x + \theta_y J_y}$, then the order of the rotations would not matter, and the rotations would commute.

An important result when working with Lie algebras is the Campbell-Baker-Hausdorff (CBH) theorem, which states that for two elements A and B of a Lie algebra:

$$\mathrm{e}^A\mathrm{e}^B = \mathrm{e}^C, \tag{2.21}$$

where

$$C = A + B + \frac{1}{2}[A, B] + \frac{1}{12}([A, [A, B]] - [B, [A, B]]) + \text{higher order terms}. \tag{2.22}$$

This formula applies to any two square matrices A and B, but the series in equation (2.22) is only guaranteed to converge when A and B are sufficiently small (in the sense of a matrix norm), as is the case for transformations close to the identity.

The CBH formula highlights the significance of the commutator. Returning to the rotations R_x and R_y, if they commuted, we would have $R_yR_x = R_xR_y$, which implies $(R_xR_y)^{-1}\,R_yR_x = R_y^{-1}R_x^{-1}R_yR_x = I$. In other words, performing the x-rotation, then the y-rotation, then undoing the x-rotation and finally undoing the y-rotation would return the system to its initial state. We can evaluate this product explicitly, noting that $R_x^{-1} = \mathrm{e}^{-\theta_x J_x}$ and $R_y^{-1} = \mathrm{e}^{-\theta_y J_y}$. Using equations (2.21) and (2.22) to second order in the (small) angles gives:

$$\begin{aligned} R_yR_x &= \mathrm{e}^{\theta_x J_x + \theta_y J_y + \frac{1}{2}\theta_x\theta_y[J_y,J_x]}, \\ R_y^{-1}R_x^{-1} &= \mathrm{e}^{-\theta_x J_x - \theta_y J_y + \frac{1}{2}\theta_x\theta_y[J_y,J_x]}. \end{aligned}$$

Therefore:

$$R_y^{-1}R_x^{-1}R_yR_x = \mathrm{e}^{\theta_x\theta_y[J_y,J_x]} = \mathrm{e}^{-\theta_x\theta_y J_z},$$

which represents a rotation by an angle $-\theta_x\theta_y$ around the z-axis. The commutator therefore specifies the infinitesimal transformation that arises when the order of two non-commuting rotations is reversed.

2.4 Galilean and Lorentz transformations in mechanics

Consider three observers moving with respect to each other. The laws of physics as seen by observer j are a transformation $T(i \to j)$ of the laws seen by observer i. Transforming from observer 1 to observer 2, and then from observer 2 to observer 3, must yield the same result as transforming directly from observer 1 to observer 3. In other words:

$$T(2 \to 3)T(1 \to 2) = T(1 \to 3).$$

This indicates that *the transformations T form a group*. Consequently, we can generate a representation of this group using Lie algebra to derive the transformations.

2.4.1 The Galilean group

The Galilean group consists of transformations between reference frames that preserve the time interval between two non-simultaneous events, as well as the distances between two events occurring simultaneously.

Boost in one dimension

Consider a frame (S), with origin O, in which a point mass is located at position x on the x-axis at time t. Now, consider another frame (S$'$) that moves along the x-axis with velocity $-v$ relative to (S), that is to say towards decreasing x for $v > 0$, with its origin O' coinciding with O at $t = 0$. We synchronise the time origin in (S$'$) with that in (S). This type of transformation, where only the velocity of the reference frame changes, is called a *boost*.

Since O' is located at $x = -vt$ in (S), a Galilean transformation between these two frames is characterised by $t' = t$ and $x' = x + vt$.[4] If we define a spacetime position vector in (S) as $\mathbf{r} = \begin{pmatrix} t \\ x \end{pmatrix}$, then the Galilean transformation from frame (S) to frame (S′) can be represented by the matrix:

$$D(v) = \begin{pmatrix} 1 & 0 \\ v & 1 \end{pmatrix}, \tag{2.23}$$

which acts on the vector $\mathbf{r}$, since $D(v)\,\mathbf{r} = \mathbf{r}'$. The property $D(v')D(v) = D(v+v')$ reflects the fact that transforming to a frame moving at velocity v, and then to another frame moving at velocity v' relative to (S′), is equivalent to moving directly to a frame moving at velocity $v + v'$ relative to (S). This is the well-known composition of the velocities.

If we consider a transformation to a frame moving with an infinitesimal velocity δv, then $D = I + \delta v X$, where:

$$X = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \tag{2.24}$$

is the generator of the group of Galilean transformations in one dimension. Since $X^2 = 0$:

$$D(v) = \mathrm{e}^{vX} = \sum_{k=0}^{+\infty} \frac{(vX)^k}{k!} = I + vX,$$

as expected.

Boost and translations in one dimension

Now, consider a boost together with a *translation in space and time*. Let frame (S′) have its origin O' at $x = -x_0$ at $t = 0$ ($x_0 > 0$), with the origin of time in frame (S′) corresponding to $t = t_0$ in frame (S). The Galilean transformation then results in $x' = x + vt + x_0$ and $t' = t - t_0$.

We observe that while the boost results only in a homogeneous term, the translations introduce inhomogeneous terms. To manage these, we extend the usual spacetime position vector to $\mathbf{s} = \begin{pmatrix} 1 \\ t \\ x \end{pmatrix}$. This extension allows us to incorporate translations into linear transformations via matrix operations.

[4] We will use the opposite convention for the Lorentz transformation in section 2.4.3, where (S′) moves towards increasing x with velocity $+v$.

Specifically, the Galilean transformation from frame (S) to frame (S′) can be represented by the following matrix:

$$M\left(v, t_0, x_0\right) = \begin{pmatrix} 1 & 0 & 0 \\ -t_0 & 1 & 0 \\ x_0 & v & 1 \end{pmatrix}, \tag{2.25}$$

which acts on the vector $\mathbf{s}$. Transforming from frame (S) to frame (S′) and then to frame (S″) transforms $\mathbf{s}$ into $\mathbf{s}''$ such that:

$$\begin{pmatrix} 1 \\ t'' \\ x'' \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ -t_0' & 1 & 0 \\ x_0' & v' & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ -t_0 & 1 & 0 \\ x_0 & v & 1 \end{pmatrix} \begin{pmatrix} 1 \\ t \\ x \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ -t_0 - t_0' & 1 & 0 \\ x_0 + x_0' - v't_0 & v + v' & 1 \end{pmatrix} \begin{pmatrix} 1 \\ t \\ x \end{pmatrix},$$

where v', x_0' and t_0' are related to the transformation from frame (S′) to frame (S″). The 3×3 matrix on the right-hand side represents the transformation from frame (S) to frame (S″), characterised by the following parameters:

$$t_0'' = t_0 + t_0', \quad v'' = v + v', \quad x_0'' = x_0 + x_0' - v't_0.$$

It is clear that the group is not commutative, meaning the order in which the transformations are performed matters.

We can construct the Lie algebra by writing the matrices associated with each of the infinitesimal transformations[5]:

$$M\left(\delta v, 0, 0\right) = I + \delta v X_K, \quad M\left(0, \delta t_0, 0\right) = I + \delta t_0 X_H, \quad M\left(0, 0, \delta x_0\right) = I + \delta x_0 X_P,$$

with:

$$X_K = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}, \quad X_H = \begin{pmatrix} 0 & 0 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad X_P = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}. \tag{2.26}$$

For an arbitrary infinitesimal transformation, we have:

$$M\left(\delta v, \delta t_0, \delta x_0\right) = I + \delta v X_K + \delta t_0 X_H + \delta x_0 X_P.$$

As for rotations, a finite transformation can be obtained by composing N infinitesimal transformations with parameters v/N, t_0/N and x_0/N, and taking the limit $N \to +\infty$:

$$\lim_{N \to +\infty} \left[M\left(\frac{v}{N}, \frac{t_0}{N}, \frac{x_0}{N} \right) \right]^N = \lim_{N \to +\infty} \left[I + \frac{v}{N} X_K + \frac{t_0}{N} X_H + \frac{x_0}{N} X_P \right]^N = \mathrm{e}^{v X_K + t_0 X_H + x_0 X_P}$$

[5]The subscript 'K' for the generator associated with a boost signifies *kinematic* transformations, as boosts are related to changes in velocity. The subscript 'P' is used for the generators of spatial translations because they are generated by the *momentum* operator. Finally, the subscript 'H' refers to the *Hamiltonian*, which is the generator of time evolution.

To evaluate this exponential, we use its series expansion, as defined in equation (2.9). Since $X_K^2 = X_H^2 = X_P^2 = 0$ and the product of any two of these matrices is zero, except for $X_K X_H = -X_P$, we have $(vX_K + t_0 X_H + x_0 X_P)^2 = -vt_0 X_P$, and all higher powers vanish. Therefore:

$$\mathrm{e}^{vX_K + t_0 X_H + x_0 X_P} = I + vX_K + t_0 X_H + \left(x_0 - \tfrac{1}{2} v t_0\right) X_P = M\left(v, t_0, x_0 - \tfrac{1}{2} v t_0\right).$$

This is indeed a Galilean transformation and, since x_0 is arbitrary, any transformation $M(v, t_0, x_0)$ can be obtained in this way, as $\mathrm{e}^{vX_K + t_0 X_H + (x_0 + vt_0/2) X_P}$. However, the parameters do not coincide with those of equation (2.25): because boosts and time translations do not commute ($X_K X_H - X_H X_K = -X_P$), performing them simultaneously in infinitesimal steps generates an additional spatial translation. This is analogous to the remark made in section 2.3.2 about rotations about the x-, y- and z-axes.

To recover the parameters of equation (2.25) directly, we can instead perform the three transformations in turn rather than simultaneously: a spatial translation by x_0, followed by a boost with velocity v, followed by a time translation by t_0. Each of these is represented by the exponential of a single generator, which reduces to its first two terms since the square of each generator vanishes, e.g. $\mathrm{e}^{vX_K} = I + vX_K$. Since the matrix on the right acts first on $\mathbf{s}$, the general transformation is then written as the product of exponentials:

$$M(v, t_0, x_0) = \mathrm{e}^{t_0 X_H}\, \mathrm{e}^{vX_K}\, \mathrm{e}^{x_0 X_P} = (I + t_0 X_H)(I + vX_K)(I + x_0 X_P). \tag{2.27}$$

The order of the factors matters, since the Galilean group is non-commutative. For example:

$$\begin{aligned} &(I + t_0 X_H)(I + vX_K)(I + x_0 X_P) \quad \text{yields} \quad x' = x + vt + x_0, \\ &(I + vX_K)(I + t_0 X_H)(I + x_0 X_P) \quad \text{yields} \quad x' = x + v(t - t_0) + x_0. \end{aligned}$$

Galilean transformations in three dimensions

So far, we have only considered a single spatial dimension. However, in three dimensions, the coordinate axes of (S′) could be rotated relative to those of (S), the velocity of (S′) relative to (S) could have any direction, and the origin of (S′) could be shifted by a vector (x_0, y_0, z_0). Consequently, the spacetime position vector and the transformation matrix take the following form:

$$\mathbf{s} = \begin{pmatrix} 1 \\ t \\ x \\ y \\ z \end{pmatrix}, \quad M = \left(\begin{array}{cc|ccc} 1 & 0 & 0 & 0 & 0 \\ -t_0 & 1 & 0 & 0 & 0 \\ \hline x_0 & v_x & & & \\ y_0 & v_y & & R & \\ z_0 & v_z & & & \end{array}\right), \tag{2.28}$$

where R is the rotation matrix. There are now 10 parameters describing a transformation: 3 for rotations, 3 for boosts, 3 for spatial translations and 1 for the time translation.

Consequently, there are 10 generators in the Lie algebra, each represented by a 5×5 matrix. We can derive these generators by considering an infinitesimal transformation associated with each parameter in turn.

A translation of time by an infinitesimal amount δt_0 results in a matrix transformation $M = I + \delta t_0 X_H$, where X_H is a 5-dimensional extension of the 3×3 matrix given in the previous section:

$$X_H = \left(\begin{array}{cc|ccc} 0 & 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & & & \\ 0 & 0 & & 0 & \\ 0 & 0 & & & \end{array}\right). \tag{2.29}$$

Similarly, spatial translations are associated with the generators X_{Px}, X_{Py} and X_{Pz}, while boosts are associated with the generators X_{Kx}, X_{Ky} and X_{Kz}. All of these generators can be derived in the same manner as X_P and X_K were calculated above. For example:

$$X_{Px} = \left(\begin{array}{cc|ccc} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ \hline 1 & 0 & & & \\ 0 & 0 & & 0 & \\ 0 & 0 & & & \end{array}\right), \quad X_{Kx} = \left(\begin{array}{cc|ccc} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 1 & & & \\ 0 & 0 & & 0 & \\ 0 & 0 & & & \end{array}\right). \tag{2.30}$$

For the rotations, the generators X_{J_i}, with $i = x, y, z$, are those calculated in section 2.3.2 for the SO(3) group, but extended to 5×5 matrices:

$$X_{J_i} = \left(\begin{array}{cc|ccc} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & & & \\ 0 & 0 & & J_i & \\ 0 & 0 & & & \end{array}\right). \tag{2.31}$$

General transformation matrices can be derived from these 10 generators using the same method applied in the simpler cases discussed above.

We can gain insights into the nature of the Galilean transformations by examining their commutation relations. If two generators commute, it means that the associated transformations do not interfere with each other. Conversely, if they do not commute, their effects are interdependent.

- $[X_{J_i}, X_{J_j}] = \epsilon_{ijk} X_{J_k}$: performing infinitesimal rotations about the i and j axes, and then undoing them in the same order, results, to second order, in a rotation about the k axis,

- $[X_{J_i}, X_{P_j}] = \epsilon_{ijk} X_{P_k}$: rotations mix the directions of spatial translations,
- $[X_{J_i}, X_{K_j}] = \epsilon_{ijk} X_{K_k}$: rotations mix boost directions, reflecting the fact that boosts transform as vectors under rotations,
- $[X_{P_i}, X_{P_j}] = 0$: spatial translations in different directions are independent,
- $[X_{P_i}, X_{K_j}] = 0$: boosts and spatial translations operate independently, consistent with v and x_0 being independent parameters in the transformation matrix (2.25),
- $[X_{K_i}, X_{K_j}] = 0$: successive boosts in different directions do not interfere, a key feature of Galilean relativity,
- $[X_{J_i}, X_H] = 0$ and $[X_{P_i}, X_H] = 0$: time translations are decoupled from spatial transformations, again consistent with Galilean relativity,
- $[X_{K_i}, X_H] = -X_{P_i}$: boosts and time translations are related to spatial translations, as seen from the transformation matrix (2.25), where the term vt produces a spatial shift.

All these commutators are linear combinations of the generators, that is to say elements of the Lie algebra. This closure of the Lie algebra under commutators is consistent with the fact that the Galilean transformations form a group: the product of any two elements (whether they are rotations, boosts, translations, or a combination) is another element of the Galilean group.

2.4.2 Relativistic mechanics: invariance of the spacetime interval

The fundamental principle of the Galilean transformation is the universality of time. As discussed earlier, a Galilean transformation preserves, on the one hand, the time interval between two events and, on the other hand, the distance $\left(\Delta x^2 + \Delta y^2 + \Delta z^2\right)^{1/2}$ between two events occurring simultaneously. In three-dimensional Euclidean space, the position vector has coordinates (x, y, z) and the invariance of the distance between two events can be seen as the invariance of the length of the displacement vector between two locations in space.

In contrast, the Lorentz transformation is based on the inseparability of time and space, as we exist in a four-dimensional Minkowski spacetime. In this spacetime, there are four coordinates (ct, x, y, z), with time intrinsically linked to the spatial coordinates. Additionally, because the geometry of Minkowski spacetime is hyperbolic rather than Euclidean, the quantity preserved under a Lorentz transformation is the spacetime interval, $\Delta s^2 = -c^2\Delta t^2 + \Delta x^2 + \Delta y^2 + \Delta z^2$, rather than the more familiar squared displacement length, $c^2\Delta t^2 + \Delta x^2 + \Delta y^2 + \Delta z^2$. This means that in a spacetime diagram (typically with one spatial dimension x and the time dimension t), points undergoing a Lorentz boost move along hyperbolic curves. This will be discussed further in chapter 3. For this reason, Lorentz boosts are sometimes described as *hyperbolic rotations*, as will become apparent in section 2.4.3, where they are expressed in terms of hyperbolic functions.

The need for the invariant Δs^2 to include a negative sign for the time term can be understood by considering electromagnetic waves, which propagate in a vacuum at a constant speed c regardless of the observer's frame of reference. Consider a frame (S) with spatial coordinates (x, y, z) and time t. In another frame (S′) moving relative to (S), the spatial coordinates become (x', y', z') and time becomes t'. If the position of the wavefront of an electromagnetic wave advances by $\Delta r = \left(\Delta x^2 + \Delta y^2 + \Delta z^2\right)^{1/2}$ during the time interval Δt in (S), then $\Delta r = c\Delta t$, i.e. $\Delta s^2 = -c^2\Delta t^2 + \Delta x^2 + \Delta y^2 + \Delta z^2 = 0$. Similarly, in (S′), where the position of the wavefront advances by $\Delta r'$ during the time interval $\Delta t'$, we have $\Delta \mathrm{s}'^2 = -c^2\Delta t'^2 + \Delta x'^2 + \Delta y'^2 + \Delta z'^2 = 0$. Therefore, the Lorentz transformation from (t, x, y, z) to (t', x', y', z') leaves Δs^2 invariant for electromagnetic waves in vacuum. This result is actually much more general, as will be discussed in chapter 3, and any Δs^2 is invariant under a Lorentz transformation.

We define:

$$\mathbf{s} = \begin{pmatrix} ct \\ x \\ y \\ z \end{pmatrix}, \quad \text{and} \quad g = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \tag{2.32}$$

where the spacetime position vector $\mathbf{s}$ has ct as its first coordinate instead of just t, for unit consistency, and where g is known as the *Minkowski metric*. The space defined by the vectors $\mathbf{s}$ is called Minkowski space. The quantity that remains invariant under the Lorentz transformation, Δs^2, can be expressed as the product $\Delta\mathbf{s}^\top g \Delta\mathbf{s}$. Letting L represent the Lorentz transformation matrix, we have $\Delta\mathbf{s}' = L\Delta\mathbf{s}$. Thus, $\Delta\mathbf{s}'^\top g \Delta\mathbf{s}' = (L\Delta\mathbf{s})^\top gL\Delta\mathbf{s} = \Delta\mathbf{s}^\top L^\top gL\Delta\mathbf{s}$. Invariance requires this expression to be equal to $\Delta\mathbf{s}^\top g\Delta\mathbf{s}$. Since this condition must hold for any $\Delta\mathbf{s}$, and both $L^\top gL$ and g are symmetric matrices, it implies the following requirement:

$$\boxed{L^\top gL = g.} \tag{2.33}$$

As with Galilean transformations, the transformations between inertial frames in special relativity can involve boosts (changes in velocity), rotations, and space and time translations. The group that includes both boosts and rotations is known as the Lorentz group $\mathrm{SO}(1,3)$, while the group that also includes translations is called the Poincaré group $\mathrm{ISO}(1,3)$ (Inhomogeneous Special Orthogonal group, where 'inhomogeneous' reflects the fact that translations introduce inhomogeneities in the equations). Below, we describe the Lorentz group.

Transformations that preserve the commonly defined squared length of the displacement vector, $c^2\Delta t^2 + \Delta x^2 + \Delta y^2 + \Delta z^2$, form the $\mathrm{O}(4)$ group. In contrast, Lorentz transformations preserve Δs^2, where the time coordinate carries a different sign than the spatial coordinates. This difference is why the group of Lorentz transformations is labelled as $(1,3)$ instead of (4).

2.4.3 The Lorentz group SO$(1,3)$

Just as we derived the matrices R of SO(3) by solving $R^\top R = I$, we now seek the matrices L that satisfy equation $L^\top g L = g$. Taking the determinant on both sides of this equation yields $\det(L)^2 = 1$, so $\det(L) = \pm 1$. As noted in section 2.3.1, the Lie algebra, which generates L from infinitesimal transformations, only produces matrices with a determinant of 1. These transformations are smoothly connected to the identity transformation, whereas matrices with a determinant of -1 are not. Unlike SO(3), however, SO$(1,3)$ also contains matrices that are not smoothly connected to the identity, such as $-I$, which reverses both time and space. The Lie algebra therefore only generates the subgroup of SO$(1,3)$ that is connected to the identity, known as the *proper orthochronous* Lorentz group SO$^+(1,3)$, whose elements also preserve the direction of time (the top-left coefficient of L is at least 1). As is common in physics, we will refer to this subgroup simply as the Lorentz group and, with a slight abuse of notation, denote it by SO$(1,3)$ in the rest of these notes. The remaining elements of O$(1,3)$ can be obtained by multiplying the matrices in SO$^+(1,3)$ by g (time reversal), $-g$ (spatial inversion) or $-I$, similar to how the matrices of O(2) are derived from those of SO(2) by multiplying them by the matrix P, as defined in equation (2.3).

We now verify that the set of matrices satisfying equation (2.33) with a determinant of 1 forms a group:

(a) Closure: consider two matrices L_1 and L_2 that satisfy equation (2.33). Using the associativity of matrix multiplication, we find $(L_1 L_2)^\top g (L_1 L_2) = L_2^\top \left(L_1^\top g L_1\right) L_2 = L_2^\top g L_2 = g$, showing that $L_1 L_2$ also satisfies equation (2.33). Additionally, if $\det(L_1) = \det(L_2) = 1$, then $\det(L_1 L_2) = \det(L_1)\det(L_2) = 1$.

(b) Associativity: matrix multiplication is associative in general, and thus it remains associative for matrices that satisfy equation (2.33) in particular.

(c) Identity element: the identity matrix I satisfies $I^\top g I = g$ and its determinant is 1.

(d) Inverse element: if L satisfies $L^\top g L = g$, multiplying both sides by $\left(L^\top\right)^{-1}$ on the left and by L^{-1} on the right gives $g = \left(L^\top\right)^{-1} g L^{-1} = \left(L^{-1}\right)^\top g L^{-1}$, showing that L^{-1} also satisfies the condition. Additionally, since $\det\left(L^{-1}\right) = 1/\det(L)$, if $\det(L) = 1$, then $\det\left(L^{-1}\right) = 1$.

This confirms that the matrices satisfying equation (2.33) with a determinant of 1 form a group.

Generators

Each Lorentz transformation is characterised by 6 parameters: three for rotations and three for boosts. Consequently, there are six generators in the Lie algebra. Rotations leave time unchanged, whereas boosts mix the time and spatial coordinates, so that Δs^2, rather than the time interval and the spatial distance separately, remains invariant.

We first observe that if a transformation matrix L leaves time invariant, the requirement that it preserves $-c^2\Delta t^2+\Delta r^2$ implies that it must also preserve Δr^2. This condition, along with $\det(L) = 1$, implies that the spatial part of L corresponds to a rotation matrix R from SO(3). Therefore, any transformation that affects only the spatial coordinates must be a rotation, and its matrix takes the following form:

$$L = \left(\begin{array}{c|ccc} 1 & 0 & 0 & 0 \\ \hline 0 & & & \\ 0 & & R & \\ 0 & & & \end{array}\right). \tag{2.34}$$

The generators of these transformations, denoted as X_{J_i} for $i = x, y, z$, are those derived in section 2.3.2 for the SO(3) group, but extended to 4×4 matrices:

$$X_{J_i} = \left(\begin{array}{c|ccc} 0 & 0 & 0 & 0 \\ \hline 0 & & & \\ 0 & & J_i & \\ 0 & & & \end{array}\right). \tag{2.35}$$

Next, we determine the generators associated with boost transformations. For an infinitesimal boost δv_x along the x-axis, the transformation matrix is expressed as $L = I + \epsilon_x K_x$, where K_x is a 4×4 matrix and ϵ_x is an infinitesimal quantity. Here, we choose both ϵ_x and the coefficients of K_x to be dimensionless. As discussed in section 2.3.1 in the context of two-dimensional rotations, we do not have prior knowledge of the exact relationship between ϵ_x and δv_x. We only know that $\epsilon_x \propto \delta v_x$ to first order in δv_x.

Equation (2.33) then implies that $[I + \epsilon_x K_x]^\top g [I + \epsilon_x K_x] = g$. To first order in δv_x, this yields $K_x^\top g = -g K_x$. Since the boost is along the x-axis, we have $y' = y$ and $z' = z$, so K_x must take the form:

$$K_x = \begin{pmatrix} a & b & 0 & 0 \\ d & e & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \tag{2.36}$$

Applying the condition $K_x^\top g = -g K_x$ leads to $a = e = 0$ and $b = d$. This allows us to determine the generator X_{K_x}, from which K_x can be derived:

$$X_{K_x} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \tag{2.37}$$

Similarly, the generators for boosts along the y and z-axes are:

$$X_{K_y} = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \quad \text{and} \quad X_{K_z} = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix} \tag{2.38}$$

General transformations

Having established all the generators of the Lorentz group, we can now express any transformation matrix L, incorporating both rotations and boosts, as follows:

$$L = \mathrm{e}^{\theta_x X_{J_x} + \theta_y X_{J_y} + \theta_z X_{J_z} - \zeta_x X_{K_x} - \zeta_y X_{K_y} - \zeta_z X_{K_z}}. \tag{2.39}$$

Here $(\zeta_x, \zeta_y, \zeta_z)$ is known as the *boost vector*. This means that a boost by ζ_x in the x-direction is achieved through a sequence of infinitesimal boosts ϵ_x, and similarly for the y and z-directions. The choice of the minus sign will be explained later.

In the case of the Galilean transformations, knowing the transformation in advance allowed us to directly equate the boost vector with the velocity vector. For Lorentz transformations, two successive boosts along the same axis combine by adding their parameters, since $\mathrm{e}^{-\zeta_x X_{K_x}} \mathrm{e}^{-\zeta'_x X_{K_x}} = \mathrm{e}^{-(\zeta_x + \zeta'_x) X_{K_x}}$. If the boost parameter were equal to the velocity, velocities would simply add when transitioning through successive moving frames, and could exceed the speed of light. Consequently, the boost vector cannot be directly equated with the velocity vector. The relationship between the two will be derived below.

Commutation relations

As is evident from the form of the generators, rotations only affect the spatial coordinates, while boosts mix spatial and time coordinates. The rotations alone form the SO(3) group. However, boosts alone do not constitute a group. This is illustrated by the commutator $[X_{K_x}, X_{K_y}] = -X_{J_z}$, which shows that performing successive boosts in different directions results in a rotation.

More generally, the commutation relations of the Lorentz algebra are:

- $[X_{J_i}, X_{J_j}] = \epsilon_{ijk} X_{J_k}$: rotations about different axes do not commute, as in the Galilean group,
- $[X_{J_i}, X_{K_j}] = \epsilon_{ijk} X_{K_k}$: rotations mix boost directions, as in the Galilean case,
- $[X_{K_i}, X_{K_j}] = -\epsilon_{ijk} X_{J_k}$: unlike in the Galilean group, this commutator is non-zero, showing that successive boosts in different directions generate a rotation.

As for the Galilean group, these commutators are linear combinations of the generators, that is to say elements of the Lie algebra. This closure of the Lie algebra under commutators is consistent with the fact that the Lorentz transformations form a group: the combination of any two elements (whether rotations, boosts, or their combinations) results in another element of the Lorentz group.

Boosts

Now, assuming there is no rotation and only a finite boost ζ_x along the x-direction, we denote the transformation matrix as Λ_x instead of L. Following the same method used in section 2.3.1 for the SO(2) group, we obtain:

$$\Lambda_x = \mathrm{e}^{-\zeta_x X_{K_x}} = I + \sum_{k=1}^{+\infty} \frac{(-\zeta_x X_{K_x})^k}{k!}, \tag{2.40}$$

where I is the 4×4 identity matrix. Given that $(X_{K_x})^2 = I_{2\times 2}$, where:

$$I_{2\times 2} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \tag{2.41}$$

we find that, for $k \geq 1$, $(X_{K_x})^{2k} = I_{2\times 2}$ and $(X_{K_x})^{2k-1} = X_{K_x}$.

Therefore, we can separate the contributions from the odd and even powers in the exponential series (2.40), which yields:

$$\Lambda_x = I + \underbrace{\sum_{k=1}^{+\infty} \frac{(-\zeta_x)^{2k}}{(2k)!}}_{\cosh(-\zeta_x) - 1} I_{2\times 2} + \underbrace{\sum_{k=0}^{+\infty} \frac{(-\zeta_x)^{2k+1}}{(2k+1)!}}_{\sinh(-\zeta_x)} X_{K_x}. \tag{2.42}$$

Substituting the values of $I_{2\times 2}$ and X_{K_x}, we get:

$$\Lambda_x = \begin{pmatrix} \cosh\zeta_x & -\sinh\zeta_x & 0 & 0 \\ -\sinh\zeta_x & \cosh\zeta_x & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \tag{2.43}$$

The same calculation for a finite boost in the y- or z-directions yields:

$$\Lambda_y = \begin{pmatrix} \cosh\zeta_y & 0 & -\sinh\zeta_y & 0 \\ 0 & 1 & 0 & 0 \\ -\sinh\zeta_y & 0 & \cosh\zeta_y & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad \Lambda_z = \begin{pmatrix} \cosh\zeta_z & 0 & 0 & -\sinh\zeta_z \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -\sinh\zeta_z & 0 & 0 & \cosh\zeta_z \end{pmatrix}. \quad (2.44)$$

A more general matrix Λ can be obtained for a boost in any direction by calculating $\Lambda = \mathrm{e}^{-\zeta_x X_{K_x} - \zeta_y X_{K_y} - \zeta_z X_{K_z}}$.

The final step in obtaining the Lorentz transformation is to relate the boost to the velocity. Since the transformation involves only a boost and no translation, the origin O' of the coordinate system (S′) coincides with the origin O of the coordinate system (S) at time $t = 0$. When considering a boost in the x-direction, this means that the trajectory of O' is $x = v_x t$ in frame (S), while it is $x' = 0$ in frame (S′). Now, using $\mathbf{s}' = \Lambda_x \mathbf{s}$, we obtain $x' = (\cosh\zeta_x)\, x - (\sinh\zeta_x)\, ct$, which, as stated, must be 0 when $x = v_x t$. This yields $\tanh\zeta_x = v_x/c$. It now becomes clear why we chose the boost to have a minus sign in equation (2.39): this way, the boost has the same sign as the velocity component. From here on, we denote $\beta_x = v_x/c$. Therefore:

$$\cosh\zeta_x = \frac{1}{\sqrt{1-\tanh^2\zeta_x}} = \frac{1}{\sqrt{1-\beta_x^2}}, \quad \text{and} \quad \sinh\zeta_x = \cosh\zeta_x \tanh\zeta_x = \frac{\beta_x}{\sqrt{1-\beta_x^2}}.$$

Dropping the subscript x, we can thus rewrite the transformation matrix as:

$$\boxed{\Lambda = \begin{pmatrix} \gamma & -\beta\gamma & 0 & 0 \\ -\beta\gamma & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad \text{with} \quad \gamma = \frac{1}{\sqrt{1-\beta^2}}, \quad \text{and} \quad \beta = \frac{v}{c}.} \quad (2.45)$$

Similar results hold for Λ_y and Λ_z. The parameter:

$$\boxed{\zeta = \tanh^{-1}(\beta)\,,} \quad (2.46)$$

is called the *rapidity*.

Chapter 3

Fundamental concepts in special relativity

In the previous chapter, we derived the Lorentz transformation using the invariance of the spacetime interval and group theory formalism. This allowed us to obtain a comprehensive form of the transformation, including rotations and boosts in any direction. In this chapter, we will start with a brief review of a simple derivation of the transformation for a boost along one axis, as this emphasises the fundamental principles underlying the Lorentz transformation. Following this, we will discuss some core aspects of special relativity before introducing the important concept of the 4-vector.

Events are described within a *reference frame* (or frame, for short), which is a coordinate system that incorporates both space and time. A specific frame can be visualised as a collection of metre sticks to measure distances and a network of clocks to track time at every point in space. These clocks must be *synchronised*, meaning they all display the same time at a given moment and advance at the same rate when observed from within the frame. A particularly important type of frame is the *inertial* reference frame, where a particle not subject to any force moves at a constant speed in a straight line (or remains at rest if initially stationary).

The first postulate of relativity states:

The laws of physics are the same in all inertial reference frames.

In the context of mechanics, this principle was already recognised by Galileo and Newton. Einstein's contribution was to extend it to all the laws of physics, including electromagnetism. For Newton, time was universal, the same for all observers in all frames. This implied that a light ray moving with velocity c in one frame would be perceived as moving with velocity $c + v$ by an observer moving with velocity v relative to that frame, in the direction opposite to the ray. However, by the end of the 19^{th} century, precise measurements showed that the speed of light remained constant for observers in different inertial frames.

This led Einstein to propose the second postulate of relativity:

The speed of light in a vacuum is the same in all inertial reference frames.

This implies that *time cannot be universal*, and that if the clocks in two different frames are synchronised at one moment, they will not remain synchronised if the frames move relative to each other. Space and time coordinates are inseparable, and their combination is referred to as *spacetime*, or *Minkowski space*. A point with coordinates (t, x, y, z) is called an *event* (not an actual occurrence, but simply a point in spacetime, representing something that could happen). A path traced in spacetime is called a *worldline*.

3.1 Deriving the Lorentz transformation for a boost

Consider a frame (S) with origin O and coordinates (t, x) and another frame (S′) with origin O' and coordinates (t', x'), where the x- and x'-axes are parallel. Frame (S′) moves with velocity v towards increasing x relative to (S). We set $t' = 0$ to coincide with $t = 0$ and O' to coincide with O at that moment. How are (t', x') related to (t, x)?

The transformation must be linear, because uniform motion in a straight line in one inertial frame must also be uniform motion in a straight line in another inertial frame, and because the origins of the two frames coincide at $t = t' = 0$, which excludes any constant term. Moreover, O' has a trajectory $x = vt$ in (S), and $x' = 0$ in (S′). Since x' is a linear function of x and t that vanishes when $x = vt$, it must depend on x and t only through the combination $x - vt$. We can then write this as:

$$x' = (x - vt)\,\gamma(v), \tag{3.1}$$

where γ is a function of v. If (S′) were moving with velocity $-v$ (towards decreasing x), we would have $x' = (x + vt)\,\gamma(-v)$. In the first case, the point O at some time t_0 measured in (S) is viewed in (S′) as having coordinate $x' = -vt_0\gamma(v)$, whereas in the second case, it is viewed as having coordinate $x' = vt_0\gamma(-v)$. Since space is isotropic, the second situation is simply the mirror image of the first under the reflection $x \to -x$, $x' \to -x'$. The two coordinates must therefore be opposite, so that we must have $\gamma(-v) = \gamma(v)$.

Returning to the case where (S′) moves with velocity v, this is equivalent to (S) moving with velocity $-v$ relative to (S′). By the first postulate, the transformation from (t', x') to x must therefore take the same form as equation (3.1), with v replaced by $-v$. Using $\gamma(-v) = \gamma(v)$, this yields:

$$x = \left(x' + vt'\right)\gamma(v). \tag{3.2}$$

With $\gamma = 1$, this would correspond to a Galilean transformation.

Next, we consider a ray of light which travels along the trajectory $x = ct$ in (S). According to equation (3.1), for this ray viewed in (S′), $x' = (c - v)\, t\gamma(v)$. The second postulate of relativity states that this ray also travels with velocity c in (S′), so that $t' = x'/c$. Substituting these two expressions into equation (3.2) yields $x = \left(1 - v^2/c^2\right) ct\gamma^2(v)$. Since $x = ct$, we must have $\gamma^2(v)\left(1 - v^2/c^2\right) = 1$. Taking the positive root, since the transformation must reduce to the identity when $v = 0$, we obtain:

$$\gamma(v) = \frac{1}{\sqrt{1 - v^2/c^2}}. \tag{3.3}$$

To obtain t' as a function of x and t, we combine equations (3.1) and (3.2) to eliminate x', resulting in:

$$t' = \left(t - \frac{vx}{c^2}\right)\gamma(v). \tag{3.4}$$

Equations (3.1) and (3.4), with γ given by equation (3.3), together constitute the Lorentz transformation. Since the frames move relative to each other along the x-axis, $y' = y$ and $z' = z$.

By defining $\beta = v/c$, and using the coordinate ct instead of t, we can express the Lorentz transformation in a compact form:

$$\boxed{\begin{aligned} ct' &= \gamma\left(ct - \beta x\right), \qquad (3.5)\\ x' &= \gamma\left(x - \beta ct\right), \qquad (3.6)\\ y' &= y, \qquad (3.7)\\ z' &= z, \qquad (3.8) \end{aligned}}$$

where the dependence of γ on v is understood. Thus, we recover the transformation given by the matrix Λ in equation (2.45).

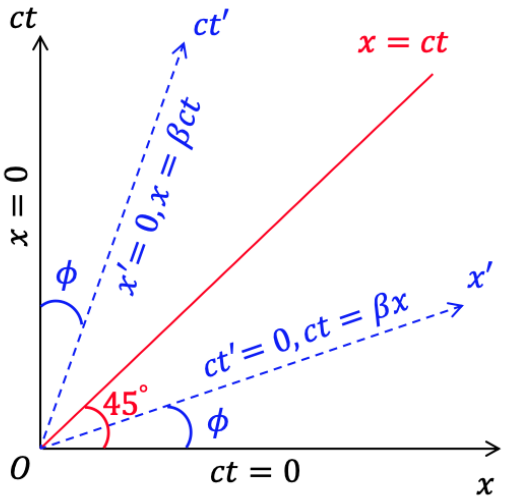
Since frame (S) moves with velocity $-v$ relative to frame (S′), we can also express the transformation from (S′) to (S) by reversing the sign of the velocity. This results in:

$$\boxed{\begin{aligned} ct &= \gamma\left(ct' + \beta x'\right), \qquad (3.9)\\ x &= \gamma\left(x' + \beta ct'\right), \qquad (3.10)\\ y &= y', \qquad (3.11)\\ z &= z', \qquad (3.12) \end{aligned}}$$

where $\beta = v/c$ as defined above. This implies that Λ^{-1} is obtained from Λ by replacing β with $-\beta$ wherever it appears.

3.2 Spacetime diagram

To conveniently represent events and worldlines in a frame (S), we use a diagram where the horizontal axis represents x (considering only one spatial coordinate) and the vertical axis represents time, typically as ct. Each event is indicated by a coordinate pair (ct, x), specifying its position in spacetime. The worldline of an object at rest in frame (S) is depicted as a vertical line.



In this diagram, the worldline of a light ray, $x = ct$, is shown as a line inclined at 45 degrees to both the x- and ct-axes (red line in the diagram). Next, we aim to plot the x' and ct' axes of frame (S′). The equation of the x'-axis in (S′) is $ct' = 0$. In frame (S), according to equation (3.5), this becomes $ct = \beta x$. Similarly, the equation for the ct'-axis, which is $x' = 0$ in (S′), becomes $x = \beta ct$ in (S).

Therefore, in the x-ct diagram, the x'-axis makes an angle $\phi = \tan^{-1}\beta$ with the x-axis, and the ct'-axis makes the same angle ϕ with the ct-axis. They are represented as the blue dashed lines in the diagram, drawn for $v > 0$. The x' and ct' axes are therefore symmetrical with respect to the $x = ct$ line, and close in on it as $\beta \to 1$.
Note that the x' and ct' axes are not perpendicular on the diagram, and that the units along them do not have the same length as along the x and ct axes. For example, according to equations (3.9) and (3.10), the event with coordinates $x' = 1$, $ct' = 0$ is located at $x = \gamma$, $ct = \gamma\beta$ in (S), that is to say at a distance $\gamma\sqrt{1+\beta^2} = \sqrt{(1+\beta^2)/(1-\beta^2)}$ from O on the diagram, which is larger than 1 for $\beta \neq 0$. Distances measured along the primed axes on the diagram therefore cannot be read with the same ruler as along the unprimed axes.

3.3 Relativistic effects on measurements

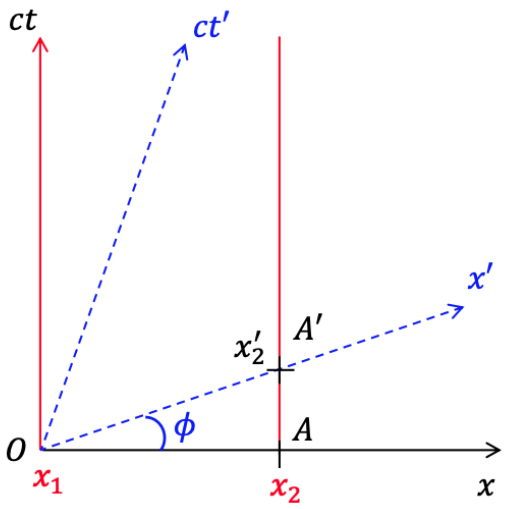
3.3.1 Proper length and length contraction

Proper length refers to the length an object has in a frame (S) in which it is at rest. Suppose this object is a rod lying along the x-axis, between the coordinates x_1 and $x_2 > x_1$. By definition, $L_0 = x_2 - x_1$ is the proper length. Now, consider a frame (S′) moving with velocity v along the x-axis. The length of the rod in (S′) is defined as the distance L between its endpoints when their positions are measured simultaneously in (S′).

Let t' be the time at which the measurement is made in (S′), with the endpoints located at x'_1 and x'_2 such that $x'_2 - x'_1 = L$. Using equation (3.10), we get $x_2 = \gamma\,(x'_2 + \beta ct')$ and $x_1 = \gamma\,(x'_1 + \beta ct')$. Subtracting these two equations yields:

$$\boxed{L = L_0/\gamma.} \tag{3.13}$$

Thus, the length in a frame moving relative to the rest frame of the object along its length is always *shorter* than the proper length. Note that if frame (S′) were moving along the y or z-axes, the length of the object in (S′) would be the same as that in (S).



To understand the significance of this result, let us represent the endpoints of the rod as measured in both frames on the diagram. In (S), assume the rod lies between points O and A at time $t = 0$, so that $x_1 = 0$. Now, assume we make the measurement in (S′) at $t' = 0$. According to equations (3.5) and (3.6), the endpoint which is at O in (S) at $t = 0$ is at $x' = 0$ in (S′) at $t' = 0$. The red ct-axis is the worldline of this endpoint.

The other endpoint must lie on the x'-axis, since we are considering events that are simultaneous in (S′) at $t' = 0$. Since the rod is at rest in (S), this endpoint stays at $x = x_2$, so that its worldline is the vertical red line through A. The event at which it is measured in (S′), denoted as A', is therefore the intersection of this vertical line with the x'-axis, located at $x = x_2$ and $ct = \beta x_2$ in (S). According to equation (3.10) with $t' = 0$, its coordinate in (S′) is $x'_2 = x_2/\gamma$. Therefore, in frame (S), the proper length of the rod is $L_0 = x_2$, measured between the events O and A, while in frame (S′), the length of the rod is $L = x'_2 = x_2/\gamma$, measured between the events O and A'. The two lengths are measured between different pairs of events, which is why they differ.

Note that, on the diagram, A' lies further from O than A, at a distance $x_2\sqrt{1+\beta^2}$, even though $x'_2 < x_2$. As discussed in section 3.2, distances along the x'-axis on the diagram cannot be read with the same ruler as along the x-axis.

3.3.2 Proper time and time dilation

Proper time is the time interval measured by a clock that is present at both of the events being timed. In other words, it is the time elapsed between two events in the frame where both events occur at the same location. Instead of referring to events, we can also consider a particle moving along a worldline in a spacetime diagram, travelling from one event to

another. If we move with the particle along its trajectory, our clock measures the particle's proper time. This can also be described by saying that, *in the rest frame of the particle, where its worldline is vertical (indicating a constant spatial location), the time measured is the proper time.* We denote the proper time as τ.

Suppose two events with the same coordinate x in (S) occur at times t_1 and $t_2 > t_1$. By definition, $t_2 - t_1 = \tau$ is the proper time. In frame (S′), these events occur at times t'_1 and t'_2, with the time interval between them being $\Delta t' = t'_2 - t'_1$. According to equation (3.5), we have $t'_1 = \gamma\,(t_1 - \beta x/c)$ and $t'_2 = \gamma\,(t_2 - \beta x/c)$. Subtracting these equations gives:

$$\boxed{\Delta t' = \gamma\tau.} \tag{3.14}$$

Thus, the time interval between the two events in any other frame is always *longer* than the proper time, measured in the frame where they occur at the same location.

3.4 The invariance of the spacetime interval

In section 2.4.2 of chapter 2, we began by asserting the *invariance of the spacetime interval*:

$$\boxed{\Delta s^2 \equiv -c^2\Delta t^2 + \Delta x^2 + \Delta y^2 + \Delta z^2,} \tag{3.15}$$

where Δx, Δy, Δz and Δt describe the separations in space and time between two events in the four-dimensional spacetime. We then derived the Lorentz transformation directly from this invariant using Lie algebra. Here, we have derived the Lorentz transformation using the two postulates of relativity. It is straightforward to verify that this transformation, given by equations (3.5)-(3.8), ensures $-c^2\Delta t'^2 + \Delta x'^2 + \Delta y'^2 + \Delta z'^2 = -c^2\Delta t^2 + \Delta x^2 + \Delta y^2 + \Delta z^2$, consistent with the spacetime interval being invariant.

A useful geometrical consequence is that all events with the same interval from the origin lie on a hyperbola in the spacetime diagram. For example, the event with coordinates $x' = \ell$, $ct' = 0$ in (S′) satisfies $-c^2t^2 + x^2 = -c^2t'^2 + x'^2 = \ell^2$ in (S), so that it lies on the hyperbola $x^2 - c^2t^2 = \ell^2$, which crosses the x-axis at $x = \ell$. This hyperbola therefore provides the correct calibration of the x'-axis in the diagrams of section 3.2, and similarly the hyperbola $c^2t^2 - x^2 = \ell^2$ for the ct'-axis.

3.4.1 Classification of worldlines

Worldlines can be classified based on the sign of $\mathrm{d}s^2 = -c^2\mathrm{d}t^2 + \mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2$, calculated for an infinitesimal displacement along a worldline. Since this value is invariant, the classification remains the same in all inertial frames.

Timelike worldlines

For massive particles travelling at speeds $v < c$, $\mathrm{d}r \equiv \left(\mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2\right)^{1/2} = v\mathrm{d}t$ and therefore $\mathbf{ds^2 < 0}$. This is known as a timelike interval. In the rest frame of the particle, where the time is the proper time:

$$\boxed{\mathrm{d}s^2 = -c^2\mathrm{d}\tau^2.} \tag{3.16}$$

Note that while $\mathrm{d}s^2$ can be negative, $\mathrm{d}\tau^2$ is always positive. Since $\mathrm{d}s^2$ is invariant, this relation holds in any inertial frame, not only in the rest frame of the particle.
Since $c\,\mathrm{d}t = (c/v)\,\mathrm{d}r > \mathrm{d}r$, the worldline of a massive particle in a one-dimensional spacetime diagram is everywhere inclined at less than 45 degrees to the ct-axis. If the particle's velocity changes over time, the proper time elapsed along its worldline can be obtained from $c^2\mathrm{d}\tau^2 = c^2\mathrm{d}t^2 - \mathrm{d}x^2 - \mathrm{d}y^2 - \mathrm{d}z^2$, which gives $\mathrm{d}\tau = \mathrm{d}t\sqrt{1 - v^2/c^2} = \mathrm{d}t/\gamma(v)$, where v is the instantaneous speed of the particle.

Lightlike (null) worldlines

As mentioned in chapter 2, the front of an electromagnetic wave advances by $\mathrm{d}r = \left(\mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2\right)^{1/2}$ during a time interval $\mathrm{d}t$ such that $\mathrm{d}r = c\mathrm{d}t$. Therefore, for light rays (or massless particles), $\mathbf{ds^2 = 0}$. This is known as a lightlike, or null, interval. Consequently, the worldline of a light ray in a one-dimensional spacetime diagram is a straight line inclined at 45 degrees to the axes.

There is no inertial frame in which a photon is at rest since, according to the second postulate, light travels with velocity c in every inertial frame. Consequently, unlike massive particles, a photon has no rest frame in which a clock could measure its proper time. Along its worldline, $\mathrm{d}\tau = 0$: in this sense, no proper time elapses for a photon. This can also be understood by taking the limit $\gamma \to +\infty$ in equations (3.13) and (3.14), which show that, as seen from a frame moving at a velocity approaching c, lengths contract to nearly zero while time intervals stretch to infinity.

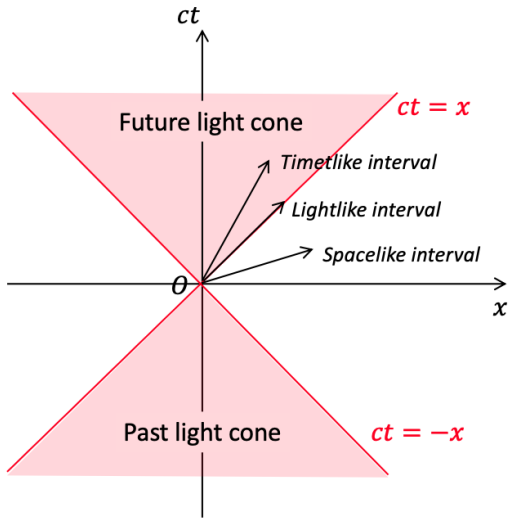
Spacelike worldlines

For hypothetical particles travelling faster than light (known as tachyons, although they are believed not to exist), we have $\mathbf{ds^2 > 0}$. This is known as a spacelike interval. In a one-dimensional spacetime diagram, such a worldline would be inclined at more than 45 degrees to the ct-axis.

Note that there is no frame in which two events separated by a spacelike interval occur at the same location. If such a frame existed, $\mathrm{d}s^2$ would be equal to $-c^2\mathrm{d}\tau^2 \leq 0$, which is not possible since $\mathrm{d}s^2 > 0$. Conversely, there is always a frame in which two such events are simultaneous and their order in time depends on the frame in which they are observed.

3.4.2 Causality in relativity

Consider an event (ct, x, y, z) in the spacetime of an inertial frame. Can this event influence another event with coordinates $[c(t+\Delta t), x+\Delta x, y+\Delta y, z+\Delta z]$? In other words, can a signal sent from (x, y, z) at time t reach the location $(x+\Delta x, y+\Delta y, z+\Delta z)$ before time $t+\Delta t$? This is only possible if $\left(\Delta x^2+\Delta y^2+\Delta z^2\right)^{1/2} \leq c\Delta t$, which means $\Delta s^2 \leq 0$. Therefore, *there cannot be any causal relationship between events separated by spacelike intervals*. This is consistent with the fact, noted above, that the order in time of two such events depends on the frame: if one could influence the other, there would be frames in which the effect precedes its cause.



This diagram illustrates the concepts discussed above. The origin O represents an event in the present ($t = 0$). It can influence any event within the so-called future light cone and can receive information from any event within the past light cone. Communication between events located on the boundaries of the cones must occur at the speed of light. However, the event at O cannot influence or be influenced by events outside of the light cones.

Similar cones can be drawn from any event in spacetime, not just from O.

3.5 Relativistic velocity

3.5.1 Velocity composition

In section 2.4.3 of chapter 2, we discussed that performing successive boosts in different directions results in a combination of boosts and a rotation, rather than another boost. This implies that a boost in a direction such as $\hat{\mathbf{x}}+\hat{\mathbf{y}}$, where $\hat{\mathbf{x}}$ and $\hat{\mathbf{y}}$ are unit vectors along the x- and y-axes, cannot be decomposed into a boost in the x-direction followed by a boost in the y-direction. However, when the boosts are aligned in the same direction, the final result is also a boost.

Consider a frame (S) and a frame (S′) moving with velocity v along the x-axis relative to (S). Now, consider a third frame (S″) moving with velocity v' along the x-axis relative to (S′). To calculate the transformation from (S) to (S″), it is easier to use the rapidity and the transformation matrix Λ_x as defined in equation (2.43). Let $\zeta = \tanh^{-1}(v/c)$ and $\zeta' = \tanh^{-1}(v'/c)$.

The transformation matrix from (S) to (S″) is:

$$\Lambda'' = \begin{pmatrix} \cosh\zeta' & -\sinh\zeta' & 0 & 0 \\ -\sinh\zeta' & \cosh\zeta' & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \cosh\zeta & -\sinh\zeta & 0 & 0 \\ -\sinh\zeta & \cosh\zeta & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \tag{3.17}$$

This results in:

$$\Lambda'' = \begin{pmatrix} \cosh\zeta'' & -\sinh\zeta'' & 0 & 0 \\ -\sinh\zeta'' & \cosh\zeta'' & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \tag{3.18}$$

with $\zeta'' = \zeta + \zeta'$. Hence, the transformation from (S) to (S″) is a boost with velocity v'' along the x-axis such that $v''/c = \tanh\zeta'' = \tanh(\zeta + \zeta')$. This can be expressed as $v''/c = (\tanh\zeta + \tanh\zeta') / (1 + \tanh\zeta \tanh\zeta')$, leading to the relativistic velocity composition formula:

$$\boxed{\frac{v''}{c} = \frac{v/c + v'/c}{1 + vv'/c^2}.} \tag{3.19}$$

While this formula can also be derived using the transformation matrix given by equation (2.45), the calculation is more involved.

Since $|\tanh\zeta''| < 1$ for any finite ζ'', combining two velocities smaller than c always gives a velocity smaller than c: it is the rapidities, not the velocities, that add, as anticipated in section 2.4.3 of chapter 2. In the limit $v, v' \ll c$, we recover the Galilean composition $v'' = v + v'$.

3.5.2 Velocity transformation

We now examine how the velocity components of a particle change when observed from different inertial frames. Let $\mathbf{u}$ be the velocity of the particle in frame (S), and $\mathbf{u}'$ the velocity in frame (S′), which moves with velocity v along the x-axis relative to (S). By definition:

$$u_x = \frac{\mathrm{d}x}{\mathrm{d}t}, \quad u'_x = \frac{\mathrm{d}x'}{\mathrm{d}t'},$$

and similarly for the y and z-components.

From equations (3.5)-(3.8), we have:

$$\begin{aligned} c\mathrm{d}t' &= \gamma\left(c\mathrm{d}t - \beta\mathrm{d}x\right) = \gamma\left(c - \beta u_x\right)\mathrm{d}t, \\ \mathrm{d}x' &= \gamma\left(\mathrm{d}x - \beta c\mathrm{d}t\right) = \gamma\left(u_x - \beta c\right)\mathrm{d}t, \\ \mathrm{d}y' &= \mathrm{d}y = u_y\mathrm{d}t, \\ \mathrm{d}z' &= \mathrm{d}z = u_z\mathrm{d}t. \end{aligned}$$

Dividing these last three equations by the first and using $\beta = v/c$, we get the relativistic velocity transformation[1]:

$$\boxed{u'_x = \frac{u_x - v}{1 - u_x v/c^2}, \quad u'_y = \frac{u_y}{\gamma\left(1 - u_x v/c^2\right)}, \quad u'_z = \frac{u_z}{\gamma\left(1 - u_x v/c^2\right)}.} \tag{3.20}$$

Note that the transverse components u'_y and u'_z differ from u_y and u_z, even though $\mathrm{d}y' = \mathrm{d}y$ and $\mathrm{d}z' = \mathrm{d}z$, because the time intervals $\mathrm{d}t'$ and $\mathrm{d}t$ are different.

3.6 Four-vectors

In this chapter, we have so far focused on Lorentz transformations that are specifically boosts. The following discussion, however, applies to any general Lorentz transformation. To maintain consistency and simplicity, we will nevertheless continue to denote the transformation matrix as Λ, avoiding the introduction of new notation.

3.6.1 Overview and definition

Since time and space are intrinsically linked through Minkowski spacetime, it is natural to define a vector where time and the three spatial coordinates are treated equally. This is called a *4-vector*, and it is represented as a column vector. For example, the position 4-vector is expressed as:

$$X^\mu = \begin{pmatrix} x^0 \\ x^1 \\ x^2 \\ x^3 \end{pmatrix} \tag{3.21}$$

Here, $x^0 = ct$ represents the time component (scaled by c for unit consistency), while $x^1 = x$, $x^2 = y$ and $x^3 = z$ are the spatial components. The convention is to use Greek letters, such as μ or ν, taking the values 0 to 3, to denote the components of 4-vectors, and

[1] To make the first equation easier to remember, it can be expressed as:

$$v_{P/\mathrm{S}'} = \frac{v_{P/\mathrm{S}} + v_{\mathrm{S}/\mathrm{S}'}}{1 + v_{P/\mathrm{S}} v_{\mathrm{S}/\mathrm{S}'}/c^2},$$

where $v_{P/\mathrm{S}} = u_x$ and $v_{P/\mathrm{S}'} = u'_x$ are the velocities of particle P in frame (S) and (S′), respectively, and $v_{\mathrm{S}/\mathrm{S}'} = -v$ is the velocity of frame (S) relative to frame (S′).

Latin letters, such as i, j or k, taking the values 1, 2 and 3, to denote spatial components. We will also use capital letters to denote 4-vectors. Thus, X^μ refers to the 4-vector as a whole, with no specific value assigned to μ, while x^μ refers to the individual components of the 4-vector, with μ taking values from 0 to 3.

By definition, the position 4-vector transforms according to the Lorentz transformation. This means that if X^μ, as defined above, is a position 4-vector in frame (S), then in frame (S′), which moves with velocity v along the x-axis relative to (S), the position 4-vector is $X'^\mu = \Lambda X^\mu$, where Λ is the transformation matrix given by equation (2.45).

As we will see later in this course, various physical quantities transform between frames in the same way as the spacetime position vector. Consequently, these quantities can be combined into a 4-vector that, like the position 4-vector, follows the Lorentz transformation rules. In fact, *this defines a 4-vector: any set of four components that transform according to the Lorentz transformation.* Formally, this means that if A^μ is a 4-vector in frame (S), then in frame (S′) it transforms as:

$$\boxed{A'^\mu = \Lambda A^\mu.} \tag{3.22}$$

This implies that if A^μ in frame (S) represents some physical quantity, then its Lorentz-transformed 4-vector A'^μ represents the same physical quantity in frame (S′).

3.6.2 Minkowski metric: length and scalar product

Previously we noted that, under a Lorentz transformation, the invariant quantity is not the commonly defined squared length of the spacetime interval, $c^2\Delta t^2 + \Delta x^2 + \Delta y^2 + \Delta z^2$, but rather $\Delta s^2 = -c^2\Delta t^2 + \Delta x^2 + \Delta y^2 + \Delta z^2$. In fact, Δs^2 *is* the square of the length of the spacetime interval in Minkowski spacetime, whose geometry is hyperbolic rather than Euclidean. This concept is formally expressed using a metric, which is represented by a matrix.

A metric is a mathematical construct that defines the distance between points in a given space. It is represented by a matrix that encodes information about the shape, angles, and volumes within the space:

- In Newtonian physics, time is universal and space is Euclidean. The metric of space is represented by the 3×3 identity matrix, which yields the commonly defined length of a vector (corresponding to the familiar Pythagorean distance formula). This description is valid in the limit of low velocities and weak gravitational fields.

- In special relativity, time and space are combined into a four-dimensional spacetime, which is flat: the effects of gravity on its geometry are neglected, so that the theory is valid in the limit of weak gravitational fields. It accurately describes phenomena at velocities close to the speed of light. Its laws take their simplest form in inertial frames, but it can also describe accelerated motion, as we will see in chapter 4.

- In general relativity, the metric determines how distances and angles are measured in a spacetime that is curved by the presence of mass and energy. This becomes important in strong gravitational fields. Examples of curved spacetimes include the Schwarzschild metric around a non-rotating spherical mass and the Kerr metric around a rotating mass.

The Minkowski metric, used in special relativity, has components that reflect the inseparability of space and time and defines the interval $\mathrm{d}s^2$, which is preserved under Lorentz transformations. We have encountered this matrix in chapter 2 but we recall it here for easier reference:

$$\boxed{g = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} : \text{ Minkowski metric.}} \tag{3.23}$$

In Euclidean space, the squared length of a 3-vector is given by $\|\mathbf{r}\|^2 = \mathbf{r}^\top I \mathbf{r}$, where I is the identity matrix, serving as the metric. In Minkowski space, the squared length of a 4-vector A^μ is defined analogously, but with the Minkowski metric g:

$$\boxed{(A^\mu)^2 \equiv (A^\mu)^\top g A^\mu.} \tag{3.24}$$

Given:

$$A^\mu = \begin{pmatrix} a^0 \\ a^1 \\ a^2 \\ a^3 \end{pmatrix}, \quad \text{and} \quad (A^\mu)^\top = \left(a^0, a^1, a^2, a^3\right),$$

this results in the squared length of the 4-vector A^μ being:

$$(A^\mu)^2 = -\left(a^0\right)^2 + \left(a^1\right)^2 + \left(a^2\right)^2 + \left(a^3\right)^2. \tag{3.25}$$

Despite the notation, this squared length is not necessarily positive: as for the interval in section 3.4.1, a 4-vector is called timelike, lightlike (null) or spacelike if $(A^\mu)^2$ is negative, zero or positive, respectively.

Based on this definition of the length, we have $\Delta s^2 = (\Delta X^\mu)^2$. This quantity remains invariant for any interval between two events, including the interval between the origin O (where both t and the spatial coordinates are zero) and a point with arbitrary coordinates. Consequently, $(X^\mu)^2$ is invariant for any position 4-vector.

Using this invariant, we can recover the relationship between the matrices g and Λ that we derived in chapter 2 (see eq. [2.33]) and used to derive the general expression for Λ. We recall this result here for easy reference. If we transform into a frame (S′), the invariance of the square of the length means that $(X'^{\mu})^2 = (X^{\mu})^2$, that is to say $(X'^{\mu})^{\top} g X'^{\mu} = (X^{\mu})^{\top} g X^{\mu}$. Using $X'^{\mu} = \Lambda X^{\mu}$, this yields $(X^{\mu})^{\top} \Lambda^{\top} g \Lambda X^{\mu} = (X^{\mu})^{\top} g X^{\mu}$. Since this must be satisfied for any position 4-vector X^{μ}, and both $\Lambda^{\top} g \Lambda$ and g are symmetric matrices, it follows that:

$$\boxed{\Lambda^{\top} g \Lambda = g.} \tag{3.26}$$

Now, consider an arbitrary 4-vector A^{μ}. In frame (S′), it is given by $A'^{\mu} = \Lambda A^{\mu}$. This yields $(A'^{\mu})^2 \equiv (A'^{\mu})^{\top} g A'^{\mu} = (\Lambda A^{\mu})^{\top} g \Lambda A^{\mu} = (A^{\mu})^{\top} \Lambda^{\top} g \Lambda A^{\mu} = (A^{\mu})^{\top} g A^{\mu} \equiv (A^{\mu})^2$. Thus, we conclude that:

Any 4-vector $\mathbf{A^{\mu}}$ has an invariant squared length $\mathbf{(A^{\mu})^2}$.

We now define the scalar product of two 4-vectors A^{μ} and B^{μ} as:

$$\boxed{A^{\mu} \cdot B^{\mu} \equiv (A^{\mu})^{\top} g B^{\mu}.} \tag{3.27}$$

It can easily be verified that:

$$A^{\mu} \cdot B^{\mu} = -a^0 b^0 + a^1 b^1 + a^2 b^2 + a^3 b^3, \tag{3.28}$$

where b^{μ} are the components of B^{μ}. This scalar product is commutative, as it satisfies $A^{\mu} \cdot B^{\mu} = B^{\mu} \cdot A^{\mu}$. Under a Lorentz transformation, the scalar product transforms as $A'^{\mu} \cdot B'^{\mu} = (\Lambda A^{\mu}) \cdot (\Lambda B^{\mu}) = (\Lambda A^{\mu})^{\top} g \Lambda B^{\mu}$. Using equation (3.26), this gives $A'^{\mu} \cdot B'^{\mu} = A^{\mu} \cdot B^{\mu}$. Thus,

The scalar product of two 4-vectors is invariant.

While matrix notation is convenient for deriving the Lorentz transformation via Lie algebra, index notation, which explicitly handles matrix indices, is generally more suitable for working with the equations of special (and general) relativity. This approach will be discussed in the following section.

3.6.3 Index notation, covariant and contravariant four-vectors

In the previous section, we defined the scalar product (and length) in matrix form (eq. [3.27]). Now, we introduce a more compact and convenient expression:

$$A^{\mu} \cdot B^{\mu} = g_{\mu\nu} a^{\mu} b^{\nu}, \tag{3.29}$$

where $g_{\mu\nu}$ are the coefficients of the matrix g. As usual, the first index μ indicates a row, and the second index ν indicates a column.

The above expression utilises the Einstein summation convention for repeated indices, which means it is a shorthand for:

$$A^\mu \cdot B^\mu = \sum_{\mu=0}^{3} \sum_{\nu=0}^{3} g_{\mu\nu} a^\mu b^\nu.$$

It is easy to verify that equations (3.27) and (3.29) are equivalent.

Covariant and contravariant four-vectors

To simplify the notation further, for any column 4-vector A^μ, we define the corresponding **row 4-vector** denoted A_μ, with components a_μ given by:

$$\boxed{a_\mu = g_{\mu\nu} a^\nu.} \tag{3.30}$$

For example, given the column 4-vector:

$$A^\mu = \begin{pmatrix} a^0 \\ a^1 \\ a^2 \\ a^3 \end{pmatrix},$$

this transformation yields the row 4-vector:

$$\boxed{A_\mu \equiv (gA^\mu)^\top = (a_0, a_1, a_2, a_3) = \left(-a^0, a^1, a^2, a^3\right).} \tag{3.31}$$

The expression (3.30) provides the relationship between the components of A_μ and those of A^μ. However, we often write $A_\mu = g_{\mu\nu} A^\nu$, mixing index and vector notation, as a shorthand for equation (3.30). This practice allows us to use the generic name of the 4-vector as a proxy notation for its components, bypassing the need to explicitly define those components.

Just as we derived A_μ from A^μ, we can reverse the process. Let $g^{\mu\nu}$ represent the coefficients of the matrix g^{-1}:

$$\boxed{g^{\mu\alpha} g_{\alpha\nu} = \delta^\mu_\nu,} \tag{3.32}$$

where δ^μ_ν is the Kronecker symbol (equal to 1 if $\mu = \nu$, and 0 otherwise). Even though $g^{-1} = g$, we use distinct notations for the coefficients to enhance clarity. Specifically, $g^{00} = g_{00} = -1$, while all other non-zero components are equal to 1. Using equation (3.30), we obtain $g^{\mu\nu} a_\nu = g^{\mu\nu} g_{\nu\xi} a^\xi = \delta^\mu_\xi a^\xi$. Therefore:

$$\boxed{a^\mu = g^{\mu\nu} a_\nu.} \tag{3.33}$$

This expression can also be derived using equation (3.31), which gives $A^\mu = g^{-1} (A_\mu)^\top$. Once again, the notation $A^\mu = g^{\mu\nu} A_\nu$ is often used as a shorthand for equation (3.33).

The column 4-vector A^μ is known as a **contravariant** vector, while the row 4-vector A_μ is known as a **covariant** vector. The operations described by equations (3.30) and (3.33) are commonly referred to as 'lowering' and 'raising' indices, respectively.

Transformation of a covariant four-vector

We have defined a contravariant 4-vector A^μ as one that transforms according to the Lorentz transformation, such that in frame (S′) we have $A'^\mu = \Lambda A^\mu$. Additionally, we have referred to the covariant vector A_μ as a 4-vector. To confirm that A_μ is indeed a 4-vector, we need to examine its transformation properties.
Recall that A_μ is a row vector. According to equation (3.31), it can be expressed in matrix notation as:

$$A_\mu = (gA^\mu)^\top = (A^\mu)^\top g^\top.$$

The Minkowski metric satisfies $g^\top = g$, as does any metric, which is symmetric by definition. We nevertheless keep $g^\top$ in the derivation below, to make each step explicit. When transforming to frame (S′), the expression above becomes $A'_\mu = (A'^\mu)^\top g^\top = (\Lambda A^\mu)^\top g^\top = (A^\mu)^\top \Lambda^\top g^\top$. Taking the transpose of equation (3.26) yields $g^\top = \Lambda^\top g^\top \Lambda$, which implies $\Lambda^\top g^\top = g^\top \Lambda^{-1}$. Substituting this into the expression for A'_μ, we obtain $A'_\mu = (A^\mu)^\top g^\top \Lambda^{-1}$, leading to:

$$\boxed{A'_\mu = A_\mu \Lambda^{-1}.} \tag{3.34}$$

Equation (3.34) shows that a covariant 4-vector does not transform with Λ, as a contravariant 4-vector does, but with Λ^{-1}, acting from the right. Since Lorentz transformations form a group, Λ^{-1} is itself a Lorentz transformation: for a boost, it corresponds to a boost in the opposite direction. The transformation of A_μ is therefore entirely determined by the Lorentz transformation, so that A_μ is also a 4-vector, in a broader sense than equation (3.22). These two different transformation laws are precisely what distinguishes covariant from contravariant 4-vectors and, as we will see below, they ensure that the scalar product $A_\mu B^\mu$ is invariant.

Squared length of a covariant four-vector

The squared length of A^μ is defined by equation (3.24). Using equation (3.31), we have $gA^\mu = (A_\mu)^\top$ and $(A^\mu)^\top = A_\mu \left(g^{-1}\right)^\top$. Substituting these into equation (3.24), we obtain:

$$(A^\mu)^2 = A_\mu \left(g^{-1}\right)^\top (A_\mu)^\top.$$

Since this expression is now in terms of A_μ, it provides a definition for $(A_\mu)^2$, which is thus equal to $(A^\mu)^2$:

$$(A_\mu)^2 \equiv A_\mu \left(g^{-1}\right)^\top (A_\mu)^\top = \left(A_\mu\, g^{-1}\, (A_\mu)^\top\right)^\top. \tag{3.35}$$

Since this is equal to $(A^\mu)^2$, it represents an invariant quantity.

Scalar products using index notation

Using equation (3.30), we can replace $g_{\mu\nu}b^\nu$ with b_μ in equation (3.29). This yields $a^\mu b_\mu$, which, in matrix notation, is the product of the row vector B_μ by the column vector A^μ: $A^\mu \cdot B^\mu = B_\mu A^\mu$. Alternatively, this can be derived directly in matrix notation by writing $A^\mu \cdot B^\mu = (A^\mu)^\top gB^\mu = (A^\mu)^\top gg^{-1}(B_\mu)^\top = (B_\mu A^\mu)^\top = B_\mu A^\mu$, where the final equality follows from the fact that the transpose of a scalar is the scalar itself.

Since the Minkowski metric is symmetric, meaning $g_{\mu\nu} = g_{\nu\mu}$, we can also replace $g_{\mu\nu}a^\mu$ with a_ν instead in equation (3.29), which yields $A^\mu \cdot B^\mu = A_\mu B^\mu$.

We summarize these results as:

$$\boxed{A^\mu \cdot B^\mu = B_\mu A^\mu = A_\mu B^\mu = \sum_{\nu=0}^{3} a^\nu b_\nu = \sum_{\nu=0}^{3} a_\nu b^\nu.} \tag{3.36}$$

The first term serves as a formal notation for the scalar product, the second and third terms represent the standard product of a row vector by a column vector, and the fourth and fifth terms are expressed in terms of the individual components of these vectors.

Note that the product of a column vector with another column vector, such as $A^\mu B^\mu$, is not defined according to the matrix and vector multiplication rules we have used so far. The same is true for $A_\mu B_\mu$. These products will be dealt with using tensor operations.

Naturally, the above expression remains valid when $B^\mu = A^\mu$, thus $A^\mu \cdot A^\mu = A_\mu A^\mu$.

The process of summing over a pair of one upper (contravariant) and one lower (covariant) index is called *index contraction*. Performing index contraction between two 4-vectors results in a scalar.

Conversely, **if A_μ is a covariant 4-vector and the scalar product $A_\mu B^\mu$ remains invariant under transformation for any choice of A_μ, then B^μ must be a contravariant 4-vector**. Indeed, the invariance of the scalar product requires $A'_\mu B'^\mu = A_\mu B^\mu$. Using equation (3.34), this leads to $A_\mu \Lambda^{-1} B'^\mu = A_\mu B^\mu$, or $A_\mu \left(\Lambda^{-1} B'^\mu - B^\mu\right) = 0$. Since this relation must hold for any choice of A_μ, the expression in parentheses must vanish, implying $B'^\mu = \Lambda B^\mu$. This demonstrates that B^μ is a contravariant 4-vector. Similarly, **if B^μ is a contravariant 4-vector and $A_\mu B^\mu$ is invariant for any choice of B^μ, then A_μ must be a covariant 4-vector**.

Until now, we have carefully distinguished between row and column 4-vectors. However, except when using matrix notation to multiply 4-vectors by matrices, this distinction is not essential, just as it is unnecessary to specify whether a spatial vector in 3-dimensional Euclidean space is a row or a column vector. Therefore, **from this point onwards, we will represent all covariant and contravariant 4-vectors as row vectors**, understanding that $A_\mu B^\mu$ denotes the sum $a_\mu b^\mu$ of the products of their components, analogous to the scalar product of two spatial vectors. By using indices to refer to the components of vectors and matrices, we eliminate the need for explicit matrix notation. This approach will also prove to be much more effective for handling tensors.

3.6.4 Covariant and contravariant derivatives

In this section, we extend the various derivative operators defined in Euclidean space to Minkowski spacetime.

Gradient

Consider a scalar function $f\left(x^0, x^1, x^2, x^3\right)$ that depends on the spacetime coordinates in frame (S). The change in f between two infinitesimally separated points in spacetime is given by the differential:

$$\mathrm{d}f_S = \frac{\partial f}{\partial x^\mu}\mathrm{d}x^\mu. \tag{3.37}$$

In frame (S′), the same scalar field is described by a function f' of the coordinates $\left(x'^0, x'^1, x'^2, x'^3\right)$, such that $f'\left(x'^\mu\right) = f\left(x^\mu\right)$ at every event: the value of a scalar at a given event is the same in all frames, although the function f' generally differs from f. Its differential is:

$$\mathrm{d}f'_{\mathrm{S}'} = \frac{\partial f'}{\partial x'^\mu}\mathrm{d}x'^\mu. \tag{3.38}$$

Since x'^μ is itself a function of the coordinates in (S), we have:

$$\mathrm{d}x'^\mu = \frac{\partial x'^\mu}{\partial x^\nu}\mathrm{d}x^\nu, \tag{3.39}$$

which yields:

$$\mathrm{d}f'_{\mathrm{S}'} = \frac{\partial f'}{\partial x'^\mu}\frac{\partial x'^\mu}{\partial x^\nu}\mathrm{d}x^\nu = \frac{\partial f}{\partial x^\nu}\mathrm{d}x^\nu. \tag{3.40}$$

The second equality follows from the chain rule, since $f\left(x^\mu\right) = f'\left(x'^\mu\left(x^\nu\right)\right)$. Therefore $\mathrm{d}f'_{\mathrm{S}'} = \mathrm{d}f_S$, indicating that the differential of a scalar function is invariant.

Equation (3.37) expresses the fact that the scalar product of the contravariant 4-vector $\mathrm{d}X^\mu$, whose components are $\mathrm{d}x^\mu$, with the vector whose components are $\partial f/\partial x^\mu$, is invariant. Thus, this vector is a **covariant 4-vector, denoted $\boldsymbol{\partial_\mu f}$**. More generally, we define the gradient operator as a covariant 4-vector:

$$\partial_\mu = \left(\frac{\partial}{\partial x^0}, \frac{\partial}{\partial x^1}, \frac{\partial}{\partial x^2}, \frac{\partial}{\partial x^3}\right) = \left(\frac{1}{c}\frac{\partial}{\partial t}, \frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z}\right). \tag{3.41}$$

Introducing the standard gradient operator in 3-dimensional space, $\boldsymbol{\nabla}$, this can be written in a more compact form as:

$$\boxed{\partial_\mu = \left(\frac{1}{c}\frac{\partial}{\partial t}, \boldsymbol{\nabla}\right).} \tag{3.42}$$

Since $\partial_\mu f$ is a covariant 4-vector, in frame (S′) it becomes $\partial'_\mu f' = (\partial_\mu f)\,\Lambda^{-1}$, where ∂'_μ involves derivatives with respect to the coordinates in (S′).
More generally, we can write:

$$\boxed{\partial'_\mu = \partial_\mu \Lambda^{-1}.} \tag{3.43}$$

Similarly, we can view f as a function of the coordinates (x_0, x_1, x_2, x_3), leading to the **contravariant 4-vector, denoted $\boldsymbol{\partial^\mu f}$**, whose components are $\partial f/\partial x_\mu$. This now defines the gradient operator as a contravariant 4-vector:

$$\partial^\mu = \left(\frac{\partial}{\partial x_0}, \frac{\partial}{\partial x_1}, \frac{\partial}{\partial x_2}, \frac{\partial}{\partial x_3}\right) = \left(-\frac{1}{c}\frac{\partial}{\partial t}, \frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z}\right), \tag{3.44}$$

or, in a more compact form:

$$\boxed{\partial^\mu = \left(-\frac{1}{c}\frac{\partial}{\partial t}, \boldsymbol{\nabla}\right).} \tag{3.45}$$

In this case, $\partial'^\mu f' = \Lambda \partial^\mu f$ and, more generally:

$$\boxed{\partial'^\mu = \Lambda \partial^\mu.} \tag{3.46}$$

Although ∂_μ and ∂^μ are operators, they can be manipulated as if they were ordinary 4-vectors. This allows us to raise and lower their indices with the metric and write $\partial_\mu = g_{\mu\nu}\partial^\nu$, or $\partial^\mu = g^{\mu\nu}\partial_\nu$, where the 4-vector name serves as a proxy notation for its components, as discussed earlier.

The results above hold for any Lorentz transformation. In the specific case of a boost, $\partial'^\mu = \Lambda \partial^\mu$ can be explicitly written as:

$$\begin{pmatrix} -\frac{1}{c}\frac{\partial}{\partial t'} \\ \frac{\partial}{\partial x'} \\ \frac{\partial}{\partial y'} \\ \frac{\partial}{\partial z'} \end{pmatrix} = \begin{pmatrix} \gamma & -\beta\gamma & 0 & 0 \\ -\beta\gamma & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} -\frac{1}{c}\frac{\partial}{\partial t} \\ \frac{\partial}{\partial x} \\ \frac{\partial}{\partial y} \\ \frac{\partial}{\partial z} \end{pmatrix}. \tag{3.47}$$

This can be verified by directly calculating the derivatives using $X'^\mu = \Lambda X^\mu$.

Divergence

So far, we have considered cases where the 4-vector gradient operator acts on a scalar function, yielding a 4-vector. This can be expressed by stating that the product of a gradient 4-vector and a scalar function results in a 4-vector. Now, we consider the scalar product of the gradient 4-vector with another 4-vector, which results in a scalar. We start with a contravariant 4-vector A^μ, whose components a^μ depend on the spacetime coordinates. The divergence of this 4-vector is defined as:

$$\partial^\mu \cdot A^\mu = \partial_\mu A^\mu = \frac{\partial a^\mu}{\partial x^\mu} = \frac{\partial a^0}{\partial x^0} + \frac{\partial a^1}{\partial x^1} + \frac{\partial a^2}{\partial x^2} + \frac{\partial a^3}{\partial x^3} = \frac{1}{c}\frac{\partial a^0}{\partial t} + \frac{\partial a^1}{\partial x} + \frac{\partial a^2}{\partial y} + \frac{\partial a^3}{\partial z}.$$

Introducing $\mathbf{A} = \left(a^1, a^2, a^3\right)$ as the spatial part of A^μ, i.e., $A^\mu = \left(a^0, \mathbf{A}\right)$, we can write the divergence as:

$$\boxed{\partial^\mu \cdot A^\mu = \partial_\mu A^\mu = \frac{1}{c}\frac{\partial a^0}{\partial t} + \boldsymbol{\nabla} \cdot \mathbf{A}.} \tag{3.48}$$

The scalar product can also be written as:

$$\partial^\mu \cdot A^\mu = \partial^\mu A_\mu = \frac{\partial a_\mu}{\partial x_\mu} = \frac{\partial a_0}{\partial x_0} + \frac{\partial a_1}{\partial x_1} + \frac{\partial a_2}{\partial x_2} + \frac{\partial a_3}{\partial x_3} = -\frac{1}{c}\frac{\partial a_0}{\partial t} + \frac{\partial a_1}{\partial x} + \frac{\partial a_2}{\partial y} + \frac{\partial a_3}{\partial z}.$$

In the Minkowski metric, A_μ has the same spatial components as A^μ, leading to:

$$\boxed{\partial^\mu \cdot A^\mu = \partial^\mu A_\mu = -\frac{1}{c}\frac{\partial a_0}{\partial t} + \boldsymbol{\nabla} \cdot \mathbf{A}.} \tag{3.49}$$

Since $a_0 = -a^0$, this is identical to the previous expression.

The divergence remains invariant as it represents the scalar product of two 4-vectors. Consequently, a conservation law expressed as a vanishing 4-divergence, $\partial_\mu A^\mu = 0$ (a continuity equation), **holds in all inertial frames** if it holds in one.

D'Alembertian

Now, consider the effect of taking the divergence of the 4-vector $\partial^\mu f$. We denote the resulting operator (known as d'Alembertian, or wave operator) by the symbol $\Box$, and we have:

$$\Box f = \partial^\mu \cdot \partial^\mu f = \partial_\mu \partial^\mu f = -\frac{1}{c^2}\frac{\partial^2 f}{\partial t^2} + \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} + \frac{\partial^2 f}{\partial z^2}. \tag{3.50}$$

We can easily verify that $\partial_\mu \partial^\mu f = \partial^\mu \partial_\mu f$.

Introducing the standard Laplacian operator in 3-dimensional space, ∇^2, we obtain:

$$\boxed{\Box = \partial^\mu \cdot \partial^\mu = \partial_\mu \partial^\mu = \partial^\mu \partial_\mu = -\frac{1}{c^2}\frac{\partial^2}{\partial t^2} + \nabla^2.} \tag{3.51}$$

Since **the d'Alembertian operator** $\partial_\mu \partial^\mu$ represents the squared length of a 4-vector, it remains **invariant under Lorentz transformations**. This invariance ensures that **the wave equation retains its form in any inertial frame**. Specifically, if a scalar function f satisfies the wave equation $\Box f = 0$ in frame (S), then the function f' describing the same scalar field in frame (S′) satisfies $\Box' f' = 0$, where $\Box'$ involves derivatives with respect to the coordinates x'^μ.

3.6.5 Transformations of four-vectors in index notation

Transformations using Lorentz matrices

We have introduced index notation for 4-vectors and the metric matrix g, but not yet for the Lorentz transformation matrix. This section addresses that. Consider a contravariant 4-vector A^μ. In matrix notation, it transforms as $A'^\mu = \Lambda A^\mu$. In previous courses, we referred to the coefficients of a matrix M as M_{ij}, where i indicates a row and j indicates a column. However, using only lower indices to label the coefficients of Λ would imply that both indices are covariant. While this might not pose a problem for 4-vectors, it can lead to confusion when dealing with tensors, as we will see later in the course.
Thus, in index notation, the relationship between A'^μ and A^μ is written as:

$$\boxed{A'^\mu = \Lambda^\mu{}_\nu A^\nu.} \tag{3.52}$$

Here, A^ν serves as a proxy notation for its components a^ν. Thus, the right-hand side is the sum of the products of the coefficients of Λ and the components of A^ν. Since A^ν is a column 4-vector, consistency with matrix notation requires that the contravariant (upper) index μ in $\Lambda^\mu{}_\nu$ corresponds to a row, while the covariant (lower) index ν corresponds to a column[2]. Staggering the indices clarifies the notation.

To make the notation clear, we explicitly write:

$$\Lambda = \begin{pmatrix} \Lambda^0{}_0 & \Lambda^0{}_1 & \Lambda^0{}_2 & \Lambda^0{}_3 \\ \Lambda^1{}_0 & \Lambda^1{}_1 & \Lambda^1{}_2 & \Lambda^1{}_3 \\ \Lambda^2{}_0 & \Lambda^2{}_1 & \Lambda^2{}_2 & \Lambda^2{}_3 \\ \Lambda^3{}_0 & \Lambda^3{}_1 & \Lambda^3{}_2 & \Lambda^3{}_3 \end{pmatrix}. \tag{3.53}$$

[2]For the metric, both indices are either covariant ($g_{\mu\nu}$) or contravariant ($g^{\mu\nu}$). Since the metric is symmetric, the order of the two indices does not matter, so there is no need to stagger them. Note that a metric with one upper and one lower index would simply be the Kronecker symbol: $g^{\mu\alpha} g_{\alpha\nu} = \delta^\mu_\nu$, according to equation (3.32).

The matrix Λ has dimension 4×4, but by contracting the lower index (multiplying by A^ν and summing over ν), we obtain a 4-vector.

With this notation, we can express the product of the matrices $g\Lambda$ as $g_{\lambda\kappa}\Lambda^\kappa{}_\nu$. Consequently, equation (3.26) becomes:

$$\boxed{\Lambda^\lambda{}_\mu g_{\lambda\kappa}\Lambda^\kappa{}_\nu = g_{\mu\nu}.} \tag{3.54}$$

Consider now a covariant 4-vector A_μ. It transforms as follows:

$$A'_\mu = g_{\mu\kappa}A'^\kappa = g_{\mu\kappa}\Lambda^\kappa{}_\lambda A^\lambda = g_{\mu\kappa}\Lambda^\kappa{}_\lambda g^{\lambda\nu}A_\nu.$$

This can be written as:

$$\boxed{A'_\mu = \Lambda_\mu{}^\nu A_\nu \quad \text{where we define} \quad \Lambda_\mu{}^\nu = g_{\mu\kappa}\Lambda^\kappa{}_\lambda g^{\lambda\nu}.} \tag{3.55}$$

By multiplying both sides of the second equation by $g^{\rho\mu}$ and $g_{\nu\pi}$ (and summing over μ and ν), and using equation (3.32), we obtain:

$$g^{\rho\mu}\Lambda_\mu{}^\nu g_{\nu\pi} = g^{\rho\mu}g_{\mu\kappa}\Lambda^\kappa{}_\lambda g^{\lambda\nu}g_{\nu\pi} = \delta^\rho_\kappa\Lambda^\kappa{}_\lambda\delta^\lambda_\pi = \Lambda^\rho{}_\pi. \tag{3.56}$$

This notation extends the concept of lowering and raising indices from 4-vectors to matrices.
From equation (3.34), we also have $A'_\mu = A_\nu \left(\Lambda^{-1}\right)^\nu{}_\mu$. Comparing this with equation (3.55) implies $\Lambda_\mu{}^\nu = \left(\Lambda^{-1}\right)^\nu{}_\mu$. Using the definition of $\Lambda_\mu{}^\nu$ from equation (3.55), we can verify that $\Lambda_\mu{}^\nu\Lambda^\mu{}_\zeta = g_{\mu\kappa}\Lambda^\kappa{}_\lambda g^{\lambda\nu}\Lambda^\mu{}_\zeta = g_{\zeta\lambda}g^{\lambda\nu} = \delta^\nu_\zeta$, where we have used equation (3.54).
We summarise these results:

$$\boxed{\Lambda_\mu{}^\nu = \left(\Lambda^{-1}\right)^\nu{}_\mu \quad \text{or equivalently} \quad \Lambda_\mu{}^\nu\Lambda^\mu{}_\zeta = \delta^\nu_\zeta.} \tag{3.57}$$

Transformations using the Jacobian matrix

Since $\mathrm{d}X^\mu$ is a contravariant 4-vector, it transforms as $\mathrm{d}X'^\mu = \Lambda^\mu{}_\nu \mathrm{d}X^\nu$. Using equation (3.39), this implies that:

$$\boxed{\frac{\partial x'^\mu}{\partial x^\nu} \equiv \partial_\nu x'^\mu = \Lambda^\mu{}_\nu,} \tag{3.58}$$

meaning that **the Lorentz transformation matrix Λ is the Jacobian matrix of the transformation** from the coordinates x^ν to x'^μ.

We can thus define a 4-vector A^μ as any set of four components that transform according to:

$$\boxed{A'^\mu = \frac{\partial x'^\mu}{\partial x^\nu} A^\nu = A^\nu \partial_\nu x'^\mu.} \tag{3.59}$$

Equation (3.59) provides the **most general definition of a 4-vector**, as it encompasses all possible transformations from one set of coordinates x^μ to another set x'^μ. The (linear) Lorentz transformation is merely a special case applicable to special relativity. In contrast, general relativity allows for transformations that are not necessarily linear and can involve any smooth mapping between coordinate systems. In that case, the Jacobian matrix depends on the position in spacetime, so that a 4-vector is defined at a given event, with the Jacobian evaluated at that event. Note also that the displacement $\mathrm{d}X^\mu$ always transforms according to equation (3.59), whereas the position X^μ itself does so only when the transformation is linear: the position 4-vector is a special feature of special relativity.

The inverse transformation gives $\mathrm{d}X^\mu = \left(\Lambda^{-1}\right)^\mu{}_\nu \, \mathrm{d}X'^\nu = \Lambda_\nu{}^\mu \mathrm{d}X'^\nu$. This implies that:

$$\boxed{\frac{\partial x^\mu}{\partial x'^\nu} \equiv \partial'_\nu x^\mu = \Lambda_\nu{}^\mu.} \tag{3.60}$$

From equation (3.55), we have $A'_\mu = \Lambda_\mu{}^\nu A_\nu$. Therefore, equation (3.60) provides the transformation of a covariant vector as:

$$\boxed{A'_\mu = \frac{\partial x^\nu}{\partial x'^\mu} A_\nu = A_\nu \partial'_\mu x^\nu.} \tag{3.61}$$

In principle, we could define coordinate transformations in a covariant form involving terms like $\partial x'_\mu / \partial x_\nu$. However, this is not commonly used.

Invariance of the four-volume element

Since Λ is the Jacobian matrix of the transformation from the coordinates x^ν to x'^μ (eq. [3.58]), the four-volume element $\mathrm{d}^4x \equiv \mathrm{d}x^0 \, \mathrm{d}x^1 \, \mathrm{d}x^2 \, \mathrm{d}x^3 = c \, \mathrm{d}t \, \mathrm{d}^3x$ transforms, by the usual rule for changing variables in a multiple integral, as:

$$\mathrm{d}^4x' = |\det \Lambda| \, \mathrm{d}^4x = \mathrm{d}^4x, \tag{3.62}$$

where we have used $\det \Lambda = \pm 1$ (section 2.4.3). The four-volume element is therefore invariant under Lorentz transformations. This result will be used in chapter 5 to construct Lorentz-invariant actions for fields.

Chapter 4

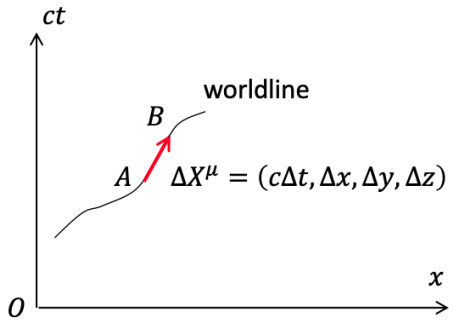
Dynamics and conservation laws

In this chapter, the spacetime position 4-vector is represented as $X^\mu = (ct, \mathbf{r})$, where $\mathbf{r} = (x, y, z)$ is the usual spatial 3-vector. We also use the notation $X^\mu = \left(x^0, x^1, x^2, x^3\right)$ when convenient.

4.1 Four-velocity

4.1.1 Derivation of the four-velocity

We begin by constructing a 4-vector for velocity, extending the concept from 3-dimensional space to spacetime.



Consider two points, A and B, along a worldline in the spacetime diagram, separated by the displacement $\Delta X^\mu = (c\Delta t, \Delta \mathbf{r})$. The ordinary velocity at point A is given by:

$$\mathbf{u} = \lim_{\Delta t \to 0} \frac{\Delta \mathbf{r}}{\Delta t} = \frac{\mathrm{d}\mathbf{r}}{\mathrm{d}t}.$$

However, $\Delta X^\mu / \Delta t$ is not a 4-vector, since Δt is not invariant, so we cannot directly construct a 4-vector from $\mathbf{u}$. Instead, we define the velocity 4-vector, or 4-velocity, U^μ as:

$$\boxed{U^\mu = \frac{\mathrm{d}X^\mu}{\mathrm{d}\tau},} \tag{4.1}$$

where τ is the *proper time* along the worldline. Since $\mathrm{d}\tau$ is invariant and $\mathrm{d}X^\mu$ is a 4-vector, U^μ is indeed a 4-vector. Explicitly, this is written as:

$$U^\mu = \left(c\frac{\mathrm{d}t}{\mathrm{d}\tau}, \frac{\mathrm{d}x}{\mathrm{d}\tau}, \frac{\mathrm{d}y}{\mathrm{d}\tau}, \frac{\mathrm{d}z}{\mathrm{d}\tau}\right) = \left(c\frac{\mathrm{d}t}{\mathrm{d}\tau}, \mathbf{u}\frac{\mathrm{d}t}{\mathrm{d}\tau}\right).$$

As discussed in section 3.4.1, a massive particle moves along a timelike worldline, corresponding to $\mathrm{d}s^2 = -c^2\mathrm{d}t^2 + \mathrm{d}r^2 = -c^2\mathrm{d}\tau^2$. This gives: $\mathrm{d}t/\mathrm{d}\tau = \mathrm{d}t/\sqrt{\mathrm{d}t^2 - \mathrm{d}r^2/c^2} = 1/\sqrt{1-u^2/c^2}$, which can be written as:

$$\boxed{\frac{\mathrm{d}t}{\mathrm{d}\tau} = \gamma_u,} \tag{4.2}$$

where $\gamma_u = 1/\sqrt{1-u^2/c^2}$. Therefore:

$$\boxed{U^\mu = (\gamma_u c, \gamma_u \mathbf{u})\,.} \tag{4.3}$$

Since U^μ is a 4-vector, its squared length is an invariant. In the rest frame of the particle, where $\mathbf{u} = \mathbf{0}$, we have $U^\mu = (c, \mathbf{0})$, resulting in:

$$\boxed{(U^\mu)^2 = -c^2.} \tag{4.4}$$

This can also be checked directly from equation (4.3): $(U^\mu)^2 = -\gamma_u^2 c^2 + \gamma_u^2 u^2 = -c^2$.

The definition of 4-velocity cannot be extended to massless particles, for which $\mathrm{d}\tau = 0$. These particles travel at $u = c$, leading to $\gamma_u \to +\infty$, which makes U^μ undefined. For such particles, we instead define a 4-momentum, as will be discussed below.

4.1.2 Velocity transformation using four-vectors

Consider a particle with velocity $\mathbf{u}$ in frame (S) and velocity $\mathbf{u}'$ in frame (S′), where (S′) moves with velocity v along the x-axis relative to (S). The 4-velocity in (S′) is given by $U'^\mu = \Lambda U^\mu$, which can be explicitly written as:

$$\begin{pmatrix} \gamma_{u'}c \\ \gamma_{u'}u'_x \\ \gamma_{u'}u'_y \\ \gamma_{u'}u'_z \end{pmatrix} = \begin{pmatrix} \gamma_v & -\beta\gamma_v & 0 & 0 \\ -\beta\gamma_v & \gamma_v & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \gamma_u c \\ \gamma_u u_x \\ \gamma_u u_y \\ \gamma_u u_z \end{pmatrix}, \tag{4.5}$$

where $\beta = v/c$ and the subscript 'v' on γ indicates that it is related to the velocity of frame (S′). This transformation yields the following equations:

$$\begin{aligned} \gamma_{u'}c &= \gamma_v\gamma_u\,(c - \beta u_x)\,, \\ \gamma_{u'}u'_x &= \gamma_v\gamma_u\,(-\beta c + u_x)\,, \\ \gamma_{u'}u'_y &= \gamma_u u_y, \\ \gamma_{u'}u'_z &= \gamma_u u_z. \end{aligned}$$

Dividing the last three equations by the first one, we can recover the components of $\mathbf{u}'$ calculated in section 3.5.2. The first equation also gives the transformation of the Lorentz factor of the particle:

$$\gamma_{u'} = \gamma_v \gamma_u \left(1 - \frac{u_x v}{c^2}\right). \tag{4.6}$$

4.2 Four-momentum

In his seminal 1905 paper *On the Electrodynamics of Moving Bodies*, Einstein derived the kinetic energy of an electron, $(\gamma_u - 1)mc^2$, and in a second paper published the same year, he showed that a body's mass changes when it emits energy, establishing the connection between mass and energy. In 1906, Planck developed the dynamics of relativistic particles using the momentum $\mathbf{p} = \gamma_u m\mathbf{u}$ and obtained the corresponding relativistic Lagrangian. In this section, we will take the reverse route, using the principle of least action to construct a relativistic Lagrangian and derive from it the relativistic momentum, $\mathbf{p} = \gamma_u m\mathbf{u}$, and energy, $E = \gamma_u mc^2$.

4.2.1 Lagrangian in special relativity

As discussed in section 1.2.4, a system moves from point A to point B by following the trajectory that makes the action stationary. The action is defined as the time integral of the Lagrangian from A to B. By requiring that the action is stationary, we derive the Euler-Lagrange equations, which are equivalent to the equations of motion. From these equations, we can identify the momentum. We now extend this concept to the relativistic case.

Lagrangian for a single particle

The determination of the appropriate Lagrangian for a physical system is guided by symmetry principles, among other factors. In classical mechanics, the action should be invariant under Galilean transformations, while in special relativity, it should be invariant under Lorentz transformations, ensuring the consistency of physical laws across all inertial frames. If the equations of motion are already known, we can work backwards to identify a corresponding action that reproduces these equations. Alternatively, proposing an action based on symmetry principles, physical intuition, or other considerations can lead to new equations of motion that may predict previously unknown phenomena.

In special relativity, we already know that the equation of motion is $\mathrm{d}\mathbf{p}/\mathrm{d}t = \mathbf{f}$, with $\mathbf{p} = \gamma_u m\mathbf{u}$, which serves as a benchmark for constructing a relativistic Lagrangian. Once such a Lagrangian is established, it can be used to further develop the theory of special relativity, particularly in the context of electromagnetism.

To extend the concept of action to the relativistic case, we must ensure that it is invariant under changes of reference frame. This means that if the action is stationary in one frame, it must also be stationary in all frames. For a single particle, this condition is satisfied if $L\,\mathrm{d}t$ is proportional to the proper time interval $\mathrm{d}\tau = \sqrt{-\mathrm{d}s^2/c}$ along the worldline, which is invariant. Note that the Lagrangian L itself is then not invariant, since $\mathrm{d}t$ is not. Thus, we express the action as:

$$S \propto \int_{\tau_A}^{\tau_B} \mathrm{d}\tau.$$

For a particle moving with velocity u, and using equation (4.2), this becomes $S \propto \int_{t_A}^{t_B} \sqrt{1-u^2/c^2}\,\mathrm{d}t$. In the limit $u/c \ll 1$, this integral reduces to $\int_{t_A}^{t_B} \left[1 - u^2/(2c^2)\right] \mathrm{d}t$, to the first non-zero order in u/c. In classical mechanics, in the absence of external forces, the Lagrangian is given by the kinetic energy $mu^2/2$. This expression should match the relativistic Lagrangian in the limit $u/c \ll 1$, up to an additive constant (since adding a constant to the Lagrangian does not alter the equations of motion). Therefore, the proportionality coefficient in the relativistic Lagrangian is $-mc^2$.

If the particle interacts with an external force, an additional term L_{int} must be included in the Lagrangian. We assume here that it depends on $\mathbf{r}$ and t, but not on $\dot{\mathbf{r}}$, which is sufficient to describe the motion in a given frame, such as the frame in which an external potential is specified. In general, however, such a term does not lead to a Lorentz-invariant action. A Lorentz-invariant interaction must depend on the velocity of the particle, as for the electromagnetic interaction discussed in section 5.7.1.

The relativistic Lagrangian for one particle is then:

$$\boxed{L = -mc^2\sqrt{1 - \frac{\dot{x}^2+\dot{y}^2+\dot{z}^2}{c^2}} + L_{\text{int}}\left(\mathbf{r}, t\right).} \tag{4.7}$$

Lagrangian for a system of particles

Consider a system of N particles. If they do not interact with each other, the Lagrangian for the system is simply the sum of the individual Lagrangians. When mutual interactions are present, they are incorporated into L_{int} such that the Lagrangian becomes:

$$L = \sum_{k=1}^{N} -m_k c^2\sqrt{1 - \frac{\dot{x}_k^2+\dot{y}_k^2+\dot{z}_k^2}{c^2}} + L_{\text{int}}\left(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N, t\right). \tag{4.8}$$

Here, the subscript k labels the k-th particle, and does not indicate a covariant coordinate. Therefore, $\mathbf{r}_k$ represents the position vector of the k-th particle. Note that an interaction term depending on the positions of all the particles at the same time t, such as the Coulomb interaction between charges, describes an instantaneous interaction at a distance, which is not compatible with special relativity. Such a description can only be valid in a given frame, as an approximation. A fully relativistic description of interactions requires fields, which propagate at finite speed, as we will see in chapter 5.

4.2.2 Relativistic momentum and energy

Momentum

In chapter 1, we noted that the Euler-Lagrange equations are applicable regardless of the explicit form of the Lagrangian, and therefore remain valid for the relativistic Lagrangian. These equations (1.20) state that, for each particle k:

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial L}{\partial \dot{x}_k}\right) = \frac{\partial L}{\partial x_k}, \quad \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial L}{\partial \dot{y}_k}\right) = \frac{\partial L}{\partial y_k}, \quad \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial L}{\partial \dot{z}_k}\right) = \frac{\partial L}{\partial z_k}. \tag{4.9}$$

In the absence of internal or external forces, the right-hand side of these equations is zero, implying that the quantities $\partial L/\partial \dot{x}_k$, $\partial L/\partial \dot{y}_k$ and $\partial L/\partial \dot{z}_k$ are conserved. These conserved quantities are the components of momentum. Therefore, for each particle, and now dropping the subscript k, we express the spatial momentum vector as $\mathbf{p} = (p_x, p_y, p_z)$ with $p_x = \partial L/\partial \dot{x} = \gamma_u m u_x$, and similarly for the other components, where we have used L given by equation (4.7). This gives:

$$\boxed{\mathbf{p} = \gamma_u m \mathbf{u}.} \tag{4.10}$$

In the limit $u/c \ll 1$, this simplifies to the familiar $\mathbf{p} = m\mathbf{u}$.

In section 1.3.2, we also noted that Noether's theorem remains valid for the relativistic Lagrangian. This implies that, when the system is invariant under spatial translation, which is the case when there are **no external forces** acting on the system, the **total momentum is conserved**.

Energy

Furthermore, when the Lagrangian does not explicitly depend on time, there exists a conserved quantity, the energy E, given by equation (1.68). For a single particle, this equation is:

$$E = \frac{\partial L}{\partial \dot{x}}\dot{x} + \frac{\partial L}{\partial \dot{y}}\dot{y} + \frac{\partial L}{\partial \dot{z}}\dot{z} - L = \text{const.} \tag{4.11}$$

Assuming $L_{\text{int}} = 0$ in the expression (4.7) for the Lagrangian, this is equal to $\mathbf{p} \cdot \mathbf{u} + mc^2/\gamma_u = \gamma_u m \left(u^2 + c^2/\gamma_u^2\right)$, which simplifies to:

$$\boxed{E = \gamma_u m c^2.} \tag{4.12}$$

This represents the total energy of a single particle in the absence of external forces and is commonly referred to as the **relativistic energy**. In the limit $u/c \ll 1$, this reduces to:

$$E \simeq \frac{1}{2}mu^2 + mc^2.$$

The constant term mc^2 is the rest energy of the particle. It represents the energy that the particle possesses due to its mass alone, independent of its motion. This term is significant in relativity, illustrating the concept that mass and energy are equivalent. It does not appear in classical mechanics, where energy is purely a function of motion and position, and does not account for the intrinsic energy of mass.

Since mc^2 is the energy of the particle in its rest frame, which all observers can identify, it has the same value for all observers: it is Lorentz invariant. This will also follow directly from the invariance of the squared length of the 4-momentum, as shown by equation (4.18). Therefore, the mass m is an intrinsic property of the particle.

By subtracting the rest energy from the total energy E, we obtain the kinetic energy (in the absence of external forces and therefore of other forms of energy, such as potential energy):

$$K = (\gamma_u - 1)\, mc^2. \tag{4.13}$$

For a system of N particles, equation (4.11) still applies, provided the first three terms are summed over all the particles. If there are neither external nor internal forces ($L_{\text{int}} = 0$), the conserved quantity is the total energy:

$$E_{\text{total}} = \sum_{k=1}^{N} \gamma_{u_k} m_k c^2. \tag{4.14}$$

Energy flux and momentum density

Consider a volume V containing N particles labelled by $k = 1, \ldots, N$, with the k-th particle having mass m_k, velocity $\mathbf{u}_k$ and momentum $\mathbf{p}_k$.

The energy flux carried by this collection of particles is given by:

$$\mathbf{S} \equiv \frac{1}{V} \sum_{k=1}^{N} E_k \mathbf{u}_k = \frac{1}{V} \sum_{k=1}^{N} \gamma_{u_k} m_k c^2 \mathbf{u}_k = \frac{c^2}{V} \sum_{k=1}^{N} \mathbf{p}_k. \tag{4.15}$$

Defining the momentum density as $\mathbf{g} \equiv \frac{1}{V} \sum_{k=1}^{N} \mathbf{p}_k$, this becomes:

$$\boxed{\mathbf{S} = \mathbf{g} c^2.} \tag{4.16}$$

This relation also applies to the electromagnetic field, as will be shown in section 5.8.4.

4.2.3 Four-momentum and energy-momentum relation

Multiplying the 4-velocity by m yields a 4-vector that includes the momentum: $mU^\mu = (\gamma_u mc, \mathbf{p})$. The first component is E/c. This 4-vector is referred to as the 4-momentum:

$$\boxed{P^\mu = mU^\mu = \left(\frac{E}{c}, \mathbf{p}\right).} \tag{4.17}$$

Note that E here includes only the rest and kinetic energies. The 4-momentum is defined this way even when additional energy, such as potential energy from external forces, is present, meaning that potential energy does not directly contribute to the 4-momentum.

Since P^μ is a 4-vector, its squared length $-E^2/c^2 + p^2$, where $p^2 = \mathbf{p} \cdot \mathbf{p}$, is invariant. In the rest frame of the particle, $p = 0$ and $E = mc^2$, so we obtain:

$$\boxed{(P^\mu)^2 = -m^2c^2 \text{ is invariant.}} \tag{4.18}$$

This yields the **energy-momentum relation for a particle**:

$$\boxed{E^2 = m^2c^4 + p^2c^2.} \tag{4.19}$$

This expression holds for any particle with mass m. It can be easily checked that it is consistent with $p = \gamma_u mu$ and $E = \gamma_u mc^2$.

The 4-momentum, being a 4-vector, transforms according to the Lorentz transformation. We denote the 4-momentum of a particle in frame (S) as P^μ and in frame (S′), which moves with velocity $\mathbf{v}$ relative to (S), as P'^μ. By aligning the x-axis with $\mathbf{v}$, the transformation $P'^\mu = \Lambda P^\mu$ gives:

$$E' = \gamma_v \left(E - vp_x\right), \quad p'_x = \gamma_v \left(-\frac{v}{c^2}E + p_x\right), \quad p'_y = p_y, \quad p'_z = p_z. \tag{4.20}$$

4.2.4 Four-momentum of massless particles

Massless particles travel with speed $u = c$, leading to a situation where the momentum $\gamma_u m\mathbf{u}$ involves multiplying an infinite quantity (γ_u) by zero (m). In this case, we define the momentum in terms of energy rather than velocity. The energy-momentum relation (4.19) shows that $E^2 \to p^2c^2$ as $m \to 0$. We assume that this expression remains valid for $m = 0$. For photons[1], this will be confirmed independently in section 5.8.4, where Maxwell's equations show that the momentum density of the electromagnetic field is its energy flux

[1] Gluons, which mediate the strong interaction, are also massless, but they are not observed as free particles.

divided by c^2. For light, which carries energy at speed c, this gives a momentum equal to the energy divided by c. This leads to:

$$\boxed{p = \frac{E}{c}.} \tag{4.21}$$

Thus, the 4-momentum of a massless particle is given by:

$$P^\mu = \left(\frac{E}{c}, \frac{E}{c}\hat{\mathbf{n}}\right), \tag{4.22}$$

where $\hat{\mathbf{n}}$ is the unit vector in the direction of propagation of the particle. This implies:

$$\boxed{(P^\mu)^2 = 0.} \tag{4.23}$$

The energy of a photon is given by $E = h\nu$, where ν is the frequency. For a photon moving along the x-axis, the 4-momentum in frame (S) is $P^\mu = (h\nu/c, h\nu/c, 0, 0)$. In a frame (S′) moving with velocity v along the x-axis relative to (S), that is to say in the direction of propagation of the photon for $v > 0$, and using equation (4.20) along with $\beta = v/c$, we find that $h\nu'/c = \gamma_v (1 - \beta) h\nu/c$, which simplifies to:

$$\boxed{\nu' = \sqrt{\frac{1-\beta}{1+\beta}}\,\nu.} \tag{4.24}$$

This equation represents the relativistic Doppler shift: for $v > 0$, the observer in (S′) moves in the same direction as the photon, so that $\nu' < \nu$ (redshift). It can also be derived by combining the classical Doppler effect, due to the change in distance travelled by successive wavefronts, with time dilation. More generally, for a photon propagating at an angle θ to the x-axis in (S), equation (4.20) gives $\nu' = \gamma_v (1 - \beta \cos\theta)\, \nu$.

4.2.5 Wave four-vector

Photons can be described as electromagnetic waves and, similarly, massive particles also exhibit wave-like behaviour. For both massless and massive particles, the energy E of the particle is related to the frequency ν of the wave through Planck's relation $E = h\nu$, and the momentum $\mathbf{p}$ is related to the wavevector $\mathbf{k}$ by de Broglie's relation $\mathbf{p} = \hbar\mathbf{k}$.

The wave properties of a particle can be conveniently described using the concept of the wave 4-vector, which expresses the frequency and wavevector in a relativistically covariant form. This 4-vector can be derived from the 4-momentum P^μ by dividing it by $\hbar$:

$$K^\mu = \frac{P^\mu}{\hbar} = \left(\frac{E}{\hbar c}, \frac{\mathbf{p}}{\hbar}\right). \tag{4.25}$$

Since $\hbar$ is a frame-invariant constant, K^μ, defined in this way, is indeed a 4-vector and therefore transforms according to Lorentz transformations. Using the relations above, we can express the wave 4-vector as:

$$\boxed{K^\mu = \left(\frac{\omega}{c}, \mathbf{k}\right),} \tag{4.26}$$

where $\omega = 2\pi\nu$ is the angular frequency, and the wavevector is related to the wavelength by $k = 2\pi/\lambda$.

Since K^μ is a 4-vector, its squared length $(K^\mu)^2 = -\omega^2/c^2 + k^2$, where $k^2 = \mathbf{k} \cdot \mathbf{k}$, is invariant. For massless particles, $(P^\mu)^2 = 0$ and therefore $(K^\mu)^2 = 0$. This leads to the condition $\omega/k = c$ (equivalent to $p = E/c$), which shows that massless particles are described by waves propagating at the speed of light.

For massive particles, in the rest frame where $\mathbf{p} = \mathbf{0}$ and $E = mc^2$, the invariant calculated from equation (4.25) is $-(mc/\hbar)^2$. This gives the following *dispersion relation*:

$$\omega^2 = c^2k^2 + \left(\frac{mc^2}{\hbar}\right)^2. \tag{4.27}$$

The phase velocity ω/k is then larger than c, but this does not violate relativity, since no information is carried at the phase velocity. The group velocity:

$$\frac{\mathrm{d}\omega}{\mathrm{d}k} = \frac{c^2k}{\omega} = \frac{pc^2}{E} = u, \tag{4.28}$$

is the velocity of the particle.

Consider the spacetime position 4-vector $X^\mu = (ct, \mathbf{r})$. The scalar product $X^\mu \cdot K^\mu = \mathbf{k} \cdot \mathbf{r} - \omega t$ is invariant. This quantity represents the *phase* of the wave associated with the particle. Conversely, since the phase determines the number of wave crests passing a given event, all observers must agree on its value. This shows that $(\omega/c, \mathbf{k})$ is a 4-vector for any plane wave, such as an electromagnetic wave, even when it is not associated with a particle.

4.3 Four-force

The Euler-Lagrange equations (4.9) show that, when a system of N particles is subject to external and/or internal forces, the momentum of the k-th particle changes according to:

$$\frac{\mathrm{d}\mathbf{p}_k}{\mathrm{d}t} = \mathbf{f}_k, \tag{4.29}$$

where $\mathbf{f}_k = \boldsymbol{\nabla}_{\mathbf{r}_k} L_{\text{int}}$ is the force acting on the k-th particle. The subscript on the gradient operator indicates the coordinates with respect to which the derivatives are calculated.

In the same way that we cannot create a velocity 4-vector by dividing $\mathrm{d}X^\mu$ by $\mathrm{d}t$, even though $\mathbf{u} = \mathrm{d}\mathbf{r}/\mathrm{d}t$, we also cannot create a force 4-vector by dividing $\mathrm{d}P^\mu$ by $\mathrm{d}t$, even though the force is given by $\mathrm{d}\mathbf{p}/\mathrm{d}t$. Instead, we define the force 4-vector, or 4-force, F^μ, as:

$$\boxed{F^\mu = \frac{\mathrm{d}P^\mu}{\mathrm{d}\tau}.} \tag{4.30}$$

Using $\mathrm{d}t/\mathrm{d}\tau = \gamma_u$, this yields $F^\mu = \gamma_u \left(\mathrm{d}P^\mu/\mathrm{d}t\right)$, which can be written as:

$$\boxed{F^\mu = \left(\gamma_u \frac{1}{c}\frac{\mathrm{d}E}{\mathrm{d}t}, \gamma_u \mathbf{f}\right),} \tag{4.31}$$

where $\mathbf{f}$ is the force acting on the particle (we now omit the subscript k). These two equations establish the relativistic form of the equation of motion, valid in any inertial frame.

The scalar product of the two 4-vectors U^μ and F^μ is invariant and equal to its value in the particle's rest frame. Therefore:

$$U^\mu \cdot F^\mu = \gamma_u^2 \left(-\frac{\mathrm{d}E}{\mathrm{d}t} + \mathbf{f}\cdot\mathbf{u}\right) = -c^2 \frac{\mathrm{d}m}{\mathrm{d}\tau}, \tag{4.32}$$

where we have used $E = mc^2$ and $t = \tau$ in the rest frame. This can also be obtained without referring to a specific frame: since $P^\mu = mU^\mu$ and $\left(U^\mu\right)^2 = -c^2$ is constant, we have:

$$U^\mu \cdot F^\mu = U^\mu \cdot \frac{\mathrm{d}\left(mU^\mu\right)}{\mathrm{d}\tau} = \frac{\mathrm{d}m}{\mathrm{d}\tau}\left(U^\mu\right)^2 + \frac{m}{2}\frac{\mathrm{d}\left(U^\mu\right)^2}{\mathrm{d}\tau} = -c^2\frac{\mathrm{d}m}{\mathrm{d}\tau}.$$

According to equation (4.32), the mass of a particle could vary along its worldline if the particle interacted with a field whose value contributes to its mass. This is the role of the Higgs field: elementary particles, such as the electron, owe their mass to their interaction with it. However, the Higgs field has a uniform, constant and non-zero vacuum expectation value, so that the masses it generates are constant. For the interactions considered in this course, the mass of a particle therefore remains constant, so that $\mathrm{d}m/\mathrm{d}\tau = 0$. Equation (4.32) then implies:

$$\frac{\mathrm{d}E}{\mathrm{d}t} = \mathbf{f}\cdot\mathbf{u}, \tag{4.33}$$

where E represents the relativistic energy, which includes both kinetic energy and rest mass energy. Since the mass is constant, $\mathrm{d}E/\mathrm{d}t$ is the rate of change of kinetic energy. Just as in classical mechanics, this is equal to the work done by the force per unit time.

Since F^μ is a 4-vector, it transforms according to Lorentz transformations, allowing us to calculate the force in different inertial frames.

4.4 Four-acceleration and instantaneous rest frame

As discussed earlier, the relationship between the 4-force and the 4-momentum is the relativistic equation of motion. Similar to non-relativistic mechanics, the 4-force can be expressed in terms of a 4-acceleration, though this 4-vector is not related to the spatial acceleration vector in a straightforward way.

4.4.1 Four-acceleration

Given that $P^\mu = mU^\mu$ and that the mass is constant, as discussed in section 4.3, equation (4.30) can be rewritten as:

$$F^\mu = m\frac{\mathrm{d}U^\mu}{\mathrm{d}\tau},$$

which leads us to define the 4-acceleration:

$$\boxed{A^\mu = \frac{\mathrm{d}U^\mu}{\mathrm{d}\tau}.} \tag{4.34}$$

Using $\mathrm{d}t/\mathrm{d}\tau = \gamma_u$, this gives:

$$A^\mu = \gamma_u \frac{\mathrm{d}U^\mu}{\mathrm{d}t} = \gamma_u \left(\frac{\mathrm{d}\gamma_u}{\mathrm{d}t} c, \frac{\mathrm{d}\gamma_u}{\mathrm{d}t}\mathbf{u} + \gamma_u \frac{\mathrm{d}\mathbf{u}}{\mathrm{d}t} \right). \tag{4.35}$$

We have:

$$\frac{\mathrm{d}\gamma_u}{\mathrm{d}t} = \gamma_u^3 \frac{1}{2c^2} \frac{\mathrm{d}\left(\mathbf{u}\cdot\mathbf{u}\right)}{\mathrm{d}t} = \gamma_u^3 \frac{\mathbf{u}\cdot\mathbf{a}}{c^2}, \tag{4.36}$$

where $\mathbf{a} = \mathrm{d}\mathbf{u}/\mathrm{d}t$ is the standard spatial acceleration vector.
Therefore, the 4-acceleration can be expressed as:

$$\boxed{A^\mu = \gamma_u^2 \left(\gamma_u^2 \frac{\mathbf{u}\cdot\mathbf{a}}{c}, \gamma_u^2 \frac{\mathbf{u}\cdot\mathbf{a}}{c^2}\mathbf{u} + \mathbf{a} \right).} \tag{4.37}$$

Because $\mathbf{f} = \mathrm{d}\mathbf{p}/\mathrm{d}t$ and $\mathbf{p} = \gamma_u m\mathbf{u}$, there is no simple relation between the force $\mathbf{f}$ and the acceleration $\mathbf{a}$.

4.4.2 Instantaneous rest frame

When a particle accelerates (with acceleration that may vary over time), its rest frame can be defined as the *inertial frame* in which the particle is momentarily at rest at any given instant. This frame, called the *instantaneous rest frame* (IRF), changes continuously as the particle's velocity changes. In this frame, the particle experiences the *proper acceleration*, namely the physical acceleration felt by the particle and measurable by an accelerometer comoving with it.

Consider an inertial frame (S) in which the particle has velocity $\mathbf{u}(t)$, with t the time measured in (S). At time $t = t_0$, the IRF of the particle, denoted $\mathrm{S}(t_0)$, is the inertial frame moving with the *constant* velocity $\mathbf{u}(t_0)$. At a later time t_1, the IRF $\mathrm{S}(t_1)$ is the inertial frame moving with the (generally different) *constant* velocity $\mathbf{u}(t_1)$. In $\mathrm{S}(t_0)$, the particle is momentarily at rest, and although it immediately acquires a small velocity in that frame, the coordinate time elapsed in $\mathrm{S}(t_0)$ during an infinitesimal interval is equal, to first order, to the proper-time increment $d\tau$. The same holds for $\mathrm{S}(t_1)$, and for any IRF: because an IRF is the inertial frame tangent to the particle's worldline at a given event, the time interval measured in that frame during an infinitesimal step is equal, to first order, to the proper-time increment $\mathrm{d}\tau$, i.e., the time recorded by a clock attached to the particle.

In the IRF, the instantaneous velocity of the particle is zero, so $\gamma_u = 1$, and $\mathrm{d}E/\mathrm{d}t = \mathrm{d}\left(mc^2\right)/\mathrm{d}\tau$. Therefore, the 4-force (4.31) in this frame is:

$$F^\mu = \left(c\frac{\mathrm{d}m}{\mathrm{d}\tau}, \mathbf{f}_0\right), \tag{4.38}$$

where $\mathbf{f}_0$ represents the ordinary 3-force in the IRF. Since the particle's velocity is zero in its IRF, equation (4.37) implies that the 4-acceleration there is:

$$A^\mu = (0, \mathbf{a}_0), \tag{4.39}$$

where $\mathbf{a}_0$ is, by definition, the proper acceleration. This yields:

$$(A^\mu)^2 = a_0^2.$$

Using equations (4.38) and (4.39), the time component of $F^\mu = mA^\mu$ gives $\mathrm{d}m/\mathrm{d}\tau = 0$, consistent with the mass being constant, as discussed in section 4.3. The spatial components give:

$$m\mathbf{a}_0 = \mathbf{f}_0. \tag{4.40}$$

This relation between force and acceleration holds in general *only in the IRF*. In any other inertial frame, the relation $F^\mu = mA^\mu$ is consistent with $\mathbf{f} = \mathrm{d}\mathbf{p}/\mathrm{d}t$, but it does not in general reduce to $\mathbf{f} = m\mathbf{a}$.

This can be seen more explicitly by transforming the force and acceleration from the IRF to an inertial frame (S). Assuming the particle moves in (S) with velocity $\mathbf{u}$ along the x-axis, transformation of the 4-force gives:

$$\begin{pmatrix} \gamma_u \frac{1}{c}\frac{\mathrm{d}E}{\mathrm{d}t} \\ \gamma_u f_x \\ \gamma_u f_y \\ \gamma_u f_z \end{pmatrix} = \begin{pmatrix} \gamma_u & \frac{u}{c}\gamma_u & 0 & 0 \\ \frac{u}{c}\gamma_u & \gamma_u & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 0 \\ f_{0x} \\ f_{0y} \\ f_{0z} \end{pmatrix}, \tag{4.41}$$

which yields $\mathrm{d}E/\mathrm{d}t = uf_{0x}$, $f_x = f_{0x}$, $f_y = f_{0y}/\gamma_u$, and $f_z = f_{0z}/\gamma_u$.

Similarly, transforming the 4-acceleration yields:

$$\begin{pmatrix} \gamma_u^4 \dfrac{u a_x}{c} \\ \gamma_u^2 \left(\gamma_u^2 \dfrac{u a_x}{c^2} u + a_x \right) \\ \gamma_u^2 a_y \\ \gamma_u^2 a_z \end{pmatrix} = \begin{pmatrix} \gamma_u & \dfrac{u}{c}\gamma_u & 0 & 0 \\ \dfrac{u}{c}\gamma_u & \gamma_u & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 0 \\ a_{0x} \\ a_{0y} \\ a_{0z} \end{pmatrix}, \tag{4.42}$$

which gives $a_x = a_{0x}/\gamma_u^3$, $a_y = a_{0y}/\gamma_u^2$ and $a_z = a_{0z}/\gamma_u^2$. Therefore, $m\mathbf{a}_0 = \mathbf{f}_0$ implies that $\gamma_u^3 m a_x = f_x$, $\gamma_u m a_y = f_y$ and $\gamma_u m a_z = f_z$.

Finally, we note that, in the IRF, U^μ has no spatial component while A^μ only has a spatial component. Consequently, their scalar product is zero. Since this result is invariant, it implies that the 4-velocity and 4-acceleration are always orthogonal, leading to:

$$\boxed{U^\mu \cdot A^\mu = 0.} \tag{4.43}$$

4.4.3 Proper acceleration and rapidity

Consider a particle initially at rest in an inertial frame (S). We label its velocity using the proper time τ, recorded by the clock attached to the particle. At time τ, the particle is moving with velocity $u(\tau)$ in frame (S), and its IRF, denoted as S(τ), moves relative to (S) with the same velocity.

At time $\tau + \mathrm{d}\tau$, the particle acquires in S(τ) a small velocity increment $\mathrm{d}u_0 = a_0\,\mathrm{d}\tau$, where a_0 is the proper acceleration. In frame (S), its velocity at this later time is $u(\tau + \mathrm{d}\tau)$. Using the velocity transformation given by equation (3.20), we obtain:

$$u\left(\tau + \mathrm{d}\tau\right) = \frac{a_0 \mathrm{d}\tau + u(\tau)}{1 + a_0 \mathrm{d}\tau\; u(\tau)/c^2}.$$

Here, we assume that $\mathbf{a}_0$ and $\mathbf{u}(\tau)$ are aligned, so that $\mathbf{u}\left(\tau + \mathrm{d}\tau\right)$ remains parallel to the boost velocity. When $\mathbf{a}_0$ is not parallel to $\mathbf{u}(\tau)$, successive boosts are in different directions and, as discussed in section 2.4.3 of chapter 2, they combine into a boost and a rotation, so that the simple result derived below no longer applies.

Defining $\mathrm{d}u = u\left(\tau + \mathrm{d}\tau\right) - u(\tau)$, and neglecting terms of second order in $\mathrm{d}\tau$, we obtain:

$$\frac{\mathrm{d}u}{1 - u^2/c^2} = a_0 \mathrm{d}\tau.$$

The left-hand side simplifies to $c\;\mathrm{d}\left[\tanh^{-1}\left(u/c\right)\right]$, where $\tanh^{-1}\left(u/c\right)$ is the rapidity ζ, as defined in chapter 2 (eq. [2.46]).

Integrating the equation yields:

$$\boxed{\zeta(\tau) = \frac{1}{c}\int_0^\tau a_0\left(\tau'\right)\mathrm{d}\tau',} \tag{4.44}$$

where we have used $\zeta = 0$ at $\tau = 0$. In the non-relativistic limit $u/c \ll 1$, $\zeta \simeq u/c$ and the equation simplifies to $u(\tau) = \int_0^\tau a_0\left(\tau'\right)\mathrm{d}\tau'$, as expected.

To see the utility of this relation, consider a particle that receives two brief, equal proper-acceleration pulses separated by a proper-time interval $\Delta\tau_{\text{rest}}$ during which the particle coasts:

- Pulse 1: duration $\Delta\tau_1$, constant proper acceleration a_0
- Rest interval: $\Delta\tau_{\text{rest}}$ (no acceleration)
- Pulse 2: duration $\Delta\tau_2 = \Delta\tau_1$, same a_0

Each pulse imparts a rapidity increment

$$\Delta\zeta = \frac{a_0\Delta\tau_1}{c}.$$

From equation (4.44), the total rapidity after both pulses is simply:

$$\zeta_{\text{final}} = 2\Delta\zeta = \frac{2a_0\Delta\tau_1}{c},$$

independent of the rest interval. The corresponding final velocity in frame (S) is:

$$u_f = c\tanh\zeta_{\text{final}} = c\tanh\left(\frac{2a_0\Delta\tau_1}{c}\right).$$

This additivity reflects the fact that rapidity is the natural parameter of Lorentz boosts: successive boosts combine by *adding rapidities*, whereas velocities combine nonlinearly. Indeed, if $u_1 = c\tanh\Delta\zeta$ is the velocity after the first pulse then, using $\tanh 2x = 2\tanh x/\left(1+\tanh^2 x\right)$, we obtain $u_f = 2u_1/\left(1+u_1^2/c^2\right) < 2u_1$, in agreement with the velocity composition formula (3.19).

4.5 Particle collisions

In particle physics, collisions typically refer to interactions between particles mediated by fundamental forces rather than physical contact. This includes scattering events, annihilation and particle creation, where the forces at play determine the outcome of the interaction. We now consider such processes.

4.5.1 Total four-momentum

As discussed above, in the absence of external forces, the total momentum of the system is conserved. Additionally, when particles are sufficiently far apart, either long before or after the interactions occur, there is no energy associated with interaction forces, and the total energy of the system includes only rest and kinetic energies. Therefore, the sum P^μ of the momentum 4-vectors for each of the N particles, given by:

$$P^\mu = \sum_{k=1}^{N} P_k^\mu = \left(\frac{E_{\text{total}}}{c}, \mathbf{p}_{\text{total}} \right), \tag{4.45}$$

is **conserved** if evaluated long before and long after the collision. Here the subscript k refers to the k-th particle.

This sum is also a **4-vector**. Indeed, each P_k^μ is a 4-vector, transforming to a frame (S′) as $P_k'^\mu = \Lambda P_k^\mu$, where the coefficients of Λ depend only on the velocity between frames (S) and (S′). If all P_k^μ are evaluated simultaneously in frame (S), the transformation might result in the $P_k'^\mu$ not being simultaneous in frame (S′). However, if the 4-vectors are evaluated when the particles are far apart, they are time-independent, as each particle conserves its energy and momentum. Thus, $P'^\mu = \Lambda P^\mu$, indicating that P^μ is itself a 4-vector, and therefore its squared length is invariant. Calculating it in the centre of momentum frame, where $\mathbf{p}_{\text{total}} = \mathbf{0}$, we obtain $(P^\mu)^2 = -E_{\text{total}}^2/c^2 = -\left(\sum_k \gamma_{u_k} m_k\right)^2 c^2$, where u_k is the velocity of the k-th particle in that frame. By analogy with a single particle, we write $(P^\mu)^2 = -M^2c^2$, where $M = \sum_k \gamma_{u_k} m_k$ is called the *invariant mass* of the system. Note that M is generally larger than the sum of the masses of the particles, $\sum_k m_k$, since it includes their kinetic energies in the centre of momentum frame. Since P^μ is conserved, the invariant mass is conserved in any collision.

The conservation of total energy and total momentum each provide distinct constraints on the system, but these constraints only relate quantities calculated within the same frame. We often gain additional useful information by noting that the squared length of the total 4-momentum remains invariant, as this allows us to derive relationships between quantities calculated in different frames.

We denote the total energy and momentum of the system (comprising all particles) by E and $\mathbf{p}$, respectively, and the sum of the individual particle masses by

$$m = \sum_k m_k.$$

We use the subscripts 'before' and 'after' to denote the values of these quantities before and after the collision. We assume throughout that the system is isolated.

We then have the following results, where K denotes the total kinetic energy of the particles::

- Classical collisions:
  - Momentum: $\mathbf{p}_{\text{after}} = \mathbf{p}_{\text{before}}$,
  - Mass: $m_{\text{after}} = m_{\text{before}}$,
  - Energy: the total energy is conserved, but the kinetic energy may be converted to or from internal energy (heat, deformation, chemical energy), which does not affect the mass:
    - Elastic collision: $K_{\text{after}} = K_{\text{before}}$,
    - Explosive collision: $K_{\text{after}} > K_{\text{before}}$,
    - Sticky collision: $K_{\text{after}} < K_{\text{before}}$;
- Relativistic collisions:
  - Momentum: $\mathbf{p}_{\text{after}} = \mathbf{p}_{\text{before}}$,
  - Energy (kinetic plus rest): $E_{\text{after}} = E_{\text{before}}$,
  - Mass: since the rest energy now includes the internal energy, any change in kinetic energy is balanced by a change in the sum of the masses:
    - Elastic collision: $K_{\text{after}} = K_{\text{before}}$ and $m_{\text{after}} = m_{\text{before}}$,
    - Explosive collision: $K_{\text{after}} > K_{\text{before}}$ and $m_{\text{after}} < m_{\text{before}}$,
    - Sticky collision: $K_{\text{after}} < K_{\text{before}}$ and $m_{\text{after}} > m_{\text{before}}$.

4.5.2 Centre of momentum frame

Let $P^\mu = (E_{\text{total}}/c, \mathbf{p}_{\text{total}})$ represent the 4-momentum of the entire system of particles in frame (S), and let $P^{*\mu} = (E^*/c, \mathbf{0})$ represent the 4-momentum in the centre of momentum (CM) frame. Here, E^* is the energy of the system in the CM frame. The CM frame moves with velocity v_{cm} relative to frame (S), and we choose the x-axis to be in the direction of that velocity. We denote γ_{cm} as the Lorentz factor associated with the velocity v_{cm} and Λ_{cm} as the transformation matrix from frame (S) to the CM frame.

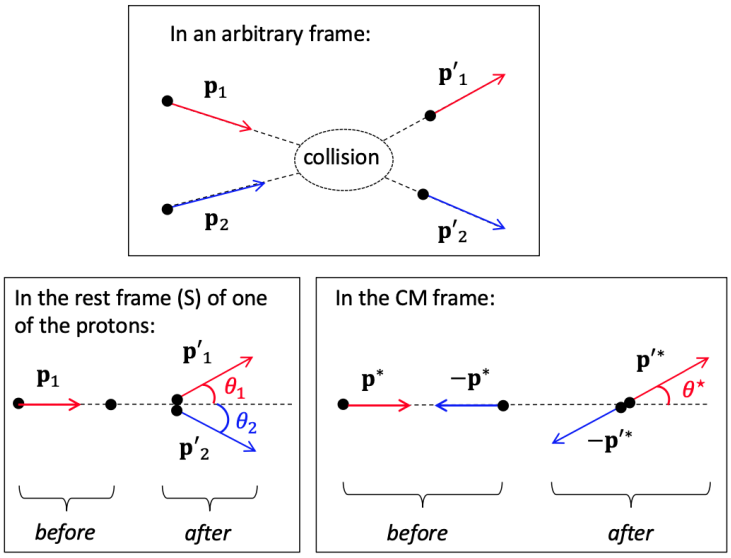
Using $\Lambda_{\text{cm}} P^\mu = P^{*\mu}$, we have $\gamma_{\text{cm}} \left(-v_{\text{cm}} E_{\text{total}}/c^2 + p_{\text{total},x}\right) = 0$ and $p_{\text{total},y} = p_{\text{total},z} = 0$. This implies that $\mathbf{p}_{\text{total}}$ is parallel to $\mathbf{v}_{\text{cm}}$, and:

$$\boxed{\mathbf{v}_{\text{cm}} = \frac{\mathbf{p}_{\text{total}} c^2}{E_{\text{total}}}.} \tag{4.46}$$

We now consider various cases of relativistic collisions.

4.5.3 Elastic collision: proton-proton scattering

Consider two protons moving along inclined paths, approaching a point of collision. We assume the protons remain intact during the collision. The diagrams below illustrate the situation before and after the collision in an arbitrary frame, in the rest frame (S) of one of the protons, and in the CM frame of the two protons:



In the diagram, we have assumed that the proton moving forwards in the CM frame also moves upwards. However, if it were to move downwards instead, the results presented here would remain unchanged. Hereafter, we denote quantities in the CM frame with a superscript '$*$' and quantities after collision with a prime. Velocities are denoted u, except for the velocity of the centre of momentum frame, which is $v_{\rm cm}$.

In classical mechanics, the conservation of momentum in frame (S) implies $(\mathbf{p}'_1 + \mathbf{p}'_2)^2 = p'^2_1 + p'^2_2 + 2\mathbf{p}'_1 \cdot \mathbf{p}'_2 = p^2_1$, and the conservation of energy implies $u'^2_1 + u'^2_2 = u^2_1$. This leads to $\mathbf{p}'_1 \cdot \mathbf{p}'_2 = 0$, indicating that the protons scatter at right angles to each other.

We will now show that this is no longer the case in relativity. Denoting the mass of the proton as m_p, we have the following relation in the CM frame:

$$p^*_1 = p^*_2 \Rightarrow \frac{m_p u^*_1}{\sqrt{1 - u^{*2}_1/c^2}} = \frac{m_p u^*_2}{\sqrt{1 - u^{*2}_2/c^2}} \Rightarrow u^*_1 = u^*_2.$$

Since $u_2 = 0$, the velocity transformation from frame (S) to the CM frame yields $\mathbf{u}^*_2 = -\mathbf{v}_{\rm cm}$. Therefore, $u^*_1 = u^*_2 = v_{\rm cm}$.

Similarly, the fact that the protons have equal and opposite momenta after the collision implies $u'^*_1 = u'^*_2$. Conservation of energy in the CM frame then implies that the γ factors of the protons must be the same before and after the collision. Consequently, the velocities after the collision are the same as before, meaning $u'^*_1 = u'^*_2 = v_{\rm cm}$, which implies $p'^* = p^* = \gamma_{\rm cm} m_p v_{\rm cm}$.

Next, we consider the transformation of the post-collision 4-momentum of the proton represented in red on the diagram. In the CM frame, it is given by:

$$P_1'^{*\mu} = \left(\frac{E_1^*}{c}, p^* \cos\theta^*, p^* \sin\theta^*, 0\right),$$

with $E_1^* = \gamma_{\text{cm}} m_p c^2$. In frame (S), it becomes:

$$P_1'^{\mu} = \left(\frac{E_1'}{c}, p_1' \cos\theta_1, p_1' \sin\theta_1, 0\right).$$

We have $P_1'^{\mu} = (\Lambda_{\text{cm}})^{-1} P_1'^{*\mu}$, where $(\Lambda_{\text{cm}})^{-1}$ is the transformation associated with the velocity $-v_{\text{cm}}$. This yields:

$$p_1' \cos\theta_1 = \gamma_{\text{cm}} \left(v_{\text{cm}} E_1^*/c^2 + p^* \cos\theta^*\right) \quad \text{and} \quad p_1' \sin\theta_1 = p^* \sin\theta^*.$$

Since $v_{\text{cm}} E_1^*/c^2 = \gamma_{\text{cm}} m_p v_{\text{cm}} = p^*$, dividing the second equation by the first one then gives:

$$\tan\theta_1 = \frac{\sin\theta^*}{\gamma_{\text{cm}}\left(1 + \cos\theta^*\right)}. \text{ Similarly, } \tan\theta_2 = \frac{\sin\theta^*}{\gamma_{\text{cm}}\left(1 - \cos\theta^*\right)}.$$

If $\theta^* = 0$ or π, the transformation shows that both θ_1 and θ_2 are also 0 or π. We have:

$$\tan\left(\theta_1 + \theta_2\right) = \frac{\tan\theta_1 + \tan\theta_2}{1 - \tan\theta_1 \tan\theta_2} = \frac{2\gamma_{\text{cm}}}{\left(\gamma_{\text{cm}}^2 - 1\right)\sin\theta^*}.$$

When $u_1/c \ll 1$ and therefore $\gamma_{\text{cm}} \to 1$, we have $\theta_1 + \theta_2 \to \pi/2$. Conversely, when $u_1/c \to 1$ and therefore $\gamma_{\text{cm}} \to +\infty$, we have $\theta_1 + \theta_2 \to 0$. For a fixed value of $\gamma_{\text{cm}} > 1$, $\theta_1 + \theta_2 < \pi/2$. The two protons therefore emerge closer to the forward direction than in classical mechanics, an effect that becomes more pronounced as the energy increases.

4.5.4 Emission and absorption

Spontaneous emission

Spontaneous emission is a process in which an excited atom (or molecule) loses energy by emitting a photon and transitioning to a lower energy state. In the rest frame of the atom before emission, the 4-momentum of the atom is $P^\mu = (Mc, \mathbf{0})$, where Mc^2 is the rest energy of the excited atom. After emission, the 4-momentum of the atom is $P'^\mu = (E'/c, \mathbf{p}')$, and that of the photon is $P_\gamma'^\mu = \left(E_\gamma'/c, \mathbf{p}_\gamma'\right)$. Let $M'c^2$ denote the rest energy of the atom after emission.

Conservation of momentum implies that the atom *recoils* (acquires some momentum) during the emission of the photon, so that $E' > M'c^2$ and the energy E_γ' of the photon is not merely the difference $E_0 \equiv (M - M')\,c^2$ between the rest energies of the excited and de-excited atom. To calculate E_γ', we apply conservation of 4-momentum:

$$P^\mu - P_\gamma'^\mu = P'^\mu \Rightarrow \left(P^\mu\right)^2 + \left(P_\gamma'^\mu\right)^2 - 2P^\mu \cdot P_\gamma'^\mu = \left(P'^\mu\right)^2.$$

By isolating P'^{μ}, we avoid involving the unknown momentum $\mathbf{p}'$ in the calculation. Using the relations $(P^{\mu})^2 = -M^2c^2$, $(P'^{\mu})^2 = -M'^2c^2$, $\left(P'^{\mu}_{\gamma}\right)^2 = 0$ and $P^{\mu} \cdot P'^{\mu}_{\gamma} = -ME'_{\gamma}$, we obtain $-M^2c^2 + 2ME'_{\gamma} = -M'^2c^2$, which simplifies to:

$$E'_{\gamma} = \frac{\left(M^2 - M'^2\right)c^2}{2M} = \left(1 - \frac{E_0}{2Mc^2}\right)E_0.$$

Radioactive decay

Radioactive decay is a natural process by which an unstable atomic nucleus loses energy by emitting radiation, which can be in the form of alpha particles, beta particles, or gamma rays:

- α-decay: The parent nucleus emits an α particle and decays into a lighter nucleus,
- β-decay: There are two types:

  (i) β^--decay: A neutron is converted into a proton, an electron, and an antineutrino. This can occur in a nucleus or with a free neutron.

  (ii) β^+-decay: A proton is converted into a neutron, a positron, and a neutrino. This occurs within a nucleus since a free proton cannot decay in this manner due to its lower mass compared to a neutron.

  In both types of β-decay within a nucleus, the electron or positron and the (anti)neutrino are emitted, resulting in the transformation of the nucleus into a different element.

- γ-decay: An unstable nucleus releases energy in the form of γ-rays, usually following α or β-decay when the daughter nucleus is produced in an excited state.

Let us consider β^+-decay. If a free proton could undergo this process, it would be represented as: $\mathrm{p} \longrightarrow \mathrm{n} + \mathrm{e}^+ + \nu_e$. In the rest frame of the proton, conservation of energy implies: $m_pc^2 = m_nc^2 + m_ec^2 + \text{kinetic energy} + E_{\nu_e}$, where m_n and m_e are the masses of the neutron and positron, respectively, and E_{ν_e} is the energy of the neutrino, whose mass is negligible. This requires $m_pc^2 > m_nc^2 + m_ec^2$. However, this condition is not satisfied, as $m_pc^2 = 938.27$ MeV and $m_nc^2 = 939.57$ MeV.

The situation is different when the proton is within a nucleus because of the binding energy. This is illustrated by the reaction: $(\mathrm{A}, \mathrm{Z}) \longrightarrow (\mathrm{A}, \mathrm{Z}-1) + \mathrm{e}^+ + \nu_e$, where (A,Z) denotes a nucleus with mass number A and atomic number Z. Energy conservation now requires the nuclear mass of (A,Z) to be greater than the sum of the nuclear mass of (A,Z-1) and the mass of the positron. This condition is satisfied, for example, for $A = 11$ and $Z = 6$, corresponding to Carbon-11, which can decay into Boron-11.

Absorption

Absorption of a particle as it collides with another particle at rest can result in either the two particles sticking to each other, or the emission of different particles. Here are a few examples of such processes:

- Photoelectric effect: Absorption of a photon by an atom, causing the atom to eject an electron.
- Excitation: Absorption of a photon by an atom, raising the energy state of the atom.
- Neutron capture: Absorption of a neutron by a nucleus, resulting in the formation of a heavier isotope.

4.5.5 Proton-proton collision and pion production: threshold energy

When two protons collide with sufficient energy, a pion can be produced:

$$\mathrm{p} + \mathrm{p} \longrightarrow \mathrm{p} + \mathrm{p} + \pi^0.$$

In the rest frame of one of the protons before the collision, the 4-momenta are $P_1^\mu = (E_1/c, \mathbf{p}_1)$ and $P_2^\mu = (m_p c, \mathbf{0})$, where E_1 is the energy of the moving proton. Conservation of energy requires the kinetic energy of the moving proton to be equal to $m_\pi c^2$ plus the kinetic energy of the particles after collision, where m_π is the mass of the pion. The minimum value of E_1 required to produce a pion is called the *threshold* energy. This corresponds to all the particles being at rest in the CM frame after the collision. The total 4-momentum in the CM frame is then $P^{*\mu} = (2m_p c + m_\pi c, \mathbf{0})$. The squared length of the total 4-momentum is invariant, meaning[2] $\left(P_1^\mu + P_2^\mu\right)^2 = \left(P^{*\mu}\right)^2$, which implies $\left(P_1^\mu\right)^2 + \left(P_2^\mu\right)^2 + 2P_1^\mu \cdot P_2^\mu = \left(P^{*\mu}\right)^2$. This yields $-2m_p^2c^2 - 2E_1 m_p = -\left(2m_p c + m_\pi c\right)^2$, which simplifies to:

$$E_1 = \frac{\left(2m_p + m_\pi\right)^2 c^2 - 2m_p^2 c^2}{2m_p}.$$

The corresponding threshold kinetic energy of the moving proton is $K = E_1 - m_p c^2 = m_\pi c^2 \left(2 + m_\pi/(2m_p)\right)$, which is about 280 MeV for the π^0 ($m_\pi c^2 \simeq 135$ MeV). This is more than twice the rest energy of the pion: much of the energy of the incident proton is needed to provide the kinetic energy associated with the motion of the centre of momentum, which is conserved.

[2] We could define the total momentum before collision as $P^\mu = (m_p c + E_1/c, \mathbf{p}_1)$ and square it, instead of squaring $P_1^\mu + P_2^\mu$. However, this would introduce p_1^2, as $\left(P^\mu\right)^2 = -\left(m_p c + E_1/c\right)^2 + p_1^2$. Using $p_1^2 c^2 = E_1^2 - m_p^2 c^4$, we obtain $\left(P^\mu\right)^2 = -2m_p^2 c^2 - 2E_1 m_p$. This result can be derived more directly by using $\left(P^\mu\right)^2 = \left(P_1^\mu + P_2^\mu\right)^2 = \left(P_1^\mu\right)^2 + \left(P_2^\mu\right)^2 + 2P_1^\mu \cdot P_2^\mu$, which avoids introducing p_1.

4.5.6 Annihilation processes

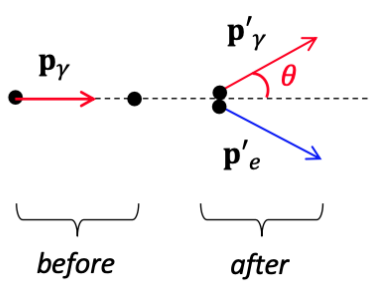
In general, when a particle collides with its antiparticle, they annihilate, producing other particles, typically photons, neutrinos or lighter particles, depending on the available energy. The number and types of particles produced depend on the specific interaction and the energy of the colliding particles. Annihilation leading to photons is a clear example of the equivalence of mass and energy: the entire rest energy of the particle and antiparticle is transformed into the energy of massless particles.

An example of annihilation producing lighter particles is the collision of a proton with an antiproton, which results in the production of pions (π^+, π^-, π^0).

When an electron and positron collide at low energies, they annihilate and produce photons: $\mathrm{e}^- + \mathrm{e}^+ \longrightarrow \gamma + \gamma$. In the centre of momentum frame of the particles before the collision, the total momentum is zero. Therefore, after the collision, momentum conservation requires the two photons to have equal and opposite momenta. In particular, a single photon cannot be produced, since its momentum could not be zero. Additionally, conservation of energy implies that each photon has an energy at least equal to the rest energy $m_e c^2$ of the electron (or positron, as their masses are identical).

4.5.7 Compton scattering

Compton scattering is a phenomenon in which a photon scatters off a charged particle, typically an electron. When a photon interacts with an electron at rest, part of the photon's energy is transferred to the electron, causing it to recoil, while the photon is scattered at an angle with reduced energy, resulting in a longer wavelength. The key result of this interaction is that the change in the photon's wavelength, known as the *Compton shift*, depends on the angle at which the photon is scattered.



The diagram shows the collision in the electron's rest frame. The photon's momentum before and after the collision is $p_\gamma = \hbar k = h/\lambda$ and $p'_\gamma = \hbar k' = h/\lambda'$, respectively. The electron's momentum after the collision is denoted by p'_e. The corresponding photon energies are $E_\gamma = h\nu = hc/\lambda$ and $E'_\gamma = h\nu' = hc/\lambda'$.

The 4-momentum vectors for the photon before and after the collision are:

$$P^\mu_\gamma = \left(\frac{h}{\lambda}, \frac{h}{\lambda}, 0, 0\right), \quad P'^\mu_\gamma = \left(\frac{h}{\lambda'}, \frac{h}{\lambda'}\cos\theta, \frac{h}{\lambda'}\sin\theta, 0\right).$$

The 4-momentum vector for the electron before the collision is $P^\mu_e = (m_e c, \mathbf{0})$, and we denote the 4-momentum after the collision as P'^μ_e.

Conservation of 4-momentum gives:

$$\left(P_\gamma^\mu + P_e^\mu - P_\gamma'^\mu\right)^2 = \left(P_e'^\mu\right)^2 .$$

By isolating $P_e'^\mu$, we avoid involving the unknown momentum $\mathbf{p}_e'$ in the calculation. Using the relations $\left(P_\gamma^\mu\right)^2 = \left(P_\gamma'^\mu\right)^2 = 0$ and $\left(P_e^\mu\right)^2 = \left(P_e'^\mu\right)^2$, we obtain:

$$P_\gamma^\mu \cdot P_e^\mu - P_\gamma^\mu \cdot P_\gamma'^\mu - P_e^\mu \cdot P_\gamma'^\mu = -\frac{h}{\lambda} m_e c + \frac{h}{\lambda}\frac{h}{\lambda'} - \frac{h}{\lambda}\frac{h}{\lambda'}\cos\theta + \frac{h}{\lambda'} m_e c = 0.$$

Finally, solving this expression yields:

$$\lambda' - \lambda = \frac{h}{m_e c}\left(1 - \cos\theta\right). \tag{4.47}$$

For forward scattering ($\theta = 0$), there is no shift in wavelength (i.e., $\lambda' = \lambda$), while for backward scattering ($\theta = \pi$), the shift is maximised and equal to $2h/(m_e c)$. The quantity $h/(m_e c) \simeq 2.43$ pm is called the *Compton wavelength* of the electron.

Chapter 5

Special relativity and electromagnetism

In the second-year Electromagnetism course, it was demonstrated that electric and magnetic fields transform into each other when moving between reference frames, indicating that they cannot be treated as separate entities. In this chapter, we will discuss how these fields combine into a tensor. We begin with an introduction to tensors.

5.1 Tensors

As mentioned earlier, one of the postulates of relativity is that the laws of physics are the same in all inertial frames. Consequently, we aim to express these physical laws in a form that remains invariant under changes in coordinates or reference frames. This is referred to as the *covariant formulation* of physical laws.

This concept was illustrated in chapter 4, where the laws of dynamics were expressed using 4-vectors, which transform according to well-defined rules between reference frames. For example, the law $F^\mu = mA^\mu$ in frame (S) retains its form $F'^\mu = mA'^\mu$ in a frame (S′), moving with velocity v relative to (S), because both 4-vectors transform according to $F'^\mu = \Lambda^\mu{}_\nu F^\nu$ and $A'^\mu = \Lambda^\mu{}_\nu A^\nu$.

More generally, tensors are mathematical objects that transform according to rules that ensure the covariant nature of the physical laws in which they appear. In particular, if all the components of a tensor vanish in one inertial frame, they vanish in all inertial frames. An equation between tensors of the same type therefore holds in every inertial frame if it holds in one.

5.1.1 Preliminary

Consider an ensemble of particles, all with the same mass m and velocity $\mathbf{u}$ in frame (S). In the rest frame (S_0) of the particles, where there are n_0 particles per unit volume, each particle has an energy of mc^2, resulting in an energy density of $n_0 mc^2 = \rho_0 c^2$, where $\rho_0 \equiv n_0 m$ is the mass density.

Frame (S) moves relative to frame (S_0) with velocity $-\mathbf{u}$. Due to length contraction along the direction of this velocity, a unit volume in the rest frame corresponds to a volume of $1/\gamma_u$ in frame (S). Therefore, the number of particles per unit volume in frame (S) is $\gamma_u n_0$. In this frame, each particle has an energy $\gamma_u mc^2$, so the energy density becomes $\gamma_u n_0 \times \gamma_u mc^2 = \gamma_u^2 \rho_0 c^2$. Notably, this equals $\rho_0 \left(u^0\right)^2$, where u^0 is the first component of the 4-velocity:

$$U^\mu = \left(u^0, u^1, u^2, u^3\right) = \left(\gamma_u c, \gamma_u u_x, \gamma_u u_y, \gamma_u u_z\right).$$

By multiplying the components of U^μ with each other and then by ρ_0, we obtain the 16 terms $\rho_0 u^\mu u^\nu$, denoted as $t^{\mu\nu}$:

$$t^{00} = \rho_0 \gamma_u^2 c^2, \quad t^{0i} = \rho_0 \gamma_u^2 c u_i, \quad t^{ij} = \rho_0 \gamma_u^2 u_i u_j,$$

where $i, j = 1 \ldots 3$. As discussed earlier, t^{00} represents the energy density of the ensemble of particles. Writing $t^{0i} = \gamma_u n_0 \times \gamma_u mc \times u_i$, we see that this represents the flux of energy in the i-direction, divided by c. Equivalently, since $\gamma_u m u_i$ is the momentum of a particle, t^{0i} is c times the momentum density in the i-direction, in agreement with the relation $S = gc^2$ derived in section 4.2.2. Finally, $t^{ij} = \gamma_u n_0 \times \gamma_u m u_i \times u_j$ represents the flux of the i–th component of momentum in the j-direction.
These quantities are all physically meaningful, so it is of interest to see how they transform from one frame to another. We denote the set of numbers $t^{\mu\nu}$ as $T^{\mu\nu}$, that is to say $T^{\mu\nu} = \rho_0 U^\mu U^\nu$. Using the transformation properties of the 4-velocity, we have:

$$T'^{\mu\nu} = \rho_0 U'^\mu U'^\nu = \rho_0 \Lambda^\mu{}_\kappa U^\kappa \Lambda^\nu{}_\lambda U^\lambda = \Lambda^\mu{}_\kappa \Lambda^\nu{}_\lambda T^{\kappa\lambda}. \tag{5.1}$$

This indicates that $T^{\mu\nu}$ transforms in a specific manner. This is an example of a contravariant tensor. The equation above can also be written in matrix notation as:

$$T' = \Lambda T \Lambda^\top, \tag{5.2}$$

where T is the matrix with components $T^{\mu\nu}$, and $\Lambda^\top$ denotes the transpose of the Lorentz transformation matrix. Although we generally use index notation, this matrix notation will occasionally be convenient.

We could alternatively define a covariant tensor $T_{\mu\nu} = \rho_0 U_\mu U_\nu$. This transforms as:

$$T'_{\mu\nu} = \rho_0 U'_\mu U'_\nu = \rho_0 \Lambda_\mu{}^\kappa U_\kappa \Lambda_\nu{}^\lambda U_\lambda = \Lambda_\mu{}^\kappa \Lambda_\nu{}^\lambda T_{\kappa\lambda}. \tag{5.3}$$

The relation between these two tensors is:

$$T_{\mu\nu} = \rho_0 U_\mu U_\nu = \rho_0 g_{\mu\kappa} U^\kappa g_{\nu\lambda} U^\lambda = g_{\mu\kappa} g_{\nu\lambda} T^{\kappa\lambda}, \tag{5.4}$$

which shows that we can 'lower' indices as in the case of 4-vectors.
This is a specific case of tensors, and we now define them more generally.

5.1.2 Definition

A tensor of type (m, n) is denoted $T^{\mu_1 \ldots \mu_m}_{\nu_1 \ldots \nu_n}$ and is defined as an object that transforms under a Lorentz transformation according to:

$$\boxed{T'^{\mu_1 \ldots \mu_m}_{\nu_1 \ldots \nu_n} = \Lambda^{\mu_1}{}_{\kappa_1} \ldots \Lambda^{\mu_m}{}_{\kappa_m} \Lambda_{\nu_1}{}^{\lambda_1} \ldots \Lambda_{\nu_n}{}^{\lambda_n} \, T^{\kappa_1 \ldots \kappa_m}_{\lambda_1 \ldots \lambda_n}} \tag{5.5}$$

Using equations (3.58) and (3.60), we can alternatively define a tensor as an object that transforms according to:

$$\boxed{T'^{\mu_1 \ldots \mu_m}_{\nu_1 \ldots \nu_n} = \frac{\partial x'^{\mu_1}}{\partial x^{\kappa_1}} \ldots \frac{\partial x'^{\mu_m}}{\partial x^{\kappa_m}} \frac{\partial x^{\lambda_1}}{\partial x'^{\nu_1}} \ldots \frac{\partial x^{\lambda_n}}{\partial x'^{\nu_n}} \, T^{\kappa_1 \ldots \kappa_m}_{\lambda_1 \ldots \lambda_n}} \tag{5.6}$$

This more general definition applies to any transformation between the sets of coordinates x^μ and x'^μ.

Note that upper and lower indices are sometimes staggered for clarity, though there is no strict convention regarding which should be positioned to the right. The type of the tensor refers to the number m of contravariant indices and the number n of covariant indices. The rank, or order, is given by $m + n$. A tensor T^μ of type $(1, 0)$ is a contravariant vector (rank 1), whereas a tensor T_ν of type $(0, 1)$ is a covariant vector (rank 1). Tensors with both covariant and contravariant indices are referred to as *mixed* tensors. A scalar is conventionally described as a tensor of rank 0. In the following sections, we discuss higher rank tensors.

Equation (5.4) shows that indices can be lowered in the special case where a tensor is formed by multiplying two 4-vectors. We now verify that indices can be raised and lowered more generally, in the same manner as for 4-vectors. For simplicity, we consider a covariant tensor of order 2, $T_{\alpha\beta}$, though the procedure generalises to any mixed type tensor of any order. The tensor transforms as:

$$T'_{\alpha\beta} = \Lambda_\alpha{}^\delta \Lambda_\beta{}^\epsilon \, T_{\delta\epsilon} = g_{\alpha\alpha'} \, \Lambda^{\alpha'}{}_{\alpha''} \, g^{\alpha''\delta} \, g_{\beta\beta'} \, \Lambda^{\beta'}{}_{\beta''} \, g^{\beta''\epsilon} \, T_{\delta\epsilon}$$

where we have used equation (3.55). Next, we multiply both sides by $g^{\zeta\alpha}$ and $g^{\eta\beta}$ (and sum over α and β):

$$\begin{aligned} g^{\zeta\alpha} \, g^{\eta\beta} \, T'_{\alpha\beta} &= g^{\zeta\alpha} \, g_{\alpha\alpha'} \, \Lambda^{\alpha'}{}_{\alpha''} \, g^{\eta\beta} \, g_{\beta\beta'} \, \Lambda^{\beta'}{}_{\beta''} \, g^{\alpha''\delta} \, g^{\beta''\epsilon} \, T_{\delta\epsilon} \\ &= \delta^\zeta_{\alpha'} \, \Lambda^{\alpha'}{}_{\alpha''} \, \delta^\eta_{\beta'} \, \Lambda^{\beta'}{}_{\beta''} \, g^{\alpha''\delta} \, g^{\beta''\epsilon} \, T_{\delta\epsilon} \end{aligned}$$

Thus:

$$\left(g^{\zeta\alpha} \, g^{\eta\beta} \, T'_{\alpha\beta} \right) = \Lambda^\zeta{}_{\alpha''} \, \Lambda^\eta{}_{\beta''} \, \left(g^{\alpha''\delta} \, g^{\beta''\epsilon} \, T_{\delta\epsilon} \right), \tag{5.7}$$

where the brackets indicate the transformed objects. Since both sides obey the defining transformation law of a rank-2 contravariant tensor (eq. [5.5]), the bracketed quantity on each side must be a contravariant tensor.

Hence we identify them as the raised-index tensors:

$$g^{\alpha''\delta}\, g^{\beta''\epsilon}\, T_{\delta\epsilon} = T^{\alpha''\beta''}, \qquad g^{\zeta\alpha}\, g^{\eta\beta}\, T'_{\alpha\beta} = T'^{\zeta\eta}.$$

By the same reasoning, we also obtain $g_{\alpha''\delta}\, g_{\beta''\epsilon}\, T^{\delta\epsilon} = T_{\alpha''\beta''}$.
We summarise the raising and lowering operations as:

$$\boxed{T_{\alpha\beta} = g_{\alpha\delta}\, g_{\beta\epsilon}\, T^{\delta\epsilon}, \quad T^{\alpha\beta} = g^{\alpha\delta}\, g^{\beta\epsilon}\, T_{\delta\epsilon}} \tag{5.8}$$

5.1.3 Operations on tensors

Multiplication

Tensors of rank 2 can be constructed by multiplying two rank 1 tensors, as illustrated by a specific example in the preliminary section (though this is not the most general method for obtaining rank 2 tensors). In chapter 3, while discussing 4-vectors, we defined the scalar product, symbolically denoted as $A^\mu \cdot B^\mu$, as the multiplication of a covariant (row) 4-vector by a contravariant (column) 4-vector, $A_\mu B^\mu$. This operation results in a scalar. We also noted that the product of two covariant 4-vectors, $A_\mu B_\mu$ or two contravariant 4-vectors, $A^\mu B^\mu$, does not follow the usual vector multiplication rules. Such products yield a tensor.

For instance, given two contravariant 4-vectors:

$$A^\mu = \left(a^0, a^1, a^2, a^3\right) \text{ and } B^\nu = \left(b^0, b^1, b^2, b^3\right),$$

we can generate $4^2 = 16$ numbers by multiplying each component of A^μ with each component of B^μ. This set of numbers, $a^\mu b^\nu$, forms a tensor, denoted as:

$$T^{\mu\nu} = A^\mu B^\nu, \tag{5.9}$$

and $a^\mu b^\nu$ are the components of the tensor. It is straightforward to verify that $T^{\mu\nu}$ satisfies the definition [5.5] of a tensor since both A^μ and B^ν transform as 4-vectors:

$$T'^{\mu\nu} = A'^\mu B'^\nu = \Lambda^\mu{}_\kappa A^\kappa \Lambda^\nu{}_\lambda B^\lambda = \Lambda^\mu{}_\kappa \Lambda^\nu{}_\lambda T^{\kappa\lambda}.$$

Just as the scalar product is symbolically represented by $A^\mu \cdot B^\mu$, the tensor product (also known as the outer product) is denoted by $A^\mu \otimes B^\nu$. It is important to note that contraction of indices occurs in the scalar product, but not in the tensor product. The tensor $T^{\mu\nu}$ is of type $(2, 0)$ and has a rank of 2. Similarly, we can generate mixed-type tensors of rank 2 by multiplying the components of a covariant 4-vector A_μ with those of a contravariant 4-vector B^ν: $T^\nu_\mu = A_\mu B^\nu$, where the components of the tensor are the numbers $a_\mu b^\nu$.

More generally, the product of n 4-vectors yields a tensor of rank n, and the product of two tensors of ranks n_1 and n_2 gives a tensor of rank $n_1 + n_2$.

Addition

Only tensors of the same type (m, n) (and therefore the same rank $m + n$) can be added together. This means they must have the same number of covariant and contravariant indices. The resulting tensor has the same type and rank as the original tensors, and is formed by adding the corresponding components of the tensors being summed.

Contraction

We have already encountered index contraction in chapter 3 while dealing with 4-vectors. Now, we extend this concept to more general tensors. Contracting a tensor involves summing over one upper index and one lower index within the same tensor or across different tensors.

Contracting a single tensor $T^{\mu_1 \ldots \mu_m}_{\nu_1 \ldots \nu_n}$ over indices μ_i and ν_j is achieved by setting $\mu_i = \nu_j$, renaming it as, say, a, and summing over all possible values of a. Using Einstein's summation convention, this is written as:

$$T^{\mu_1 \ldots \mu_{i-1}\ a\ \mu_{i+1} \ldots \mu_m}_{\nu_1 \ldots \nu_{j-1}\ a\ \nu_{j+1} \ldots \nu_n}$$

The result is a new tensor $U^{\mu_1 \ldots \mu_{i-1}\mu_{i+1} \ldots \mu_m}_{\nu_1 \ldots \nu_{j-1}\nu_{j+1} \ldots \nu_n}$ which is of order $n + m - 2$. We can verify that this is indeed a tensor by examining its transformation properties:

$$U'^{\mu_1 \ldots \mu_{i-1}\mu_{i+1} \ldots \mu_m}_{\nu_1 \ldots \nu_{j-1}\nu_{j+1} \ldots \nu_n} = T'^{\mu_1 \ldots \mu_{i-1}\ a\ \mu_{i+1} \ldots \mu_m}_{\nu_1 \ldots \nu_{j-1}\ a\ \nu_{j+1} \ldots \nu_n} =$$

$$\Lambda^{\mu_1}{}_{\kappa_1} \ldots \Lambda^{\mu_{i-1}}{}_{\kappa_{i-1}} \Lambda^{a}{}_{\kappa_i} \Lambda^{\mu_{i+1}}{}_{\kappa_{i+1}} \ldots \Lambda^{\mu_m}{}_{\kappa_m} \Lambda_{\nu_1}{}^{\lambda_1} \ldots \Lambda_{\nu_{j-1}}{}^{\lambda_{j-1}} \Lambda_{a}{}^{\lambda_j} \Lambda_{\nu_{j+1}}{}^{\lambda_{j+1}} \ldots \Lambda_{\nu_n}{}^{\lambda_n}\ T^{\kappa_1 \ldots \kappa_m}_{\lambda_1 \ldots \lambda_n}$$

Using the property of the Lorentz transformation matrix given by equation (3.57), which gives $\Lambda^{a}{}_{\kappa_i} \Lambda_{a}{}^{\lambda_j} = \delta^{\lambda_j}_{\kappa_i}$, we simplify this expression to:

$$\Lambda^{\mu_1}{}_{\kappa_1} \ldots \Lambda^{\mu_{i-1}}{}_{\kappa_{i-1}} \Lambda^{\mu_{i+1}}{}_{\kappa_{i+1}} \ldots \Lambda^{\mu_m}{}_{\kappa_m} \Lambda_{\nu_1}{}^{\lambda_1} \ldots \Lambda_{\nu_{j-1}}{}^{\lambda_{j-1}} \Lambda_{\nu_{j+1}}{}^{\lambda_{j+1}} \ldots \Lambda_{\nu_n}{}^{\lambda_n} T^{\kappa_1 \ldots \kappa_{i-1}\ b\ \kappa_{i+1} \ldots \kappa_m}_{\lambda_1 \ldots \lambda_{j-1}\ b\ \lambda_{j+1} \ldots \lambda_n}$$

where we have renamed $b = \kappa_i = \lambda_j$.

By definition, the last term on the right-hand side is equal to $U^{\kappa_1 \ldots \kappa_{i-1}\kappa_{i+1} \ldots \kappa_m}_{\lambda_1 \ldots \lambda_{j-1}\lambda_{j+1} \ldots \lambda_n}$, showing that $U^{\mu_1 \ldots \mu_{i-1}\mu_{i+1}\mu_m}_{\nu_1 \ldots \nu_{j-1}\nu_{j+1} \ldots \nu_n}$ transforms as a tensor.

This calculation assumes that m and n are greater than 1. If $m = n = 1$, contracting T^{μ}_{ν} involves calculating its trace:

$$\boxed{\mathrm{Tr}(T) = T^{\mu}_{\mu}.} \tag{5.10}$$

For a contravariant tensor $T^{\mu\nu}$ (or a covariant tensor $T_{\mu\nu}$), one index must first be lowered (or raised) with the metric, so that the trace is $T^{\mu}_{\mu} = g_{\mu\nu}T^{\mu\nu}$ (or $T^{\mu}_{\mu} = g^{\mu\nu}T_{\mu\nu}$).

When either m or n is equal to 1, contracting indices produces a tensor that is either covariant (if $m = 1$) or contravariant (if $n = 1$).

We can also contract different tensors by summing over matching upper and lower indices, as demonstrated in the following example:

$$U^{\mu\nu}_{\beta\lambda} = T^{\mu\nu}_{\alpha\beta}R^{\alpha}_{\lambda}.$$

The transformation is given by:

$$U'^{\mu\nu}_{\beta\lambda} = T'^{\mu\nu}_{\alpha\beta}R'^{\alpha}_{\lambda} = \Lambda^{\mu}{}_{\mu'}\Lambda^{\nu}{}_{\nu'}\Lambda_{\alpha}{}^{\alpha'}\Lambda_{\beta}{}^{\beta'}\Lambda^{\alpha}{}_{\alpha''}\Lambda_{\lambda}{}^{\lambda''}T^{\mu'\nu'}_{\alpha'\beta'}R^{\alpha''}_{\lambda''}$$

Using the relation $\Lambda^{\alpha}{}_{\alpha''}\Lambda_{\alpha}{}^{\alpha'} = \delta^{\alpha'}_{\alpha''}$ (eq. [3.57]), we obtain:

$$U'^{\mu\nu}_{\beta\lambda} = \Lambda^{\mu}{}_{\mu'}\Lambda^{\nu}{}_{\nu'}\Lambda_{\beta}{}^{\beta'}\Lambda_{\lambda}{}^{\lambda''}T^{\mu'\nu'}_{\alpha'\beta'}R^{\alpha'}_{\lambda''} = \Lambda^{\mu}{}_{\mu'}\Lambda^{\nu}{}_{\nu'}\Lambda_{\beta}{}^{\beta'}\Lambda_{\lambda}{}^{\lambda''}U^{\mu'\nu'}_{\beta'\lambda''}$$

demonstrating that $U^{\mu\nu}_{\beta\lambda}$ is indeed a tensor. This approach can be generalised to the contraction of any number of indices involving any number of mixed tensors.

If $m = n$, a *full* contraction can be performed by summing over all possible pairs of matching indices, as illustrated in the following example:

$$T^{\mu\nu}_{\alpha\beta}R^{\alpha\beta}_{\mu\nu}.$$

This operation reduces the rank of the resulting tensor to zero, yielding a **scalar that is invariant**. This invariance can be demonstrated with a simple case:

$$T'^{\mu\nu}R'_{\mu\nu} = \Lambda^{\mu}{}_{\alpha}\Lambda^{\nu}{}_{\beta}\Lambda_{\mu}{}^{\gamma}\Lambda_{\nu}{}^{\delta}T^{\alpha\beta}R_{\gamma\delta}.$$

Using the properties $\Lambda^{\mu}{}_{\alpha}\Lambda_{\mu}{}^{\gamma} = \delta^{\gamma}_{\alpha}$ and $\Lambda^{\nu}{}_{\beta}\Lambda_{\nu}{}^{\delta} = \delta^{\delta}_{\beta}$, we obtain $T'^{\mu\nu}R'_{\mu\nu} = T^{\alpha\beta}R_{\alpha\beta}$. This demonstrates that the scalar remains invariant under the transformation.

5.1.4 Symmetric and antisymmetric tensors

A tensor is said to be symmetric if it remains unchanged when any two indices of the same type (both contravariant or both covariant) are interchanged. Specifically, a rank-2 tensor $T^{\mu\nu}$ is symmetric if it satisfies the condition $T^{\mu\nu} = T^{\nu\mu}$.

Conversely, a tensor is said to be antisymmetric if its components change sign when any two indices of the same type are interchanged. For a rank-2 tensor $A^{\mu\nu}$, this property is expressed as $A^{\mu\nu} = -A^{\nu\mu}$, which implies that all the diagonal elements of the tensor are zero.

These properties do not depend on the reference frame: if $T^{\mu\nu} = T^{\nu\mu}$, then $T'^{\nu\mu} = \Lambda^{\nu}{}_{\kappa}\Lambda^{\mu}{}_{\lambda}T^{\kappa\lambda} = \Lambda^{\nu}{}_{\kappa}\Lambda^{\mu}{}_{\lambda}T^{\lambda\kappa} = T'^{\mu\nu}$, and similarly for antisymmetric tensors.

A symmetric rank-2 tensor in four dimensions has 10 independent components and an antisymmetric one has 6.

Any rank-2 tensor can be written as the sum of a symmetric and an antisymmetric tensor: $T^{\mu\nu} = \frac{1}{2}\left(T^{\mu\nu} + T^{\nu\mu}\right) + \frac{1}{2}\left(T^{\mu\nu} - T^{\nu\mu}\right)$.

Finally, the full contraction of a symmetric tensor $S^{\mu\nu}$ with an antisymmetric tensor $A_{\mu\nu}$ vanishes, since $S^{\mu\nu}A_{\mu\nu} = S^{\nu\mu}A_{\mu\nu} = -S^{\nu\mu}A_{\nu\mu} = -S^{\mu\nu}A_{\mu\nu}$, where the last equality follows from renaming the dummy indices.

5.2 Invariance of the electric charge

While currents depend on the motion of charges and thus vary across different inertial frames, the electric charge carried by a particle has the same value in every inertial frame. This is an experimental fact, as discussed below.

The force exerted by a moving charge on another charge depends on its velocity, and its electric field is no longer spherically symmetric, as will be shown in Chapter 6. The force on a test charge therefore depends on where the test charge is placed, and does not by itself provide a convenient measure of the moving charge. Instead, we define the electric charge Q within a volume $\mathcal{V}$ using Gauss's law:

$$Q = \epsilon_0 \int_{\Sigma} \mathbf{E} \cdot \hat{\mathbf{n}} \, \mathrm{d}\Sigma. \tag{5.11}$$

Here, the integral is taken over the surface Σ enclosing the volume $\mathcal{V}$, with $\hat{\mathbf{n}}$ representing the outward surface normal. The surface Σ is stationary in frame (S), and the electric field $\mathbf{E}$ is measured at all points of Σ at the same time t in (S), through the force on a test charge at rest in (S). The surface Σ can be chosen arbitrarily, provided it encloses the charges at the time at which the measurement is made. Gauss's law therefore provides a definition of the total charge Q that applies whether the charges are at rest or in motion.

Experimentally, it is found that the charge Q defined in this way has the same value in every inertial frame: if the charge is measured in another inertial frame (S′), using a surface at rest in that frame and enclosing the same charges, the result is again Q. This is what is meant by the **invariance of the electric charge**.[1] Equivalently, the charge of a particle does not depend on its velocity. For example, the electrons in a hydrogen molecule and in a helium atom move with very different velocities relative to the nuclei, yet both are found to be electrically neutral to about one part in 10^{20}. If the charge of a particle depended on its velocity, this could not be the case. Similarly, heating a piece of metal, which increases the speed of its electrons, does not cause it to become charged.

Although the charge is invariant, the charge density is not: combined with length contraction, the invariance of the charge implies that the charge density depends on the frame, as shown in the next section.

5.3 Four-vectors for the current and potential

5.3.1 Definition of the four-current

Consider an ensemble of charges all moving with the same velocity $\mathbf{u}$ in a frame (S). In the rest frame (S_0) of the particles, the charge density is ρ_0. Frame (S) moves relative to frame (S_0) with velocity $-\mathbf{u}$. Due to length contraction in the direction of this velocity, a unit volume in the rest frame corresponds to a volume of $1/\gamma_u$ in frame (S). Additionally, the charge of each particle remains invariant. Therefore, the charge density in (S) is $\rho = \gamma_u \rho_0$ (following the same reasoning as in section 5.1.1).
The current density in frame (S) associated with these charges is $\mathbf{j} = \rho\mathbf{u} = \gamma_u \rho_0 \mathbf{u}$. This is the spatial component of $\rho_0 U^\mu$, where $U^\mu = (\gamma_u c, \gamma_u \mathbf{u})$ is the 4-velocity of a particle. This naturally leads to the definition of the current 4-vector, or 4-current, J^μ as:

$$\boxed{J^\mu = \rho_0 U^\mu = (\gamma_u \rho_0 c, \gamma_u \rho_0 \mathbf{u}) = (\rho c, \mathbf{j})\,.} \tag{5.12}$$

More generally, the charges may have a spread of velocities, as in a gas or in a conductor. The 4-current is then the sum of the contributions from charges with different velocities. Since each contribution transforms as a 4-vector, their sum is also a 4-vector. Thus, quite generally, we may write $J^\mu = (\rho c, \mathbf{j})$ where ρ and $\mathbf{j}$ are the total charge and current densities in the frame considered.

[1] Note that invariance of the charge, i.e. the fact that it has the same value in all inertial frames, is distinct from conservation of the charge, i.e. the fact that the total charge of an isolated system does not change with time. By contrast, energy is conserved but not invariant.

Since J^μ is a 4-vector, in a frame (S′) moving with velocity v relative to (S), it transforms into $J'^\mu = \Lambda^\mu{}_\nu J^\nu$. This implies that the charge density ρ' and current density $\mathbf{j}'$ in (S′) are given by:

$$\rho' c = \Lambda^0{}_\nu J^\nu \quad \text{and} \quad j'_i = \Lambda^i{}_\nu J^\nu.$$

We align the x-axis with the velocity $\mathbf{u}$. Therefore, frame (S) moves relative to frame (S_0) with velocity $-u\hat{\mathbf{x}}$, where $\hat{\mathbf{x}}$ is a unit vector in the x-direction. Now, consider frame (S′) moving with velocity $v\hat{\mathbf{x}}$ relative to (S), and with velocity $w\hat{\mathbf{x}}$ relative to (S_0). The transformation of the 4-current from (S) to (S′) yields:

$$\rho' c = \gamma_v \left(\rho c - \frac{v}{c} j_x\right) = \gamma_v \gamma_u \rho_0 c \left(1 - \frac{vu}{c^2}\right), \tag{5.13}$$

where we have used $\rho = \gamma_u \rho_0$ and $j_x = \gamma_u \rho_0 u$ for the second equality. We can also obtain ρ' by noting that, since in frame (S′) the charges move with velocity $-w\hat{\mathbf{x}}$, the charge density in that frame is $\rho' = \gamma_w \rho_0$. We now verify that this expression for ρ' is consistent with equation (5.13). To express $\gamma_u \gamma_v$ in terms of γ_w, we use the invariance of the scalar product. Consider a particle at rest in frame (S_0). Its 4-vectors in (S_0) and (S) are $U_0^\mu = (c, \mathbf{0})$ and $U^\mu = (\gamma_u c, \gamma_u u\hat{\mathbf{x}})$, respectively. Now consider a particle at rest in (S′). Its 4-velocities in (S_0) and (S) are $V_0^\mu = (\gamma_w c, \gamma_w w\hat{\mathbf{x}})$ and $V^\mu = (\gamma_v c, \gamma_v v\hat{\mathbf{x}})$, respectively. The invariance of the scalar product, $U_0^\mu V_{0\mu} = U^\mu V_\mu$, yields $-\gamma_w c^2 = \gamma_u \gamma_v \left(-c^2 + uv\right)$. Therefore, $\rho' = \gamma_w \rho_0 = \gamma_u \gamma_v \rho_0 \left(1 - uv/c^2\right)$, which is consistent with equation (5.13).

5.3.2 Conservation law

The conservation of charge is expressed by the continuity equation:

$$\frac{\partial \rho}{\partial t} + \boldsymbol{\nabla} \cdot \mathbf{j} = 0. \tag{5.14}$$

Using the gradient 4-vector defined by equation (3.42), this can be written in covariant form as:

$$\boxed{\partial_\mu J^\mu = 0.} \tag{5.15}$$

Since $\partial_\mu J^\mu$ is an invariant, if charge is conserved in one inertial frame, it is conserved in all inertial frames.

5.3.3 Lorenz gauge

Denoting ϕ as the electric scalar potential and $\mathbf{A}$ as the magnetic vector potential, we have:

$$\mathbf{E} = -\boldsymbol{\nabla}\phi - \frac{\partial \mathbf{A}}{\partial t}, \tag{5.16}$$

$$\mathbf{B} = \boldsymbol{\nabla} \times \mathbf{A}. \tag{5.17}$$

Gauge transformation

The magnetic vector potential $\mathbf{A}$ is not uniquely defined, as different potentials can produce the same magnetic field. If both $\mathbf{A}$ and $\tilde{\mathbf{A}}$ correspond to the same field, then $\mathbf{B} = \boldsymbol{\nabla}\times\mathbf{A} = \boldsymbol{\nabla}\times\tilde{\mathbf{A}}$, which implies $\boldsymbol{\nabla}\times(\tilde{\mathbf{A}} - \mathbf{A}) = \mathbf{0}$. The vector $\tilde{\mathbf{A}} - \mathbf{A}$ is curl-free and can thus be written as the gradient of a scalar function Γ, so $\tilde{\mathbf{A}} - \mathbf{A} = \boldsymbol{\nabla}\Gamma$. Therefore:

$$\tilde{\mathbf{A}} = \mathbf{A} + \boldsymbol{\nabla}\Gamma. \tag{5.18}$$

For $\mathbf{E}$ to remain unchanged, ϕ must be transformed into $\tilde{\phi}$ such that:

$$\frac{\partial\tilde{\mathbf{A}}}{\partial t} + \boldsymbol{\nabla}\tilde{\phi} = \frac{\partial\mathbf{A}}{\partial t} + \boldsymbol{\nabla}\phi,$$

which can also be written as:

$$\boldsymbol{\nabla}\tilde{\phi} = \boldsymbol{\nabla}\phi - \frac{\partial}{\partial t}\left(\boldsymbol{\nabla}\Gamma\right).$$

This is satisfied with:

$$\tilde{\phi} = \phi - \partial\Gamma/\partial t. \tag{5.19}$$

Such a transformation from $(\phi, \mathbf{A})$ to $\left(\tilde{\phi}, \tilde{\mathbf{A}}\right)$ is called a *gauge transformation.*

Although the curl of the vector potential is specified by equation (5.17), its divergence can be chosen freely. Specifically, $\boldsymbol{\nabla}\cdot\tilde{\mathbf{A}} = \boldsymbol{\nabla}\cdot\mathbf{A} + \nabla^2\Gamma$. By choosing Γ appropriately, given $\mathbf{A}$, the vector potential $\tilde{\mathbf{A}}$ can be made to have any desired divergence.

Lorenz gauge and wave equations for the potentials

In the presence of electric charges with density ρ and electric currents with density $\mathbf{j}$, Maxwell's equations are given by:

$$\boldsymbol{\nabla}\cdot\mathbf{E} = \frac{\rho}{\epsilon_0}, \tag{5.20}$$

$$\boldsymbol{\nabla}\cdot\mathbf{B} = 0, \tag{5.21}$$

$$\boldsymbol{\nabla}\times\mathbf{E} = -\frac{\partial\mathbf{B}}{\partial t}, \tag{5.22}$$

$$\boldsymbol{\nabla}\times\mathbf{B} = \mu_0\mathbf{j} + \mu_0\epsilon_0\frac{\partial\mathbf{E}}{\partial t}. \tag{5.23}$$

With $\mathbf{E}$ and $\mathbf{B}$ given by equations (5.16) and (5.17), Maxwell's equations (5.21) and (5.22) are satisfied. Next, we substitute equation (5.16) into Gauss's law (5.20):

$$\nabla^2\phi + \frac{\partial}{\partial t}\left(\boldsymbol{\nabla}\cdot\mathbf{A}\right) = -\frac{\rho}{\epsilon_0}. \tag{5.24}$$

Finally, we substitute equations (5.16) and (5.17) into equation (5.23):

$$\boldsymbol{\nabla}\times\left(\boldsymbol{\nabla}\times\mathbf{A}\right) = \mu_0\mathbf{j} - \mu_0\epsilon_0\frac{\partial}{\partial t}\left(\frac{\partial\mathbf{A}}{\partial t} + \boldsymbol{\nabla}\phi\right).$$

Using the identity $\boldsymbol{\nabla}\times(\boldsymbol{\nabla}\times\mathbf{A}) = \boldsymbol{\nabla}(\boldsymbol{\nabla}\cdot\mathbf{A}) - \nabla^2\mathbf{A}$, the above equation can be written as:

$$\nabla^2\mathbf{A} - \mu_0\epsilon_0\frac{\partial^2\mathbf{A}}{\partial t^2} - \boldsymbol{\nabla}\left(\boldsymbol{\nabla}\cdot\mathbf{A} + \mu_0\epsilon_0\frac{\partial\phi}{\partial t}\right) = -\mu_0\mathbf{j}. \tag{5.25}$$

As previously noted, the divergence of the vector potential $\mathbf{A}$ can be chosen freely. We adopt the Lorenz gauge, defined by the condition:

$$\boxed{\boldsymbol{\nabla}\cdot\mathbf{A} = -\mu_0\epsilon_0\frac{\partial\phi}{\partial t}.} \tag{5.26}$$

To show that this condition can always be imposed, suppose that a particular choice of $(\phi, \mathbf{A})$ does not satisfy it. A gauge transformation with function Γ gives:

$$\boldsymbol{\nabla}\cdot\tilde{\mathbf{A}} + \frac{1}{c^2}\frac{\partial\tilde{\phi}}{\partial t} = \boldsymbol{\nabla}\cdot\mathbf{A} + \frac{1}{c^2}\frac{\partial\phi}{\partial t} + \nabla^2\Gamma - \frac{1}{c^2}\frac{\partial^2\Gamma}{\partial t^2}, \tag{5.27}$$

so that $(\tilde{\phi}, \tilde{\mathbf{A}})$ satisfies the Lorenz condition if Γ is a solution of the inhomogeneous wave equation $\nabla^2\Gamma - c^{-2}\partial_t^2\Gamma = -(\boldsymbol{\nabla}\cdot\mathbf{A} + c^{-2}\partial_t\phi)$. We can therefore choose Γ so that the Lorenz condition is satisfied.

Substituting this condition into equations (5.24) and (5.25), we can recast Maxwell's equations into inhomogeneous wave equations for ϕ and $\mathbf{A}$:

$$\nabla^2\phi - \frac{1}{c^2}\frac{\partial^2\phi}{\partial t^2} = -\frac{\rho}{\epsilon_0}, \tag{5.28}$$

$$\nabla^2\mathbf{A} - \frac{1}{c^2}\frac{\partial^2\mathbf{A}}{\partial t^2} = -\mu_0\mathbf{j}, \tag{5.29}$$

where we have used $\mu_0\epsilon_0 = 1/c^2$.

5.3.4 Definition of the four-potential

Given the definitions of the 4-current (5.12) and the d'Alembertian operator (3.51), we can introduce:

$$\boxed{A^\mu = \left(\frac{\phi}{c}, \mathbf{A}\right),} \tag{5.30}$$

so that equations (5.28) and (5.29) can be written as:

$$\boxed{\Box A^\mu = -\mu_0 J^\mu.} \tag{5.31}$$

Since the operator $\Box = \partial_\mu\partial^\mu$ is invariant under Lorentz transformations, as shown in section 3.6.4, and J^μ is a 4-vector, A^μ must also transform as a 4-vector. It is known as the 4-potential. Although the notation is the same as for the 4-acceleration, the context should make it clear which quantity is being referred to.

If ϕ and $\mathbf{A}$ are the scalar and vector potentials in frame (S), then in frame (S′) they transform to ϕ' and $\mathbf{A}'$ such that $A'^{\mu} = \Lambda^{\mu}{}_{\nu} A^{\nu}$. Specifically:

$$\phi'/c = \Lambda^{0}{}_{\nu} A^{\nu} \quad \text{and} \quad A'^{\,i} = \Lambda^{i}{}_{\nu} A^{\nu}.$$

Here, the potentials on both sides refer to the same event: if this event has position 4-vector X^{α} in (S) and $X'^{\alpha} = \Lambda^{\alpha}{}_{\beta} X^{\beta}$ in (S′), then $A'^{\mu}(X'^{\alpha}) = \Lambda^{\mu}{}_{\nu} A^{\nu}(X^{\alpha})$. The potentials ϕ' and $\mathbf{A}'$ are therefore expressed in terms of the coordinates in (S′), and their values mix those of ϕ and $\mathbf{A}$ in (S).

The Lorenz gauge (5.26) can also be expressed in terms of the 4-potential as:

$$\boxed{\partial_{\mu} A^{\mu} = 0.} \tag{5.32}$$

Similarly, the gauge transformation given by equations (5.18) and (5.19) can be written in covariant form as:

$$\boxed{\tilde{A}^{\mu} = A^{\mu} + \partial^{\mu} \Gamma,} \tag{5.33}$$

since $\partial^{0} \Gamma = -\frac{1}{c} \partial_{t} \Gamma$ and $\partial^{i} \Gamma = \partial_{i} \Gamma$. Under this transformation, $\partial_{\mu} \tilde{A}^{\mu} = \partial_{\mu} A^{\mu} + \Box \Gamma$, which is the covariant form of equation (5.27).

The invariance of the d'Alembertian operator $\Box$ ensures that equation (5.31), which encapsulates all of Maxwell's equations, is in a covariant form. In other words, if A^{μ} and J^{μ} transform to A'^{μ} and J'^{μ} in another inertial frame, then $\Box A'^{\mu} = -\mu_0 J'^{\mu}$ still holds. This guarantees that the wave equation for the 4-potential maintains the same form across different inertial frames, thereby preserving the consistency of physical laws in relativity. Similarly, the invariance of the scalar product ensures that the gauge condition (5.32) is also in a covariant form.

We have successfully rewritten Maxwell's equations in an explicitly covariant form without altering their original structure, as they were inherently covariant from the outset. In fact, this is closely tied to the historical development of the Lorentz transformation. Recognising that Maxwell's equations were not invariant under Galilean transformations, Lorentz derived a transformation that preserved their invariance. Subsequently, Poincaré demonstrated that these transformations formed a group, all of which occurred prior to the development of the theory of special relativity. This inherent covariance of Maxwell's equations contrasts with Newton's laws, which are invariant under Galilean transformations but not under Lorentz transformations. Deriving a covariant form of the equation of motion, as presented in chapter 4, required the introduction of a relativistic definition of momentum.

5.4 Electromagnetic tensor

We now express the electric and magnetic fields in an arbitrary inertial frame in terms of the components of the 4-potential:

$$A^\mu = \left(A^0, A^1, A^2, A^3\right) = \left(\frac{\phi}{c}, A_x, A_y, A_z\right).$$

Equations (5.16) and (5.17) can be rewritten as:

$$\frac{E_x}{c} = -\frac{\partial\left(\phi/c\right)}{\partial x} - \frac{\partial A_x}{c\partial t} = -\partial^1 A^0 + \partial^0 A^1, \tag{5.34}$$

$$\frac{E_y}{c} = -\frac{\partial\left(\phi/c\right)}{\partial y} - \frac{\partial A_y}{c\partial t} = -\partial^2 A^0 + \partial^0 A^2, \tag{5.35}$$

$$\frac{E_z}{c} = -\frac{\partial\left(\phi/c\right)}{\partial z} - \frac{\partial A_z}{c\partial t} = -\partial^3 A^0 + \partial^0 A^3, \tag{5.36}$$

$$B_x = \frac{\partial A_z}{\partial y} - \frac{\partial A_y}{\partial z} = \partial^2 A^3 - \partial^3 A^2, \tag{5.37}$$

$$B_y = \frac{\partial A_x}{\partial z} - \frac{\partial A_z}{\partial x} = \partial^3 A^1 - \partial^1 A^3, \tag{5.38}$$

$$B_z = \frac{\partial A_y}{\partial x} - \frac{\partial A_x}{\partial y} = \partial^1 A^2 - \partial^2 A^1, \tag{5.39}$$

where we have used the definition (3.45) of the contravariant 4-gradient ∂^μ. We express the derivatives using ∂^μ rather than ∂_μ to maintain a consistent pattern of signs in the expressions for the components of $\mathbf{E}$ and $\mathbf{B}$.

5.4.1 Definition of the electromagnetic tensor

Since both ∂^μ and A^ν are contravariant 4-vectors, $\partial^\mu A^\nu$ is a rank-2 contravariant tensor. We can therefore define the electromagnetic tensor as:

$$\boxed{F^{\mu\nu} = \partial^\mu A^\nu - \partial^\nu A^\mu,} \tag{5.40}$$

which can be written in matrix form as:

$$F^{\mu\nu} = \begin{pmatrix} F^{00} & F^{01} & F^{02} & F^{03} \\ F^{10} & F^{11} & F^{12} & F^{13} \\ F^{20} & F^{21} & F^{22} & F^{23} \\ F^{30} & F^{31} & F^{32} & F^{33} \end{pmatrix} = \begin{pmatrix} 0 & E_x/c & E_y/c & E_z/c \\ -E_x/c & 0 & B_z & -B_y \\ -E_y/c & -B_z & 0 & B_x \\ -E_z/c & B_y & -B_x & 0 \end{pmatrix}. \tag{5.41}$$

This tensor is antisymmetric, which means that although it has $4 \times 4 = 16$ components, only 6 of them are independent, corresponding to the components of the electric and magnetic fields. The tensor $F^{\mu\nu}$ encapsulates the electromagnetic field. Because it is a tensor, it provides a compact and elegant way to express Maxwell's equations in a covariant form, as will be shown below, and it allows the electromagnetic field to be transformed consistently between different inertial frames.

Under the gauge transformation given by equation (5.33), we obtain:

$$\tilde{F}^{\mu\nu} = F^{\mu\nu} + \partial^\mu\partial^\nu\Gamma - \partial^\nu\partial^\mu\Gamma = F^{\mu\nu}.$$

Therefore, the electromagnetic tensor is gauge invariant, as expected since it contains only the fields **E** and **B**.

Lowering both indices gives the covariant components of the electromagnetic tensor:

$$\boxed{F_{\mu\nu} = g_{\mu\kappa}g_{\nu\lambda}F^{\kappa\lambda} = \partial_\mu A_\nu - \partial_\nu A_\mu.} \tag{5.42}$$

For the Minkowski metric, which has only diagonal components, this can be written as $F_{\mu\nu} = g_{\mu\mu}g_{\nu\nu}F^{\mu\nu}$ (no summation over μ and ν), so that:

$$F_{\mu\nu} = \begin{pmatrix} 0 & -E_x/c & -E_y/c & -E_z/c \\ E_x/c & 0 & B_z & -B_y \\ E_y/c & -B_z & 0 & B_x \\ E_z/c & B_y & -B_x & 0 \end{pmatrix}. \tag{5.43}$$

The contravariant tensor can be reconstructed from the covariant one by multiplying both sides of equation (5.42) by $g^{\alpha\mu}g^{\beta\nu}$ and summing over μ and ν, resulting in:

$$g^{\alpha\mu}g^{\beta\nu}F_{\mu\nu} = F^{\alpha\beta}. \tag{5.44}$$

5.4.2 Maxwell's equations in terms of the electromagnetic tensor

We can now express Maxwell's equation (5.31) using the electromagnetic tensor:

$$\Box A^\mu \equiv \partial_\nu\partial^\nu A^\mu = \partial_\nu\left(\partial^\mu A^\nu - F^{\mu\nu}\right) = \partial^\mu\partial_\nu A^\nu - \partial_\nu F^{\mu\nu} = -\partial_\nu F^{\mu\nu},$$

where we have applied the gauge condition (5.32). This yields:

$$\boxed{\partial_\nu F^{\mu\nu} = \mu_0 J^\mu.} \tag{5.45}$$

Although the Lorenz gauge was used to derive it, this equation involves only $F^{\mu\nu}$, which does not depend on the choice of gauge. It therefore holds in any gauge, as confirmed by the direct verification below.

Equation (5.31) explicitly represents only the inhomogeneous Maxwell's equations (5.20) and (5.23). However, the use of the 4-potential naturally ensures that the homogeneous equations (5.21) and (5.22) are also satisfied. This is because these equations are automatically fulfilled when the electric and magnetic fields are expressed in terms of the potentials.

When we replace equation (5.31) with equation (5.45), we still retain the inhomogeneous equations, but the homogeneous ones are no longer guaranteed if $F^{\mu\nu}$ is regarded as the fundamental field. Let us first verify that equation (5.45) does indeed capture the inhomogeneous equations. For $\mu = 0$, this equation gives $\partial_\nu F^{0\nu} = \boldsymbol{\nabla} \cdot \mathbf{E}/c = \mu_0 \rho c$, which corresponds to Gauss's law $\boldsymbol{\nabla} \cdot \mathbf{E} = \rho/\epsilon_0$. For $\mu = i$, with $i = 1 \ldots 3$, we obtain:

$$\frac{1}{c}\frac{\partial F^{i0}}{\partial t} + \frac{\partial F^{i1}}{\partial x} + \frac{\partial F^{i2}}{\partial y} + \frac{\partial F^{i3}}{\partial z} = \mu_0 j_i.$$

By substituting the components of the electromagnetic field tensor, it can be readily shown that these three equations yield $\boldsymbol{\nabla} \times \mathbf{B} = \mu_0 \mathbf{j} + \mu_0 \epsilon_0 \left(\partial \mathbf{E}/\partial t\right)$. Thus, equation (5.45) is equivalent to the two inhomogeneous Maxwell's equations (5.20) and (5.23). To obtain the complete set of Maxwell's equations, we need to include the two homogeneous equations. These are compactly expressed by the *Bianchi identity* for the electromagnetic tensor:

$$\boxed{\partial_\lambda F_{\mu\nu} + \partial_\mu F_{\nu\lambda} + \partial_\nu F_{\lambda\mu} = 0.} \tag{5.46}$$

This relation follows directly from the definition (5.42): substituting $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$, the six terms cancel in pairs because partial derivatives commute. It holds for all index triples (λ, μ, ν). The choice $(\lambda, \mu, \nu) = (1, 2, 3)$ yields $\nabla \cdot \mathbf{B} = 0$, while $(\lambda, \mu, \nu) = (0, i, j)$ with i and j any distinct spatial indices in $(1, 2, 3)$ gives the three components of Faraday's law $\nabla \times \mathbf{E} = -\partial \mathbf{B}/\partial t$.

5.5 Lorentz transformation of the electromagnetic field

Consider an inertial frame (S), where the electric and magnetic fields are denoted by $\mathbf{E}$ and $\mathbf{B}$, respectively, and another inertial frame (S′) moving with a velocity $\mathbf{v}$ relative to (S). We align the x-axis with the direction of $\mathbf{v}$. As a rank-2 tensor, $F^{\mu\nu}$ transforms under a Lorentz transformation according to the rule:

$$F'^{\mu\nu}\left(X'^{\alpha}\right) = \Lambda^{\mu}{}_{\kappa} \Lambda^{\nu}{}_{\lambda} F^{\kappa\lambda}\left(X^{\alpha}\right),$$

where, as for the 4-potential, X^α and X'^α are the position 4-vectors of the same event in (S) and (S′). This can also be written in matrix notation as:

$$F' = \Lambda F \Lambda^{\top}.$$

This implies that the electric and magnetic fields in (S′), denoted by $\mathbf{E}'$ and $\mathbf{B}'$, are related to those in (S) through the following transformations:

$$\frac{E'_x}{c} = F'^{01} = \gamma^2 \frac{E_x}{c} - \beta^2\gamma^2 \frac{E_x}{c} = \frac{E_x}{c}, \tag{5.47}$$

$$\frac{E'_y}{c} = F'^{02} = \gamma \frac{E_y}{c} - \beta\gamma B_z = \frac{\gamma}{c}\left(E_y - vB_z\right), \tag{5.48}$$

$$\frac{E'_z}{c} = F'^{03} = \gamma \frac{E_z}{c} + \beta\gamma B_y = \frac{\gamma}{c}\left(E_z + vB_y\right), \tag{5.49}$$

$$B'_x = F'^{23} = B_x, \tag{5.50}$$

$$B'_y = F'^{31} = \beta\gamma \frac{E_z}{c} + \gamma B_y = \gamma\left(B_y + \frac{v}{c^2}E_z\right), \tag{5.51}$$

$$B'_z = F'^{12} = -\beta\gamma \frac{E_y}{c} + \gamma B_z = \gamma\left(B_z - \frac{v}{c^2}E_y\right). \tag{5.52}$$

These transformations can also be compactly expressed as:

$$\boxed{\begin{aligned} &\mathbf{E}'_{\parallel} = \mathbf{E}_{\parallel}, \qquad \mathbf{E}'_{\perp} = \gamma\left(\mathbf{E}_{\perp} + \mathbf{v}\times\mathbf{B}\right), && (5.53)\\ &\mathbf{B}'_{\parallel} = \mathbf{B}_{\parallel}, \qquad \mathbf{B}'_{\perp} = \gamma\left(\mathbf{B}_{\perp} - \frac{\mathbf{v}\times\mathbf{E}}{c^2}\right), && (5.54) \end{aligned}}$$

where the subscripts $\parallel$ and $\perp$ denote the components parallel and perpendicular to the velocity $\mathbf{v}$, respectively.

In particular, a purely electric field in one frame ($\mathbf{B} = \mathbf{0}$) gives rise to a magnetic field $\mathbf{B}' = -\gamma\,\mathbf{v}\times\mathbf{E}/c^2$ in a frame moving relative to it, and conversely. Whether a field is electric or magnetic therefore depends on the frame of reference.

These field transformations follow directly from the tensorial nature of the electromagnetic field. They can also be obtained by considering the transformation of charge densities and currents between reference frames. Although this alternative approach is more laborious, it provides a useful physical interpretation of the mixing of electric and magnetic fields. This derivation is presented in Appendix B.

Invariants of the electromagnetic field

Although $\mathbf{E}$ and $\mathbf{B}$ depend on the reference frame, certain combinations of them take the same value in all inertial frames. Since $F^{\mu\nu}$ is a tensor, its full contraction with itself, $F^{\mu\nu}F_{\mu\nu}$, is invariant. Using equations (5.41) and (5.43), we obtain:

$$F^{\mu\nu}F_{\mu\nu} = 2\left(B^2 - \frac{E^2}{c^2}\right), \tag{5.55}$$

so that:

$$\boxed{\mathbf{B}^2 - \frac{\mathbf{E}^2}{c^2} \quad \text{is invariant.}} \tag{5.56}$$

Also, as can be verified directly using the Lorentz transformations of the electric and magnetic fields derived above:

$$\boxed{\mathbf{E}\cdot\mathbf{B} \quad \text{is invariant.}} \tag{5.57}$$

For example, for a Lorentz boost along the x–direction, equations (5.47)–(5.52) give:

$$\begin{aligned}\mathbf{E}'\cdot\mathbf{B}' &= E_xB_x + \gamma^2(E_y - vB_z)\left(B_y + \frac{vE_z}{c^2}\right) + \gamma^2(E_z + vB_y)\left(B_z - \frac{vE_y}{c^2}\right)\\ &= E_xB_x + E_yB_y + E_zB_z = \mathbf{E}\cdot\mathbf{B}.\end{aligned}$$

These invariants have useful consequences. If $\mathbf{E}$ and $\mathbf{B}$ are perpendicular in one frame, they are perpendicular in all frames. If $|\mathbf{E}| = c|\mathbf{B}|$ in one frame, as for a plane electromagnetic wave, this holds in all frames. If $\mathbf{E}\cdot\mathbf{B} \neq 0$, there is no frame in which the field is purely electric or purely magnetic.

5.6 The Lorentz force

It is an experimental fact that the force acting on a particle with (invariant) charge q and velocity $\mathbf{u}$ in a reference frame where the electric and magnetic fields are $\mathbf{E}$ and $\mathbf{B}$, respectively, is given by the Lorentz force law:

$$\mathbf{f} = q\,(\mathbf{E} + \mathbf{u}\times\mathbf{B}) \tag{5.58}$$

This expression leads to the following equation for the rate of change of the particle's momentum:

$$\frac{\mathrm{d}\mathbf{p}}{\mathrm{d}t} = q\,(\mathbf{E} + \mathbf{u}\times\mathbf{B}) \tag{5.59}$$

Since the mass of the particle is constant, the rate of change of the particle's relativistic energy is given by:

$$\frac{\mathrm{d}E}{\mathrm{d}t} = \mathbf{f}\cdot\mathbf{u} = q\mathbf{E}\cdot\mathbf{u} \tag{5.60}$$

As discussed in section 4.3, these equations can be encapsulated in the relativistic form of the equation of motion:

$$F^\mu = \frac{\mathrm{d}P^\mu}{\mathrm{d}\tau}, \quad \text{with} \quad F^\mu = \left(\frac{\gamma_u}{c}\frac{\mathrm{d}E}{\mathrm{d}t}, \gamma_u\mathbf{f}\right). \tag{5.61}$$

We can also express F^μ using the electromagnetic tensor:

$$\boxed{F^\mu = qF^{\mu\nu}U_\nu,} \tag{5.62}$$

where $U_\nu = (-\gamma_u c, \gamma_u\mathbf{u})$ is the covariant 4-velocity. It is straightforward to verify that this expression correctly yields the 4-force F^μ in terms of $\mathbf{E}$ and $\mathbf{B}$. For example, for $\mu = 0$, $F^{0\nu}U_\nu = F^{0i}U_i = \gamma_u\mathbf{E}\cdot\mathbf{u}/c$, so that $F^0 = q\gamma_u\mathbf{E}\cdot\mathbf{u}/c = (\gamma_u/c)\,\mathrm{d}E/\mathrm{d}t$, in agreement with equation (5.61). The spatial components follow similarly, using $F^{i0}U_0 = \gamma_u E_i$.

Equation (5.62) yields a covariant form of the relativistic equation of motion for a particle with charge q in an electromagnetic field:

$$\boxed{\frac{\mathrm{d}P^\mu}{\mathrm{d}\tau} = qF^{\mu\nu}U_\nu.} \tag{5.63}$$

The 4-force is orthogonal to the 4-velocity: $F^\mu U_\mu = qF^{\mu\nu}U_\mu U_\nu = 0$, since this is the full contraction of the antisymmetric tensor $F^{\mu\nu}$ with the symmetric tensor $U_\mu U_\nu$ (section 5.1.4). As shown by equation (4.32), this implies that the rest mass of a particle subject to the electromagnetic force remains constant, consistent with the assumption made in deriving equation (5.60).

Both expressions (5.61) and (5.62) for F^μ are equivalent and, within an inertial frame, equally simple to evaluate since they give the force in terms of the fields and the particle's motion in that frame. The tensor form has the additional advantage that it makes the geometric structure of the force law explicit, since it is written entirely as a tensor contraction $F^\mu = qF^{\mu\nu}U_\nu$, a relation that is valid in any inertial frame.

5.7 Lagrangian formulation of electromagnetism

As discussed in Chapter 1, the Lagrangian formalism simplifies the derivation of conserved quantities such as energy and momentum. Here, we derive the Lagrangian for a relativistic particle in an electromagnetic field, as well as for the electromagnetic field itself, in covariant form.

5.7.1 Covariant relativistic Lagrangian for a particle

Although the free-particle part of the relativistic Lagrangian in equation (4.7) leads to a Lorentz-invariant action, this invariance is not immediately apparent, since the Lagrangian itself is not invariant: only $L\,\mathrm{d}t = -mc^2\,\mathrm{d}\tau$ is. To make it explicit, we reformulate the action in terms of the particle's displacement 4-vector. Consider a particle with position 4-vector X^μ. The spacetime interval along the worldline of the particle is:

$$\mathrm{d}s^2 = \mathrm{d}X^\mu \mathrm{d}X_\mu = g^{\mu\nu}\mathrm{d}X_\nu \mathrm{d}X_\mu = -c^2\mathrm{d}\tau^2.$$

This yields:

$$\sqrt{-g^{\mu\nu}\mathrm{d}X_\nu \mathrm{d}X_\mu} = c\mathrm{d}\tau. \tag{5.64}$$

Using this relation, the action for a free particle can be written as:

$$S = \int_A^B -mc^2\mathrm{d}\tau = \int_A^B -mc\sqrt{-g^{\mu\nu}\mathrm{d}X_\nu \mathrm{d}X_\mu}. \tag{5.65}$$

The worldline of the particle can be parametrised by the proper time τ, or by any other parameter θ that monotonically increases with τ. Using a general parameter θ rather than τ is essential for the variational principle. If the worldline is parametrised by τ, the four components $\mathrm{d}X^\mu/\mathrm{d}\tau$ are not independent, since they satisfy $g_{\mu\nu}\,(\mathrm{d}X^\mu/\mathrm{d}\tau)(\mathrm{d}X^\nu/\mathrm{d}\tau) = -c^2$, and the Lagrangian would reduce to the constant $-mc^2$. With a general parameter θ, the four components $\mathrm{d}X^\mu/\mathrm{d}\theta$ can be varied independently. Once the Euler-Lagrange equations have been derived, we are free to choose $\theta = \tau$.

Expressing the differentials in terms of θ, we have:

$$\mathrm{d}X_\mu = \frac{\mathrm{d}X_\mu}{\mathrm{d}\theta}\mathrm{d}\theta.$$

Substituting this into the action, we obtain:

$$S = \int_{\theta_A}^{\theta_B} -mc\sqrt{-g^{\mu\nu}\frac{\mathrm{d}X_\nu}{\mathrm{d}\theta'}\frac{\mathrm{d}X_\mu}{\mathrm{d}\theta'}}\mathrm{d}\theta', \tag{5.66}$$

where θ_A and θ_B are the values of θ at points A and B on the worldline.

Therefore, the Lagrangian for a free particle can be expressed as:

$$L_{\text{free}} = -mc\sqrt{-g^{\mu\nu}\frac{\mathrm{d}X_\nu}{\mathrm{d}\theta}\frac{\mathrm{d}X_\mu}{\mathrm{d}\theta}}. \tag{5.67}$$

It depends on the components of the 4-vector X_μ and on their derivatives with respect to θ. This form makes the Lorentz invariance of the action explicit, as it is expressed directly in terms of the spacetime metric and the particle's trajectory.

Next, we introduce the interaction with the electromagnetic field, which is expressed similarly in both classical and relativistic mechanics. The generalised potential accounting for the electromagnetic force in the classical limit is given by equation (1.35), and it remains valid here. This implies that the term to be added to the action is:

$$\int_{t_A}^{t_B} -q\,(\phi - \dot{\mathbf{r}}\cdot\mathbf{A})\,\mathrm{d}t. \tag{5.68}$$

As mentioned earlier, a gauge transformation leaves the electromagnetic field unchanged. Thus, the action S' resulting from such a transformation can differ from the original action S only by a boundary term. This ensures that the trajectory which makes S stationary also makes S' stationary. For this condition to hold, the Lagrangian after transformation, L', must either equal the original Lagrangian L or differ from it by a total time derivative.

We now verify that this condition is satisfied when the term $-q\,(\phi - \dot{\mathbf{r}}\cdot\mathbf{A})$ is added to L.

Under the gauge transformation $(\phi, \mathbf{A}) \to \left(\tilde{\phi}, \tilde{\mathbf{A}}\right)$, as defined in equations (5.18) and (5.19), the change in Lagrangian is:

$$L' - L = q\left(\frac{\partial \Gamma}{\partial t} + \dot{\mathbf{r}} \cdot \boldsymbol{\nabla}\Gamma\right) = q\frac{\mathrm{d}\Gamma}{\mathrm{d}t},$$

where the total time derivative is computed along the particle's trajectory. This confirms that gauge symmetry is satisfied.

Using the potential and displacement 4-vectors, equation (5.68) can be rewritten as:

$$\int_{t_A}^{t_B} qA^\mu \frac{\mathrm{d}X_\mu}{\mathrm{d}t}\mathrm{d}t = \int_A^B qA^\mu \mathrm{d}X_\mu.$$

Since $A^\mu dX_\mu$ is a scalar product of two 4-vectors, this term is manifestly Lorentz invariant.

To make this consistent with the previous expression (5.66) for the action, we parametrise the worldline by the same parameter θ, yielding the total action:

$$S = \int_{\theta_A}^{\theta_B} \left(-mc\sqrt{-g^{\mu\nu}\frac{\mathrm{d}X_\nu}{\mathrm{d}\theta'}\frac{\mathrm{d}X_\mu}{\mathrm{d}\theta'}} + qA^\mu \frac{\mathrm{d}X_\mu}{\mathrm{d}\theta'}\right)\mathrm{d}\theta'. \tag{5.69}$$

The Lagrangian therefore becomes:

$$\boxed{L = -mc\sqrt{-g^{\mu\nu}\frac{\mathrm{d}X_\nu}{\mathrm{d}\theta}\frac{\mathrm{d}X_\mu}{\mathrm{d}\theta}} + qA^\mu \frac{\mathrm{d}X_\mu}{\mathrm{d}\theta}.} \tag{5.70}$$

Since A^μ is a function of X_μ, here again L depends on the components of the 4-vector X_μ and on their derivatives with respect to θ.

5.7.2 Euler-Lagrange equations

As discussed in chapter 3, the notation X_μ has been used to represent both the 4-vector and its components. To make the analysis clearer, we now distinguish the position 4-vector X_μ from its individual components x_α and, similarly, the potential 4-vector A^μ from its components a^α, where $\alpha = 0 \ldots 3$.

The Lagrangian L depends on the variables x_α and their derivatives $\mathrm{d}x_\alpha/\mathrm{d}\theta$, similar to the Lagrangian we discussed in chapter 1 for classical mechanics, which depended on the generalised coordinates q_i and their time derivatives. As before, applying the principle of least action yields the Euler-Lagrange equations in the same form as equations (1.20):

$$\frac{\mathrm{d}}{\mathrm{d}\theta}\left(\frac{\partial L}{\partial\left(\frac{\mathrm{d}x_\alpha}{\mathrm{d}\theta}\right)}\right) - \frac{\partial L}{\partial x_\alpha} = 0, \quad \text{for} \quad \alpha = 0 \ldots 3. \tag{5.71}$$

Next, we verify that these equations lead to the equation of motion (5.63).

Writing the Lagrangian in terms of the variables x_α and $\mathrm{d}x_\alpha/\mathrm{d}\theta$ was essential to obtain the Euler-Lagrange equations (5.71). To proceed with the unpacking of these equations, we need to revert to the proper time.
We have:

$$\frac{\partial L}{\partial\left(\dfrac{\mathrm{d}x_\alpha}{\mathrm{d}\theta}\right)} = \frac{mc}{2}\frac{g^{\alpha\nu}\dfrac{\mathrm{d}x_\nu}{\mathrm{d}\theta} + g^{\mu\alpha}\dfrac{\mathrm{d}x_\mu}{\mathrm{d}\theta}}{\sqrt{-g^{\mu\nu}\dfrac{\mathrm{d}x_\nu}{\mathrm{d}\theta}\dfrac{\mathrm{d}x_\mu}{\mathrm{d}\theta}}} + qa^\alpha = m\frac{\dfrac{\mathrm{d}x^\alpha}{\mathrm{d}\theta}}{\dfrac{\mathrm{d}\tau}{\mathrm{d}\theta}} + qa^\alpha, \tag{5.72}$$

where we used equation (5.64) and the identities $g^{\alpha\nu}x_\nu = g^{\mu\alpha}x_\mu = x^\alpha$, applied here to the derivatives $\mathrm{d}x_\nu/\mathrm{d}\theta$. Now, writing:

$$\frac{\mathrm{d}x^\alpha}{\mathrm{d}\theta} = \frac{\mathrm{d}x^\alpha}{\mathrm{d}\tau}\frac{\mathrm{d}\tau}{\mathrm{d}\theta},$$

we obtain the first term on the left-hand side of the Euler-Lagrange equations (5.71):

$$\frac{\mathrm{d}}{\mathrm{d}\theta}\left(\frac{\partial L}{\partial\left(\dfrac{\mathrm{d}x_\alpha}{\mathrm{d}\theta}\right)}\right) = \frac{\mathrm{d}}{\mathrm{d}\theta}\left(m\frac{\mathrm{d}x^\alpha}{\mathrm{d}\tau} + qa^\alpha\right) = \left(m\frac{\mathrm{d}^2x^\alpha}{\mathrm{d}\tau^2} + q\frac{\mathrm{d}a^\alpha}{\mathrm{d}\tau}\right)\frac{\mathrm{d}\tau}{\mathrm{d}\theta}. \tag{5.73}$$

The second term on the left-hand side of the Euler-Lagrange equations is:

$$\frac{\partial L}{\partial x_\alpha} = q\frac{\partial a^\mu}{\partial x_\alpha}\frac{\mathrm{d}x_\mu}{\mathrm{d}\theta} = q\frac{\partial a^\mu}{\partial x_\alpha}\frac{\mathrm{d}x_\mu}{\mathrm{d}\tau}\frac{\mathrm{d}\tau}{\mathrm{d}\theta}. \tag{5.74}$$

Thus, the Euler-Lagrange equations become:

$$m\frac{\mathrm{d}^2x^\alpha}{\mathrm{d}\tau^2} + q\frac{\mathrm{d}a^\alpha}{\mathrm{d}\tau} - q\frac{\partial a^\mu}{\partial x_\alpha}\frac{\mathrm{d}x_\mu}{\mathrm{d}\tau} = 0. \tag{5.75}$$

Since a^α depends on the coordinates x_μ, we can write:

$$\frac{\mathrm{d}a^\alpha}{\mathrm{d}\tau} = \frac{\partial a^\alpha}{\partial x_\mu}\frac{\mathrm{d}x_\mu}{\mathrm{d}\tau},$$

which allows us to rewrite the equation as:

$$m\frac{\mathrm{d}^2x^\alpha}{\mathrm{d}\tau^2} = q\left(\frac{\partial a^\mu}{\partial x_\alpha} - \frac{\partial a^\alpha}{\partial x_\mu}\right)\frac{\mathrm{d}x_\mu}{\mathrm{d}\tau}. \tag{5.76}$$

Returning to 4-vector notation, and noting that $\partial/\partial x_\alpha = \partial^\alpha$, the term in parentheses is $\partial^\alpha a^\mu - \partial^\mu a^\alpha = F^{\alpha\mu}$ (see eq. [5.40]), and $\mathrm{d}X_\mu/\mathrm{d}\tau = U_\mu$. On the left-hand side, we have:

$$m\frac{\mathrm{d}}{\mathrm{d}\tau}\left(\frac{\mathrm{d}X^\alpha}{\mathrm{d}\tau}\right) = m\frac{\mathrm{d}U^\alpha}{\mathrm{d}\tau} = \frac{\mathrm{d}P^\alpha}{\mathrm{d}\tau}. \tag{5.77}$$

This confirms that equation (5.76), derived from the Euler-Lagrange equations, is identical to the covariant equation of motion (5.63).

5.7.3 Lagrangian for the electromagnetic field

Electromagnetic field in vacuum

In chapter 1, we discussed that the Lagrangian density for the electromagnetic field in a specific inertial frame, in the absence of charges and currents, is given by:

$$\mathcal{L} = \frac{\epsilon_0}{2}\mathbf{E}^2 - \frac{1}{2\mu_0}\mathbf{B}^2. \tag{5.78}$$

In Problem Set 1, it was shown that the Euler-Lagrange equations for this Lagrangian, when expressed in terms of the potentials ϕ and $\mathbf{A}$, do indeed yield Maxwell's equations.

As for a particle (section 4.2.1), the action must be Lorentz invariant. For a field, the action is $S = \int \mathcal{L}\,\mathrm{d}^3x\,\mathrm{d}t = c^{-1}\int \mathcal{L}\,\mathrm{d}^4x$, where $\mathrm{d}^4x = c\,\mathrm{d}t\,\mathrm{d}^3x$ is invariant under Lorentz transformations (eq. [3.62]). Therefore, unlike the Lagrangian of a particle, which is not invariant (only $L\,\mathrm{d}t$ is), the Lagrangian density $\mathcal{L}$ must itself be Lorentz invariant. This is not apparent from equation (5.78), which is expressed in terms of the fields in a particular frame. Using equation (5.55), we can rewrite it as:

$$\mathcal{L} = -\frac{1}{4\mu_0}F^{\mu\nu}F_{\mu\nu}, \tag{5.79}$$

which is manifestly invariant.

Interaction with charges and currents

In the presence of charges and currents, an interaction term must be added to the Lagrangian density (5.79). As discussed in Chapter 1, the interaction Lagrangian density for a continuous distribution of charge, with charge density ρ and current density $\mathbf{j}$, is given by equation (1.50), which we reproduce here for ease of reference:

$$\mathcal{L}_{\text{int}} = -\rho\,\phi + \mathbf{j}\cdot\mathbf{A}. \tag{5.80}$$

Using the 4-current $J^\mu = (\rho c, \mathbf{j})$ and the 4-potential $A^\mu = (\phi/c, \mathbf{A})$, we have $\mathcal{L}_{\text{int}} = J^\mu A_\mu$.

The Lagrangian density for the electromagnetic field interacting with charges and currents is thus:

$$\boxed{\mathcal{L} = -\frac{1}{4\mu_0}F^{\mu\nu}F_{\mu\nu} + J^\mu A_\mu.} \tag{5.81}$$

Both $F^{\mu\nu}F_{\mu\nu}$ and $J^\mu A_\mu$ are invariant, so this expression is manifestly covariant.

The Euler-Lagrange equations obtained by varying the 4-potential:

$$\partial_\mu\left(\frac{\partial\mathcal{L}}{\partial(\partial_\mu A_\nu)}\right) - \frac{\partial\mathcal{L}}{\partial A_\nu} = 0,$$

yield Maxwell's equations with sources, as shown in Problem Set 1.

Under the gauge transformation $\tilde{A}_\mu = A_\mu + \partial_\mu \Gamma$, the interaction term changes by $J^\mu \partial_\mu \Gamma = \partial_\mu (J^\mu \Gamma) - \Gamma\, \partial_\mu J^\mu$. The first term is a total divergence, which contributes only a boundary term to the action and therefore does not affect the field equations. The second term vanishes because the 4-current is conserved (eq. [5.15]). The action is therefore gauge invariant, up to a boundary term.

In the following section, we first consider the electromagnetic field in vacuum, for which $J^\mu = 0$, before returning to the exchange of energy and momentum between the field and charged matter.

5.8 Energy and momentum conservation for fields in vacuum

We now derive an equation for the conservation of energy and momentum in the electromagnetic field, formulated in covariant form. To identify the terms corresponding to energy density and energy flux, we make use of Noether's theorem.

5.8.1 Noether's theorem for the electromagnetic field

In the absence of charges and currents, the Lagrangian density does not depend explicitly on the spacetime coordinates, so the action is invariant under time and spatial translations. In Problem Set 1, we derived Noether's theorem for systems invariant under time translation. For a Lagrangian density $\mathcal{L}$ that depends on a field ϕ and its derivatives, this theorem is expressed as:

$$\partial_t \left(\frac{\partial \mathcal{L}}{\partial (\partial_t \phi)} \partial_t \phi - \mathcal{L} \right) + \partial_i \left(\frac{\partial \mathcal{L}}{\partial (\partial_i \phi)} \partial_t \phi \right) = 0, \tag{5.82}$$

where summation over $i = 1 \ldots 3$ is implied. Using 4-vector notation, this can be rewritten as:

$$\partial_\mu \left(\frac{\partial \mathcal{L}}{\partial (\partial_\mu \phi)} \partial_0 \phi - \delta^\mu_0 \mathcal{L} \right) = 0, \tag{5.83}$$

where summation over $\mu = 0 \ldots 3$ is implied. This relation was derived by considering a time translation and noting that it leaves the system invariant. By considering a spatial translation in the i-direction, we obtain the same result with ∂_0 replaced by ∂_i. Thus, the following relations hold for $\nu = 0 \ldots 3$:

$$\partial_\mu \left(\frac{\partial \mathcal{L}}{\partial (\partial_\mu \phi)} \partial_\nu \phi - \delta^\mu_\nu \mathcal{L} \right) = 0, \tag{5.84}$$

This result extends to cases where the Lagrangian depends on N fields ϕ_k. In such cases, ϕ in the expression above is replaced by ϕ_k, with the first term in parentheses summed over k.

For the electromagnetic field, the Lagrangian is given by equation (5.79). Using equation (5.40), it can be seen that this Lagrangian depends on the fields A^ρ, which are the components of the 4-potential. Thus, ϕ_k is replaced by A^ρ, yielding:

$$\boxed{\partial_\mu \left(\frac{\partial \mathcal{L}}{\partial \left(\partial_\mu A^\rho \right)} \partial_\nu A^\rho - \delta^\mu_\nu \mathcal{L} \right) = 0.} \tag{5.85}$$

This yields four conservation laws.

5.8.2 Canonical energy-momentum tensor

We now define the canonical *stress-energy tensor*, also referred to as the canonical *energy-momentum tensor*, as:

$$T^\mu_\lambda = -\frac{\partial \mathcal{L}}{\partial \left(\partial_\mu A^\rho \right)} \partial_\lambda A^\rho + \delta^\mu_\lambda \mathcal{L}, \tag{5.86}$$

where the choice of sign for T^μ_λ will be justified later. The fact that this is a mixed tensor will be demonstrated below. Multiplying by $g^{\nu\lambda}$ and summing over λ gives the contravariant tensor:

$$\boxed{T^{\mu\nu} = -\frac{\partial \mathcal{L}}{\partial \left(\partial_\mu A^\rho \right)} \partial^\nu A^\rho + g^{\nu\mu} \mathcal{L}.} \tag{5.87}$$

This tensor satisfies the conservation law:

$$\boxed{\partial_\mu T^{\mu\nu} = 0.} \tag{5.88}$$

The term *canonical* reflects the construction of the tensor from the Lagrangian, while the names 'stress-energy' or 'energy-momentum tensor' will be clarified below. The fact that $T^{\mu\nu}$ is a tensor follows from the invariance of $\mathcal{L}$ and the covariant transformation properties of all components in its definition, such as $\partial^\nu A^\rho$ and $g^{\nu\mu}$, under Lorentz transformations.

Using $F_{\alpha\beta} = g_{\alpha\kappa} g_{\beta\lambda} F^{\kappa\lambda}$ and $F^{\alpha\beta}$ from equation (5.40), the Lagrangian density (5.79) can be rewritten as:

$$\mathcal{L} = -\frac{1}{4\mu_0} \underbrace{g_{\alpha\kappa} g_{\beta\lambda} \left(\partial^\kappa A^\lambda - \partial^\lambda A^\kappa \right)}_{F_{\alpha\beta}} \underbrace{\left(\partial^\alpha A^\beta - \partial^\beta A^\alpha \right)}_{F^{\alpha\beta}}, \tag{5.89}$$

$$= -\frac{1}{4\mu_0} \underbrace{\left(g_{\beta\lambda} \partial_\alpha A^\lambda - g_{\alpha\kappa} \partial_\beta A^\kappa \right)}_{F_{\alpha\beta}} \underbrace{\left(g^{\alpha\gamma} \partial_\gamma A^\beta - g^{\beta\epsilon} \partial_\epsilon A^\alpha \right)}_{F^{\alpha\beta}}. \tag{5.90}$$

This yields:

$$\frac{\partial \mathcal{L}}{\partial \left(\partial_\mu A^\rho\right)} = -\frac{1}{4\mu_0}\left(g_{\beta\rho}F^{\mu\beta} - g_{\alpha\rho}F^{\alpha\mu} + g^{\alpha\mu}F_{\alpha\rho} - g^{\beta\mu}F_{\rho\beta}\right) \tag{5.91}$$

$$= -\frac{1}{2\mu_0}\left(g_{\alpha\rho}F^{\mu\alpha} + g^{\alpha\mu}F_{\alpha\rho}\right), \tag{5.92}$$

where we used the antisymmetry of $F^{\mu\alpha}$ and $F_{\mu\alpha}$ to obtain the second equality. Writing:

$$g_{\alpha\rho}F^{\mu\alpha} = g_{\alpha\rho}g^{\mu\xi}g^{\alpha\zeta}F_{\xi\zeta} = g^{\mu\xi}F_{\xi\rho},$$

equation (5.92) becomes:

$$\frac{\partial \mathcal{L}}{\partial \left(\partial_\mu A^\rho\right)} = -\frac{1}{\mu_0}g^{\alpha\mu}F_{\alpha\rho}. \tag{5.93}$$

This expression represents a mixed tensor, which supports the classification of T^μ_λ in equation (5.86) as a mixed tensor. Substituting this result into equation (5.87) yields:

$$\boxed{T^{\mu\nu} = \frac{1}{\mu_0}g^{\alpha\mu}F_{\alpha\rho}\partial^\nu A^\rho + g^{\nu\mu}\mathcal{L}.} \tag{5.94}$$

For $\nu = 0$, equation (5.88) gives $\partial_t T^{00} + c\partial_i T^{i0} = 0$, which represents energy conservation. Here, T^{00} corresponds to the energy density and cT^{i0} to the energy flux. Note that the choice of the minus sign in equation (5.86) ensures that T^{00} directly represents the energy density of the system, rather than its negative. From equation (5.94), we find:

$$T^{00} = -\frac{1}{\mu_0}F_{0\rho}\,\partial^0 A^\rho - \mathcal{L}$$

Using $F_{0\rho}$ from equation (5.43) and $\mathcal{L}$ from equation (5.78), we obtain:

$$T^{00} = -\frac{1}{\mu_0 c^2}\mathbf{E}\cdot\partial_t\mathbf{A} - \frac{\epsilon_0}{2}\mathbf{E}^2 + \frac{1}{2\mu_0}\mathbf{B}^2 = \frac{\epsilon_0}{2}\mathbf{E}^2 + \frac{1}{2\mu_0}\mathbf{B}^2 + \epsilon_0\boldsymbol{\nabla}\cdot(\phi\mathbf{E}), \tag{5.95}$$

where we used $\mathbf{E} = -\boldsymbol{\nabla}\phi - \partial_t\mathbf{A}$ and $\boldsymbol{\nabla}\cdot\mathbf{E} = 0$.

There is an additional divergence term compared to the standard definition of energy density. However, when T^{00} is integrated over the entire volume of the system (defined to fully contain the fields such that they vanish at the boundaries) this divergence term does not contribute to the integral. Thus, the total energy of the system is given by the integral of T^{00}. Similarly, the energy flux cT^{i0} and the momentum density T^{0i}/c obtained from the canonical tensor differ from the Poynting vector and from $\mathbf{S}/c^2$, respectively. As shown in the next section, these differences do not affect the total energy and momentum of the field.

5.8.3 Symmetric energy-momentum tensor

The extra terms in the canonical tensor indicate that quantities like energy density and flux are not uniquely determined, although the total energy and momentum of the fields are unambiguously defined. If we add to the energy-momentum tensor a tensor $Q^{\mu\nu}$ satisfying $\partial_\mu Q^{\mu\nu} = 0$, the conservation laws are preserved.

This flexibility allows us to construct a symmetric, traceless and gauge-invariant energy-momentum tensor $\mathcal{T}^{\mu\nu} = T^{\mu\nu} + Q^{\mu\nu}$. Gauge invariance is not satisfied by $T^{\mu\nu}$, as it involves the potential, and the tensor is neither symmetric nor traceless. Symmetry is required for the angular momentum of the field to be conserved. The vanishing of the trace is related to the invariance of the free electromagnetic field under scale transformations, $X^\mu \to \lambda X^\mu$, and more generally under conformal transformations. This invariance arises because the field contains no intrinsic mass or length scale (the photon is massless).

Using $\partial^\nu A^\rho = -F^{\rho\nu} + \partial^\rho A^\nu$ from equation (5.40) and $\mathcal{L}$ from equation (5.79), we can rewrite equation (5.94) as:

$$T^{\mu\nu} = -\frac{1}{\mu_0}\left(g^{\alpha\mu}F_{\alpha\rho}F^{\rho\nu} + \frac{1}{4}g^{\nu\mu}F^{\alpha\beta}F_{\alpha\beta}\right) + \frac{1}{\mu_0}g^{\alpha\mu}F_{\alpha\rho}\partial^\rho A^\nu. \tag{5.96}$$

The negative of the last term in the above equation can be expressed as:

$$Q^{\mu\nu} \equiv -\frac{1}{\mu_0}g^{\alpha\mu}F_{\alpha\rho}\partial^\rho A^\nu = -\frac{1}{\mu_0}F^{\mu\beta}\partial_\beta A^\nu = -\frac{1}{\mu_0}\partial_\beta\left(F^{\mu\beta}A^\nu\right), \tag{5.97}$$

where we used equation (5.45) which implies that, in vacuum, $\partial_\beta F^{\mu\beta} = 0$. Taking the divergence of $Q^{\mu\nu}$ gives:

$$\partial_\mu Q^{\mu\nu} = -\frac{1}{\mu_0}\partial_\mu\partial_\beta\left(F^{\mu\beta}A^\nu\right) = -\frac{1}{\mu_0}\partial_\beta\partial_\mu\left(F^{\beta\mu}A^\nu\right), \tag{5.98}$$

where the second equality follows from swapping the indices μ and β. Since $F^{\beta\mu} = -F^{\mu\beta}$ and partial derivatives commute, we conclude that $\partial_\mu Q^{\mu\nu} = 0$. Moreover, since $F^{00} = 0$, the components $Q^{0\nu} = -\mu_0^{-1}\partial_j\left(F^{0j}A^\nu\right)$ are spatial divergences. Integrated over a volume enclosing the whole field, they vanish. Adding $Q^{\mu\nu}$ therefore does not change the conserved quantities $\int T^{0\nu}\,\mathrm{d}^3x$ associated with equation (5.88): the total energy of the field for $\nu = 0$ (section 5.8.2) and, as shown in section 5.8.4, c times its total momentum for $\nu = i$.

Consequently, we can define the energy-momentum tensor as $\mathcal{T}^{\mu\nu} = T^{\mu\nu} + Q^{\mu\nu}$, with $T^{\mu\nu}$ and $Q^{\mu\nu}$ defined above. This results in:

$$\boxed{\mathcal{T}^{\mu\nu} = -\frac{1}{\mu_0}\left(g^{\alpha\mu}F_{\alpha\rho}F^{\rho\nu} + \frac{1}{4}g^{\nu\mu}F^{\alpha\beta}F_{\alpha\beta}\right).} \tag{5.99}$$

This tensor satisfies the conservation law:

$$\boxed{\partial_\mu \mathcal{T}^{\mu\nu} = 0,} \tag{5.100}$$

and is gauge invariant, since it depends only on $F^{\mu\nu}$. It is symmetric since both $F^{\mu\nu}$ and $F_{\mu\nu}$ are antisymmetric. Finally, it is traceless since:

$$\mathrm{Tr}(T^{\mu\nu}) = g_{\mu\nu}T^{\mu\nu} = -\frac{1}{\mu_0}\left(g_{\mu\nu}g^{\alpha\mu}F_{\alpha\rho}F^{\rho\nu} + g_{\mu\nu}g^{\nu\mu}F^{\alpha\beta}F_{\alpha\beta}\right),$$

with $g_{\mu\nu}g^{\alpha\mu}F_{\alpha\rho}F^{\rho\nu} = F_{\nu\rho}F^{\rho\nu} = -F^{\alpha\beta}F_{\alpha\beta}$ and $g_{\mu\nu}g^{\nu\mu} = 4$, which yields $\mathrm{Tr}(T^{\mu\nu}) = 0$.

5.8.4 Physical interpretation of the energy-momentum tensor

The components of $\mathcal{T}^{\mu\nu}$ are given as:

$$\boxed{\begin{aligned} \mathcal{E} \equiv \mathcal{T}^{00} &= \frac{\epsilon_0 \mathbf{E}^2}{2} + \frac{\mathbf{B}^2}{2\mu_0}, && (5.101)\\ \frac{S_i}{c} \equiv \mathcal{T}^{0i} &= \frac{1}{\mu_0 c}\left(\mathbf{E}\times\mathbf{B}\right)_i, && (5.102)\\ -\sigma_{ij} \equiv \mathcal{T}^{ij} &= -\epsilon_0 E_i E_j - \frac{1}{\mu_0}B_i B_j + \delta_{ij}\left(\frac{\epsilon_0 \mathbf{E}^2}{2} + \frac{\mathbf{B}^2}{2\mu_0}\right), && (5.103) \end{aligned}}$$

and we now discuss their physical significance.

Conservation of energy

Writing equation (5.100) for $\nu = 0$ gives Poynting's theorem:

$$\frac{\partial \mathcal{E}}{\partial t} + \boldsymbol{\nabla}\cdot\mathbf{S} = 0, \tag{5.104}$$

where $\mathcal{E}$ is the **energy density** and $\mathbf{S}$ is the Poynting vector, i.e. the **flux of energy**.

Integrating this equation over a finite volume $\mathcal{V}$ yields:

$$\int_{\mathcal{V}} \frac{\partial \mathcal{E}}{\partial t}\mathrm{d}v = -\int_{\Sigma} \mathbf{S}\cdot\hat{\mathbf{n}}\mathrm{d}\sigma, \tag{5.105}$$

where Σ is the surface enclosing the volume $\mathcal{V}$ and $\hat{\mathbf{n}}$ is the unit normal vector to the surface element $\mathrm{d}\sigma$, directed outwards. This shows that energy can only decrease (increase) in the volume by flowing out (in) through its surfaces.

Integrating over the entire volume of the system, in which the field is confined, yields the **Hamiltonian** of the system:

$$\boxed{H \equiv \int \mathcal{E}\mathrm{d}v = \text{const.}} \tag{5.106}$$

Conservation of momentum

For $\nu = i$, equation (5.100) becomes:

$$\frac{\partial}{\partial t}\left(\frac{S_i}{c^2}\right) - \partial_j \sigma_{ij} = 0. \tag{5.107}$$

This equation expresses conservation of momentum, where $\mathbf{g} \equiv \mathbf{S}/c^2$ represents the **momentum density of the field** and σ_{ij} is the **Maxwell stress** (force per unit area).

This extends to the electromagnetic field the general relation between energy flux and momentum density derived for massive particles in section 4.2.2 (eq. [4.16]). Consider, as an illustration, a locally unidirectional radiation field. Let $\mathcal{E}(\mathbf{r})$ be the energy density of the field and $\hat{\mathbf{n}}(\mathbf{r})$ its local direction of propagation. Since the energy propagates at velocity $c\hat{\mathbf{n}}(\mathbf{r})$, the energy flux is $\mathbf{S}(\mathbf{r}) = \mathcal{E}(\mathbf{r})c\hat{\mathbf{n}}(\mathbf{r})$. Each photon carries momentum equal to its energy divided by c, directed along $\hat{\mathbf{n}}$, so that the corresponding momentum density is $\mathbf{g}(\mathbf{r}) = (\mathcal{E}(\mathbf{r})/c)\,\hat{\mathbf{n}}(\mathbf{r})$. Comparing the two expressions gives:

$$\mathbf{S} = \mathbf{g}c^2. \tag{5.108}$$

Although illustrated here for the simple case of locally unidirectional radiation, this relation holds more generally for the electromagnetic field, as already follows from equation (5.107), whose derivation does not rely on the photon picture.

Integrating equation (5.107) over a finite volume $\mathcal{V}$ gives:

$$\int_{\mathcal{V}} \frac{\partial g_i}{\partial t} \mathrm{d}v = \int_{\Sigma} \sigma_{ij} n_j \mathrm{d}\sigma, \tag{5.109}$$

where Σ is the surface enclosing the volume $\mathcal{V}$ and n_j is the j-component of the unit normal vector to the surface element $\mathrm{d}\sigma$, directed outwards. The term on the left-hand side represents the rate of change of electromagnetic momentum within the volume. The right-hand side represents the net flux of electromagnetic momentum through the surface enclosing the volume. Thus, **$-\sigma_{ij}$ is the i-component of the electromagnetic momentum flux in the j-direction. Equivalently, σ_{ij} is the i-component of the force per unit area (stress) exerted across a surface element whose normal vector points in the j-direction, by the field on the side towards which the normal points, on the field on the other side**. In vacuum, this stress describes the transport of momentum by the field and is not associated with a mechanical force on matter. The stress tensor is symmetric, with $\sigma_{ij} = \sigma_{ji}$.

Integrating over the entire volume of the system gives the **total momentum** of the electromagnetic field:

$$\boxed{\mathbf{P}_{\text{field}} \equiv \int \frac{\mathbf{S}}{c^2}\, \mathrm{d}v = \textbf{const.}} \tag{5.110}$$

These properties make it clear why $\mathcal{T}^{\mu\nu}$ is referred to as either the *stress-energy* or *energy-momentum* tensor. Its common form is:

$$\boxed{\mathcal{T}^{\mu\nu} = \left(\begin{array}{c|c} \mathcal{E} & c\mathbf{g} \\ \hline c\mathbf{g} & -\sigma_{ij} \end{array}\right)} \tag{5.111}$$

5.9 Exchange of energy and momentum with charges

So far, we have considered the energy and momentum of the electromagnetic field in vacuum. In the presence of charges and currents, the field can exchange energy and momentum with matter, and the energy-momentum tensor of the electromagnetic field is therefore no longer conserved separately.

5.9.1 Energy-momentum exchange equation

The tensor $\mathcal{T}^{\mu\nu}$ defined by equation (5.99) depends only on the electromagnetic field, and we take it as the energy-momentum tensor of the field also in the presence of charges and currents. However, we cannot repeat the derivation of section 5.8 to obtain its divergence. If $J^\mu(\mathbf{r},t)$ is treated as a prescribed external 4-current, the Lagrangian density (5.81) depends explicitly on the spacetime coordinates. The system is then no longer invariant under translations in space and time, and Noether's theorem no longer yields a conservation law. Instead, we calculate the divergence of $\mathcal{T}^{\mu\nu}$ directly, using only Maxwell's equations. The result therefore holds whether the charges and currents are prescribed or move under the action of the field.

Taking the divergence of equation (5.99) gives:

$$\mu_0\partial_\mu\mathcal{T}^{\mu\nu} = \underbrace{-\,\partial_\mu\left(g^{\alpha\mu}F_{\alpha\rho}\right)F^{\rho\nu}}_{I_1}\ \underbrace{-\,g^{\alpha\mu}F_{\alpha\rho}\partial_\mu F^{\rho\nu}}_{I_2}\ \underbrace{-\frac{1}{4}\partial_\mu\left(g^{\nu\mu}F^{\alpha\beta}F_{\alpha\beta}\right)}_{I_3}. \tag{5.112}$$

First, consider I_3, which can be written as:

$$I_3 = -\frac{1}{4}g^{\nu\mu}\partial_\mu\left(F^{\alpha\beta}F_{\alpha\beta}\right) = -\frac{1}{4}\partial^\nu(F^{\alpha\beta}F_{\alpha\beta}) = -\frac{1}{2}F^{\alpha\beta}\partial^\nu F_{\alpha\beta}.$$

Next, we calculate I_1 as:

$$I_1 = -\partial_\mu\left(F^\mu_\rho\right)F^{\rho\nu} = -\partial_\mu\left(g_{\alpha\rho}F^{\mu\alpha}\right)F^{\rho\nu} = -g_{\alpha\rho}\partial_\mu\left(F^{\mu\alpha}\right)F^{\rho\nu} = g_{\alpha\rho}\partial_\mu\left(F^{\alpha\mu}\right)F^{\rho\nu},$$

where we have used the antisymmetry of $F^{\mu\alpha}$ in the last equality. From Maxwell's equations (5.45), and using again the antisymmetry of $F^{\rho\nu}$, it follows that:

$$I_1 = g_{\alpha\rho}\mu_0 J^\alpha F^{\rho\nu} = \mu_0 J_\rho F^{\rho\nu} = -\mu_0 J_\rho F^{\nu\rho}.$$

Finally, we consider I_2, which can be written as:

$$I_2 = -F^{\mu}_{\rho}\partial_\mu F^{\rho\nu} = -F^{\mu\alpha}g_{\alpha\rho}\partial_\mu F^{\rho\nu} = -F^{\mu\alpha}\partial_\mu F^{\nu}_{\alpha} = -g^{\nu\rho}F^{\mu\alpha}\partial_\mu F_{\alpha\rho}.$$

To simplify this expression, we use the Bianchi identity (5.46) in the form:

$$\partial_\mu F_{\alpha\rho} + \partial_\alpha F_{\rho\mu} + \partial_\rho F_{\mu\alpha} = 0.$$

Multiplying by $F^{\mu\alpha}$ gives:

$$F^{\mu\alpha}\partial_\mu F_{\alpha\rho} + F^{\mu\alpha}\partial_\alpha F_{\rho\mu} + F^{\mu\alpha}\partial_\rho F_{\mu\alpha} = 0.$$

The first two terms on the left-hand side are equal. Indeed, interchanging the dummy indices μ and α in the second term gives $F^{\alpha\mu}\partial_\mu F_{\rho\alpha} = F^{\mu\alpha}\partial_\mu F_{\alpha\rho}$, where we have used the antisymmetry of $F^{\mu\alpha}$ and $F_{\rho\alpha}$. The Bianchi identity therefore gives:

$$2F^{\mu\alpha}\partial_\mu F_{\alpha\rho} + F^{\mu\alpha}\partial_\rho F_{\mu\alpha} = 0,$$

or:

$$F^{\mu\alpha}\partial_\mu F_{\alpha\rho} = -\frac{1}{2}F^{\mu\alpha}\partial_\rho F_{\mu\alpha}.$$

Substituting this result into the expression for I_2, we obtain:

$$I_2 = \frac{1}{2}g^{\nu\rho}F^{\mu\alpha}\partial_\rho F_{\mu\alpha} = \frac{1}{2}F^{\mu\alpha}\partial^\nu F_{\mu\alpha}.$$

Thus $I_2 = -I_3$, and equation (5.112) reduces to:

$$\boxed{\partial_\mu \mathcal{T}^{\mu\nu} = -F^{\nu\rho}J_\rho.} \tag{5.113}$$

In vacuum, $J_\rho = 0$ and we recover the conservation law (5.100). The right-hand side involves only the sources through Maxwell's equations, while the Bianchi identity, i.e. the homogeneous Maxwell's equations, ensures that all the other terms cancel.

5.9.2 Energy and momentum conservation

We define:

$$\mathcal{F}^\nu \equiv F^{\nu\rho}J_\rho, \tag{5.114}$$

the term on the right-hand side of equation (5.113). Since it is obtained by contracting a rank-2 tensor with a 4-vector, $\mathcal{F}^\nu$ is itself a 4-vector. Writing $\mathcal{F}^\nu = \big(f^0, \mathbf{f}\big)$, we calculate:

$$cf^0 = \mathbf{j}\cdot\mathbf{E}, \tag{5.115}$$

$$\mathbf{f} = \rho\mathbf{E} + \mathbf{j}\times\mathbf{B}. \tag{5.116}$$

The time component cf^0 is the power transferred by the electromagnetic field to the charges per unit volume, whereas the spatial part $\mathbf{f}$ is the Lorentz force per unit volume.

This has the same form as the Lorentz 4-force derived in section 5.6, $F^\nu = qF^{\nu\rho}U_\rho$ (eq. [5.62]). To see this, consider charges of charge q with proper number density n_0, moving with 4-velocity U^μ, so that $J^\mu = n_0 q\, U^\mu$. Then $\mathcal{F}^\nu = n_0\, qF^{\nu\rho}U_\rho = n_0 F^\nu$. Using $F^\nu = \mathrm{d}P^\nu/\mathrm{d}\tau$, this yields $\mathcal{F}^\nu = n_0\, \mathrm{d}P^\nu/\mathrm{d}\tau$. Since the number density in the frame considered is $n = \gamma_u n_0$ and $\mathrm{d}t = \gamma_u\, \mathrm{d}\tau$, this is $\mathcal{F}^\nu = n\, \mathrm{d}P^\nu/\mathrm{d}t$: the rate of change of the 4-momentum of the charges per unit volume. Therefore, $\mathcal{F}^\nu$ is called the 4-force density.

The divergence of the electromagnetic energy-momentum tensor can then be written as:

$$\boxed{\partial_\mu \mathcal{T}^{\mu\nu} = -\mathcal{F}^\nu.} \tag{5.117}$$

With this additional term on the right-hand side, we can write equations for $\nu = 0$ and $\nu = i$ by adding the appropriate component of $-\mathcal{F}^\nu$ on the right-hand side of equations (5.104) and (5.107). This yields:

$$\frac{\partial \mathcal{E}}{\partial t} + \boldsymbol{\nabla} \cdot \mathbf{S} = -\mathbf{j} \cdot \mathbf{E}, \tag{5.118}$$

and:

$$\frac{\partial}{\partial t}\left(\frac{S_i}{c^2}\right) - \partial_j \sigma_{ij} = -f_i. \tag{5.119}$$

Equation (5.118) is Poynting's theorem in the presence of charges and currents. In the absence of charges, the energy density of the electromagnetic field can only decrease locally ($\partial\mathcal{E}/\partial t < 0$) if there is a net outward flow of energy ($\boldsymbol{\nabla} \cdot \mathbf{S} > 0$), or vice-versa. When charges are present though, energy can also be transferred to the charges, where $\mathbf{j} \cdot \mathbf{E} > 0$ indicates power transferred from the field to the charges. The magnetic field does not appear in this equation because the magnetic force does no work.

Integrating equation (5.118) over the entire system, with the boundary chosen so that there is no energy flux through it, gives:

$$\frac{\mathrm{d}U_{\text{field}}}{\mathrm{d}t} + \int \mathbf{j} \cdot \mathbf{E}\, \mathrm{d}v = 0, \tag{5.120}$$

where $U_{\text{field}} = \int \mathcal{E}\, \mathrm{d}v$ is the energy in the electromagnetic field. The second term is the rate at which the field does work on the charges, i.e. the rate of change of their total energy U_{charges}. The equation above can therefore be rewritten as:

$$\boxed{U_{\text{field}} + U_{\text{charges}} = \text{const.}} \tag{5.121}$$

With $\mathbf{g} = \mathbf{S}/c^2$, we similarly integrate equation (5.119) over the entire system, which yields:

$$\int \frac{\partial \mathbf{g}}{\partial t} \mathrm{d}v + \int \mathbf{f} \mathrm{d}v = 0, \tag{5.122}$$

The first term on the left-hand side can be written as $\mathrm{d}\mathbf{P}_{\text{field}}/\mathrm{d}t$, where $\mathbf{P}_{\text{field}}$ defined by equation (5.110) is the total momentum of the electromagnetic field. The second term is the total force exerted by the field on the charges, and therefore represents the rate of change of the total momentum $\mathbf{P}_{\text{charges}}$ of the charges. The equation above can therefore be written as:

$$\boxed{\mathbf{P}_{\text{field}} + \mathbf{P}_{\text{charges}} = \textbf{const}.} \tag{5.123}$$

Equations (5.121) and (5.123) show that both energy and momentum can be transferred between the electromagnetic field and the charges.

Chapter 6

Radiation from moving charges

Electromagnetic waves are generally associated with a nonzero Poynting vector, which describes the transport of energy by the wave. The power passing through the surface of a sphere at infinity is equal to the flux of the Poynting vector through that surface. If this flux is zero, no energy is carried away to infinity. We define radiation as the energy flux that propagates to infinity.

When the electric and magnetic fields decay as the inverse square of the distance, the Poynting vector falls off as the inverse fourth power of the distance. In such cases, the power passing through a sphere at infinity is zero, and no radiation is produced.

Systems that radiate energy include accelerating charges, which we will examine in this chapter. Accelerating charges, such as those in oscillating antennas, produce electromagnetic waves that carry information over long distances, which is the basis for radio, television, and satellite communications. Electromagnetic radiation also enables us to see the stars, communicate via satellites, and receive information from the early universe through the cosmic microwave background, among many other phenomena.

In this chapter, we determine the electromagnetic field produced by moving charges, and calculate the energy they radiate.

6.1 Solution of the wave equation

Before solving the wave equation satisfied by the potential 4-vector for general charge and current densities, we first examine the electromagnetic field produced by a charge moving with uniform velocity. This specific case provides valuable insight into how relativity influences the structure and orientation of the field lines.

6.1.1 Field due to a uniformly moving charge

Consider a particle with charge q moving with uniform velocity $\mathbf{v}$ in an inertial frame (S). In its rest frame (S′), where the charge is stationary at the origin, the electromagnetic field at a point P with position vector $\mathbf{r}'$ is given by:

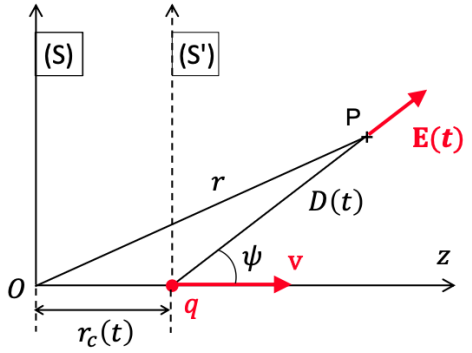
$$\mathbf{E}' = \frac{q\mathbf{r}'}{4\pi\epsilon_0 r'^3}, \quad \mathbf{B}' = \mathbf{0}, \tag{6.1}$$

where primed quantities denote the values in (S′). Using equations (5.53) and (5.54), the electromagnetic field in (S), which moves with velocity $-\mathbf{v}$ with respect to (S′), can be expressed as:

$$\mathbf{E}_\parallel = \mathbf{E}'_\parallel, \qquad \mathbf{E}_\perp = \gamma\mathbf{E}'_\perp, \qquad \mathbf{B} = \gamma\frac{\mathbf{v}\times\mathbf{E}'}{c^2}, \tag{6.2}$$

where the subscripts $\parallel$ and $\perp$ indicate the components parallel and perpendicular to the velocity $\mathbf{v}$, respectively. Aligning the z-axis with $\mathbf{v}$, the field in frame (S) is evaluated at $\mathbf{r} = (x, y, z)$, which is related to the coordinates in frame (S′) by:

$$z' = \gamma\left(z - vt\right), \qquad x' = x, \qquad y' = y.$$



In frame (S), the position of the particle at time t is represented as $\mathbf{r}_c(t) = \mathbf{v}t = vt\hat{\mathbf{z}}$, with the origin chosen such that $z = 0$ corresponds to the initial position. The relative position of point P with respect to the particle at time t is given by $\mathbf{D}(t) = \mathbf{r} - \mathbf{r}_c(t)$, and ψ denotes the angle between $\mathbf{D}(t)$ and the z-axis.

We find:

$$r'^2 = \gamma^2\left(z - vt\right)^2 + x^2 + y^2 = \gamma^2 D^2\cos^2\psi + D^2\sin^2\psi = \gamma^2 D^2\left(1 - \frac{v^2}{c^2}\sin^2\psi\right). \tag{6.3}$$

Here, r' is the distance from P to the charge in (S′), while D is the corresponding distance in (S). The difference between these two quantities arises from length contraction when transforming from the rest frame (S′) to (S).

Equations (6.2) can be expressed as $\mathbf{E} \equiv E_x\hat{\mathbf{x}} + E_y\hat{\mathbf{y}} + E_z\hat{\mathbf{z}} = \gamma E'_x\hat{\mathbf{x}} + \gamma E'_y\hat{\mathbf{y}} + E'_z\hat{\mathbf{z}}$, resulting in:

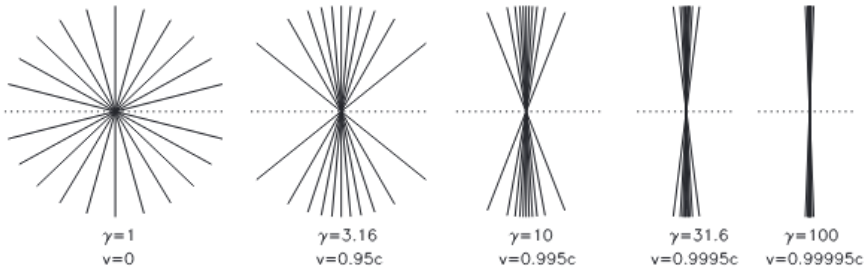
$$\mathbf{E} = \frac{q}{4\pi\epsilon_0 r'^3}\left(\gamma x'\hat{\mathbf{x}} + \gamma y'\hat{\mathbf{y}} + z'\hat{\mathbf{z}}\right) = \frac{q}{4\pi\epsilon_0 r'^3}\left(\gamma x\hat{\mathbf{x}} + \gamma y\hat{\mathbf{y}} + \gamma\left(z - vt\right)\hat{\mathbf{z}}\right) = \frac{q\gamma\mathbf{D}}{4\pi\epsilon_0 r'^3}. \tag{6.4}$$

Substituting the expression for r' from equation (6.3) into equation (6.4) gives:

$$\boxed{\mathbf{E}(\mathbf{r},t) = \frac{1}{\gamma^2\left[1-(v^2/c^2)\sin^2\psi\right]^{3/2}}\frac{q\mathbf{D}}{4\pi\epsilon_0 D^3}.} \tag{6.5}$$

The electric field points away from the instantaneous position of the charge. This is non-intuitive because the signal received at P was emitted when the particle was at a different (*retarded*) position. We will see in section 6.1.2 and the following sections that this is a special property of uniform motion, which no longer holds for an accelerated charge. In the limit $v/c \ll 1$, this reduces to the familiar field of a stationary point charge. As v/c approaches 1, the field is enhanced by a factor γ in the directions perpendicular to the velocity ($\psi = \pi/2$) and reduced by a factor $1/\gamma^2$ along the velocity ($\psi = 0$ or π), so that the field lines become concentrated in the plane perpendicular to the motion, as illustrated below.

Electric field lines for different values of γ (*A. K. Singal, 2020, J. Phys. Commun., 4*):



For the magnetic field, equations (6.2) yield:

$$\boxed{\mathbf{B} = \frac{\mathbf{v}\times\mathbf{E}}{c^2} = \frac{1}{\gamma^2\left[1-(v^2/c^2)\sin^2\psi\right]^{3/2}}\frac{\mu_0 q\mathbf{v}\times\mathbf{D}}{4\pi D^3}.} \tag{6.6}$$

In the limit $v/c \ll 1$, this simplifies to the Biot-Savart's law, where the current is replaced by $q\mathbf{v}$.

6.1.2 Retarded potentials

The electric and magnetic fields given by equations (6.5) and (6.6) are solutions to Maxwell's equations for charge and current densities given by:

$$\rho(\mathbf{r},t) = q\delta(\mathbf{r}-\mathbf{r}_c(t)), \qquad \mathbf{j}(\mathbf{r},t) = q\mathbf{v}(t)\delta(\mathbf{r}-\mathbf{r}_c(t)), \tag{6.7}$$

where the specific case of uniform velocity, $\mathbf{v}(t) = \mathrm{d}\mathbf{r}_c/\mathrm{d}t = \mathbf{const}$, was examined.

We now consider more general charge and current densities, and seek solutions to Maxwell's equations in the form of equation (5.31), which is restated here for convenience:

$$\Box A^{\mu} \equiv -\frac{1}{c^2}\frac{\partial^2 A^{\mu}}{\partial t^2} + \nabla^2 A^{\mu} = -\mu_0 J^{\mu}, \tag{6.8}$$

where the 4-potential is $A^{\mu} = (\phi/c, \mathbf{A})$, the 4-current is $J^{\mu} = (\rho c, \mathbf{j})$ and the Lorenz gauge is assumed. All functions depend on $\mathbf{r}$ and t.

We introduce the Green function G, defined by the equation:

$$\Box G\left(\mathbf{r}, \mathbf{r}_c, t, t_c\right) = -\delta(t - t_c)\delta\left(\mathbf{r} - \mathbf{r}_c\right). \tag{6.9}$$

This equation identifies G as the potential generated by a point charge that exists only instantaneously at $\mathbf{r} = \mathbf{r}_c$ at time $t = t_c$, effectively 'flashed' at that moment. This implies $G\left(\mathbf{r}, \mathbf{r}_c, t, t_c\right) = G\left(\mathbf{r} - \mathbf{r}_c, t - t_c\right)$. Since equation (6.8) is linear, its solutions can be constructed by summing over all positions and times:

$$A^{\mu}\left(\mathbf{r}, t\right) = \int_{-\infty}^{+\infty} \mathrm{d}t_c \int_{-\infty}^{+\infty} \mathrm{d}^3\mathbf{r}_c \; \mu_0 J^{\mu}\left(\mathbf{r}_c, t_c\right) G\left(\mathbf{r} - \mathbf{r}_c, t - t_c\right). \tag{6.10}$$

Solving for G thus determines A^{μ}. We next introduce the Fourier transform of G, denoted as $\tilde{G}$, which satisfies:

$$G\left(\mathbf{R}, \theta\right) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \tilde{G}\left(\mathbf{R}, \omega\right) \mathrm{e}^{-\mathrm{i}\omega\theta} \, \mathrm{d}\omega, \tag{6.11}$$

where $\mathbf{R} \equiv \mathbf{r} - \mathbf{r}_c$ and $\theta \equiv t - t_c$. Substituting this expression into equation (6.9), and noting that $\boldsymbol{\nabla}_{\mathbf{r}} = \boldsymbol{\nabla}_{\mathbf{R}}$ and $\partial_t = \partial_\theta$ (since $\mathbf{r}_c$ and t_c are fixed) results in:

$$\int_{-\infty}^{+\infty} \left(k^2 + \nabla^2\right) \tilde{G}\left(\mathbf{R}, \omega\right) \mathrm{e}^{-\mathrm{i}\omega\theta} \, \mathrm{d}\omega = -\int_{-\infty}^{+\infty} \delta\left(\mathbf{R}\right) \mathrm{e}^{-\mathrm{i}\omega\theta} \, \mathrm{d}\omega, \tag{6.12}$$

where we used $\int_{-\infty}^{+\infty} \mathrm{e}^{-\mathrm{i}\omega\theta} \, \mathrm{d}\omega = 2\pi\delta(\theta)$ and defined $k = \omega/c$. Since this holds for all t, and thus for all θ, it follows[1]:

$$\left(k^2 + \nabla^2\right) \tilde{G}\left(\mathbf{R}, \omega\right) = -\delta\left(\mathbf{R}\right). \tag{6.13}$$

[1]This relies on the fact that, if $\int_{-\infty}^{+\infty} f\left(\mathbf{r}, \omega\right) \mathrm{e}^{-\mathrm{i}\omega\theta} \, \mathrm{d}\omega = 0$ for all θ, then $f\left(\mathbf{r}, \omega\right) = 0$ for all ω. To see this, multiply the integral by $\mathrm{e}^{\mathrm{i}\omega_0\theta}$, where ω_0 is arbitrary, and integrate over θ:

$$\int_{-\infty}^{+\infty} \mathrm{d}\omega \; f\left(\mathbf{r}, \omega\right) \int_{-\infty}^{+\infty} \mathrm{d}\theta \; \mathrm{e}^{-\mathrm{i}(\omega - \omega_0)\theta} = \int_{-\infty}^{+\infty} \mathrm{d}\omega \; 2\pi f\left(\mathbf{r}, \omega\right) \delta\left(\omega - \omega_0\right) = 2\pi f\left(\mathbf{r}, \omega_0\right) = 0.$$

Since G represents the potential created by a charge that exists momentarily at $\mathbf{R} = \mathbf{0}$ at a specific instant, it depends only on $R = |\mathbf{r} - \mathbf{r}_c|$, not on the direction of the vector. Using spherical coordinates, equation (6.13) can thus be rewritten as:

$$k^2\tilde{G} + \frac{1}{R^2}\frac{\mathrm{d}}{\mathrm{d}R}\left(R^2\frac{\mathrm{d}\tilde{\mathrm{G}}}{\mathrm{dR}}\right) = k^2\tilde{G} + \frac{1}{R}\frac{\mathrm{d}^2}{\mathrm{d}R^2}\left(R\tilde{G}\right) = -\delta\left(\mathbf{R}\right). \tag{6.14}$$

Away from $R = 0$, the solution is $R\tilde{G} = C\mathrm{e}^{\pm \mathrm{i}kR}$, where C is a constant.

The constant C is determined by the behaviour of $\tilde{G}$ near $R = 0$, where the delta function acts. There, $\mathrm{e}^{\pm \mathrm{i}kR} \simeq 1$ and the term $k^2\tilde{G}$ is negligible compared with the derivative terms, so that $\tilde{G}$ must behave as the solution of $\nabla^2\tilde{G} = -\delta\left(\mathbf{R}\right)$. This is Poisson's equation $\nabla^2\tilde{G} = -\rho/\epsilon_0$ with $\rho = \epsilon_0\delta\left(\mathbf{R}\right)$, which describes the static potential created by a point charge of magnitude ϵ_0 located at $R = 0$, whose solution is $\tilde{G} = 1/(4\pi R)$. Matching this with $C\mathrm{e}^{\pm \mathrm{i}kR}/R$ near $R = 0$ gives $C = 1/(4\pi)$, for all values of k.

Therefore, the solution becomes:

$$\tilde{G}(R, \omega) = \frac{\mathrm{e}^{\pm \mathrm{i}kR}}{4\pi R}. \tag{6.15}$$

Recovering the Green function G involves the inverse Fourier transform:

$$G(R, \theta) = \frac{1}{2\pi}\int_{-\infty}^{+\infty}\tilde{G}(R, \omega)\mathrm{e}^{-\mathrm{i}\omega\theta}\,\mathrm{d}\omega = \frac{1}{8\pi^2 R}\int_{-\infty}^{+\infty}\mathrm{e}^{-\mathrm{i}\omega(\theta \mp R/c)}\,\mathrm{d}\omega = \frac{\delta\left(\theta \mp R/c\right)}{4\pi R}. \tag{6.16}$$

This shows that the Green function consists of two mathematical branches:

$$G(R, t - t_c) = \frac{1}{4\pi R}\delta(t - t_c \mp R/c),$$

corresponding to a signal propagating from the source point $\mathbf{r}_c$ to the field point $\mathbf{r}$ in either temporal direction. The solution with the "+" sign would imply that the field at time t depends on the value of the source at a later time $t_c = t + R/c$, i.e. the field would be generated before the source acts. This violates causality. Therefore, the only physically meaningful choice is the "–" sign, which states that the disturbance produced at the source at time t_c reaches the observation point at time $t = t_c + R/c$. We thus select the retarded Green function.

Using this result, equation (6.10) becomes:

$$A^\mu\left(\mathbf{r}, t\right) = \int_{-\infty}^{+\infty}\mathrm{d}t_c\int_{-\infty}^{+\infty}\mathrm{d}^3\mathbf{r}_c\;\mu_0 J^\mu\left(\mathbf{r}_c, t_c\right)\frac{\delta\left(t - t_c - R/c\right)}{4\pi R}, \tag{6.17}$$

which simplifies to:

$$\boxed{A^{\mu}\left(\mathbf{r},t\right)=\frac{\mu_0}{4\pi}\int_{-\infty}^{+\infty}\mathrm{d}^3\mathbf{r}_c\,\frac{J^{\mu}\left(\mathbf{r}_c,t-R/c\right)}{R},} \tag{6.18}$$

where $R=|\mathbf{r}-\mathbf{r}_c|$ and the integral runs over all source points $\mathbf{r}_c$. This expression shows that the potential at point $\mathbf{r}$ and time t is determined by the current density at each source point $\mathbf{r}_c$ at the earlier time $t-|\mathbf{r}-\mathbf{r}_c|/c$, where $|\mathbf{r}-\mathbf{r}_c|/c$ is the time required for the field to propagate from $\mathbf{r}_c$ to $\mathbf{r}$. For a point charge, as we will see in the next section, $\mathbf{r}_c$ becomes the position of the charge at the retarded time t_c, defined implicitly by $t_c=t-|\mathbf{r}-\mathbf{r}_c(t_c)|/c$. For this reason, the potential given in equation (6.18) is referred to as the *retarded potential*.

It is important to note that the retarded potentials depend on the position and velocity of the charge at the retarded time, through the charge and current densities, but not explicitly on its acceleration. Consequently, **the potentials at time t for a charge moving at uniform velocity and for an accelerated charge will be identical if they have the same position and the same velocity at the relevant retarded time**.

However, this equivalence does not hold for the electromagnetic fields. When the potentials are differentiated to obtain $\mathbf{E}$ and $\mathbf{B}$, time derivatives of the velocity appear, and the fields acquire explicit dependence on the acceleration of the charge, as will be shown below.

6.1.3 Liénard-Wiechert potential

The potential derived above is general and applies to any charge and current distributions. We now focus on the Liénard–Wiechert potential, which is the potential of a single accelerating point charge. This can be obtained by substituting J^{μ} into equation (6.18) for this specific case. However, it is more straightforward to revisit the situation of a charge moving with uniform velocity and use the previously established fact that the retarded potential in frame (S) is independent of whether the charge is accelerated or moving uniformly.

The case of a charge moving with uniform velocity $\mathbf{v}$ in frame (S) was discussed in section 6.1.1, where the electric and magnetic fields were calculated directly. Here, we focus on the potentials instead, as these are the quantities that remain the same for both uniformly moving and accelerated charges. In the rest frame (S′) of this charge, where it is at the (fixed) location $\mathbf{r}'_c$, the scalar and vector potentials are given by:

$$\phi'\left(\mathbf{r}'\right)=\frac{q}{4\pi\epsilon_0\left|\mathbf{r}'-\mathbf{r}'_c\right|},\qquad \mathbf{A}'\left(\mathbf{r}'\right)=\mathbf{0}. \tag{6.19}$$

Note that *these are not the potentials in the instantaneous rest frame if the charge is accelerating*, as we will discuss later.

Since A^μ is a 4-vector, the transformation to frame (S) gives:

$$\frac{\phi}{c}(\mathbf{r},t) = \gamma\frac{\phi'}{c} = \frac{\gamma q}{4\pi\epsilon_0 c\,|\mathbf{r}'-\mathbf{r}'_c|}, \tag{6.20}$$

$$\mathbf{A}_\parallel(\mathbf{r},t) = \gamma\frac{\mathbf{v}}{c}\frac{\phi'}{c} = \frac{\gamma q\mathbf{v}}{4\pi\epsilon_0 c^2\,|\mathbf{r}'-\mathbf{r}'_c|}, \tag{6.21}$$

$$\mathbf{A}_\perp(\mathbf{r},t) = \mathbf{0}, \tag{6.22}$$

where $\mathbf{A}_\parallel$ and $\mathbf{A}_\perp$ are the components of $\mathbf{A}$ parallel and perpendicular to $\mathbf{v}$, respectively.

To express $|\mathbf{r}'-\mathbf{r}'_c|$ in terms of the coordinates $\mathbf{r}$ and t in frame (S), consider the following two events in frame (S′):

- Event 1: the charge is at $\mathbf{r}'_c$ (a fixed location) and emits a field (or photon) at time t'_c,
- Event 2: the observer is at $\mathbf{r}'$ and receives the field (or photon) at time $t' = t'_c + |\mathbf{r}'-\mathbf{r}'_c|/c$.

The displacement 4-vector between these two events in (S′) is $R'^\mu = [c(t'-t'_c), \mathbf{r}'-\mathbf{r}'_c]$, and since the interval is light-like we have $(R'^\mu)^2 = 0$.

In frame (S), the corresponding events are as follows:

- Event 1: the charge is at $\mathbf{r}_c$ and emits a field (or photon) at time t_c,
- Event 2: the observer is at $\mathbf{r}$ (a fixed location) and receives the field (or photon) at time $t = t_c + |\mathbf{r}-\mathbf{r}_c(t_c)|/c$, noting that the charge has moved from $\mathbf{r}_c$ by the time the observer receives the field.

The displacement 4-vector between these two events in (S) is $R^\mu = [c(t-t_c), \mathbf{r}-\mathbf{r}_c(t_c)]$ and, again, since the interval is light-like, we have $(R^\mu)^2 = 0$.

To calculate $|\mathbf{r}'-\mathbf{r}'_c|$, we instead calculate $c(t'-t'_c)$, as these two expressions are equal. Transforming from frame (S) to frame (S′), we obtain:

$$c\left(t'-t'_c\right) = \gamma\left[c(t-t_c) - \frac{\mathbf{v}}{c}\cdot(\mathbf{r}-\mathbf{r}_c(t_c))\right]. \tag{6.23}$$

With $U^\mu = (\gamma c, \gamma\mathbf{v})$ representing the 4-velocity of the charge in frame (S), the right-hand side of the expression above can be identified as $-U^\mu R_\mu/c$. Substituting this result into equations (6.20), (6.21) and (6.22) gives:

$$\frac{\phi}{c}(\mathbf{r},t) = -\frac{1}{4\pi\epsilon_0 c}\frac{\gamma c q}{U^\mu R_\mu}, \tag{6.24}$$

$$\mathbf{A}(\mathbf{r},t) = -\frac{1}{4\pi\epsilon_0 c}\frac{\gamma q\mathbf{v}}{U^\mu R_\mu}. \tag{6.25}$$

This can be expressed in covariant form as:

$$\boxed{A^{\nu}(\mathbf{r},t) = -\frac{1}{4\pi\epsilon_0 c}\frac{qU^{\nu}}{U^{\mu}R_{\mu}},} \tag{6.26}$$

where $\mathbf{r}_c$ in R_μ must be evaluated at the retarded time $t_c = t - |\mathbf{r} - \mathbf{r}_c(t_c)|/c$.

We now rewrite the potentials in a more explicit form in terms of the coordinates. Defining $\mathbf{R} \equiv \mathbf{r} - \mathbf{r}_c(t_c)$, as introduced in section 6.1.2, equation (6.23) becomes:

$$c\left(t' - t'_c\right) = \gamma\left(1 - \frac{\mathbf{v}}{c}\cdot\widehat{\mathbf{R}}\right)R, \tag{6.27}$$

where we have used $c(t - t_c) = R$ and $\widehat{\mathbf{R}}$ is a unit vector in the direction of $\mathbf{r} - \mathbf{r}_c(t_c)$. The left-hand side, $c(t' - t'_c)$, is equal to $|\mathbf{r}' - \mathbf{r}'_c|$. Substituting this into equations (6.20), (6.21) and (6.22), we obtain:

$$\boxed{\phi(\mathbf{r},t) = \frac{q}{4\pi\epsilon_0}\frac{1}{\left(1 - \boldsymbol{\beta}\cdot\widehat{\mathbf{R}}\right)R}, \quad \mathbf{A}(\mathbf{r},t) = \frac{q}{4\pi\epsilon_0 c}\frac{\boldsymbol{\beta}}{\left(1 - \boldsymbol{\beta}\cdot\widehat{\mathbf{R}}\right)R},} \tag{6.28}$$

where $\boldsymbol{\beta} \equiv \mathbf{v}/c$ and $R = |\mathbf{r} - \mathbf{r}_c(t_c)|$, with $t_c = t - |\mathbf{r} - \mathbf{r}_c(t_c)|/c$.

The Liénard-Wiechert potentials (6.26) and (6.28) were derived for a charge with uniform velocity $\mathbf{v}$. However, as noted earlier, they remain valid for an accelerated charge that has the velocity $\mathbf{v}$ at time t_c, even though the potentials given by equation (6.19), which we used as the starting point for the calculation and represent the potentials in the rest frame of a charge moving with uniform velocity, are not equal to the potentials of the accelerated charge in its instantaneous rest frame, as will be seen in the next section.

6.1.4 Field due to an accelerated charge

The electric and magnetic fields can be determined from the potentials in equations (6.28) using the relations:

$$\mathbf{E} = -\boldsymbol{\nabla}\phi - \frac{\partial\mathbf{A}}{\partial t}, \quad \mathbf{B} = \boldsymbol{\nabla}\times\mathbf{A}, \tag{6.29}$$

where the derivatives are with respect to the coordinates $\mathbf{r}$ and t in frame (S). Since t_c depends on both $\mathbf{r}$ and t, the calculation is quite laborious. However, having already derived the fields for a uniformly moving charge in section 6.1.1, we now calculate only the terms involving the acceleration. The total fields are obtained by adding the two contributions. The component of the electric field that depends on acceleration is denoted $\mathbf{E}_{\text{rad}}$, known as the **radiation field**, as it contributes to the radiated power, as will be shown in the next section.

Using the potentials derived earlier, we have:

$$\begin{aligned} E_{\text{rad},i} &= \frac{q}{4\pi\epsilon_0 R}\left[-\frac{\widehat{\mathbf{R}}\cdot\partial_i\boldsymbol{\beta}}{\left(1-\boldsymbol{\beta}\cdot\widehat{\mathbf{R}}\right)^2}-\frac{\partial_t\beta_i}{\left(1-\boldsymbol{\beta}\cdot\widehat{\mathbf{R}}\right)c}-\frac{\beta_i\left(\widehat{\mathbf{R}}\cdot\partial_t\boldsymbol{\beta}\right)}{\left(1-\boldsymbol{\beta}\cdot\widehat{\mathbf{R}}\right)^2 c}\right] \\ &= \frac{-q\left[c\widehat{\mathbf{R}}\cdot\partial_i\boldsymbol{\beta}+\partial_t\beta_i-\partial_t\beta_i\left(\boldsymbol{\beta}\cdot\widehat{\mathbf{R}}\right)+\beta_i\left(\widehat{\mathbf{R}}\cdot\partial_t\boldsymbol{\beta}\right)\right]}{4\pi\epsilon_0\left(1-\boldsymbol{\beta}\cdot\widehat{\mathbf{R}}\right)^2 cR} \end{aligned} \tag{6.30}$$

The position of the charge at time t_c is $\mathbf{r}_c$, with velocity and acceleration given by:

$$\mathbf{v}(t_c)=\boldsymbol{\beta}c=\frac{\mathrm{d}\mathbf{r}_c}{\mathrm{d}t_c},\qquad \mathbf{a}(t_c)=\frac{\mathrm{d}^2\mathbf{r}_c}{\mathrm{d}t_c^2}.$$

Defining $\dot{\boldsymbol{\beta}}=\mathrm{d}\boldsymbol{\beta}/\mathrm{d}t_c$, the derivatives of $\boldsymbol{\beta}$ are:

$$\partial_t\boldsymbol{\beta}=\dot{\boldsymbol{\beta}}\partial_t t_c \quad\text{and}\quad \partial_i\boldsymbol{\beta}=\dot{\boldsymbol{\beta}}\partial_i t_c.$$

Writing:

$$t_c=t-\frac{R}{c}=t-\frac{|\mathbf{r}-\mathbf{r}_c(t_c)|}{c}=t-\frac{1}{c}\left[(x-x_c)^2+(y-y_c)^2+(z-z_c)^2\right]^{1/2}, \tag{6.31}$$

with $\mathbf{r}=(x,y,z)$ and $\mathbf{r}_c=(x_c,y_c,z_c)$, we calculate:

$$\partial_i t_c=-\frac{R_i}{Rc}+\frac{\boldsymbol{\beta}\cdot\mathbf{R}}{R}\partial_i t_c,\quad \text{which yields}\quad \partial_i t_c=\frac{-\widehat{R}_i}{\left(1-\boldsymbol{\beta}\cdot\widehat{\mathbf{R}}\right)c}. \tag{6.32}$$

Similarly,

$$\partial_t t_c=1+\frac{\boldsymbol{\beta}\cdot\mathbf{R}}{R}\partial_t t_c,\quad \text{which yields}\quad \partial_t t_c=\frac{1}{1-\boldsymbol{\beta}\cdot\widehat{\mathbf{R}}}. \tag{6.33}$$

Substituting these into equation (6.30) gives:

$$\mathbf{E}_{\text{rad}}(\mathbf{r},t)=\frac{q\widehat{\mathbf{R}}\times\left[\left(\widehat{\mathbf{R}}-\boldsymbol{\beta}\right)\times\dot{\boldsymbol{\beta}}\right]}{4\pi\epsilon_0\left(1-\boldsymbol{\beta}\cdot\widehat{\mathbf{R}}\right)^3 cR}. \tag{6.34}$$

To this radiation field, we add the component depending only on velocity, denoted $\mathbf{E}_{\text{nf}}$. Since it does not depend on the acceleration, it is the field of a fictitious charge moving uniformly with velocity $\mathbf{v}\equiv\mathbf{v}(t_c)$, which passes through $\mathbf{r}_c(t_c)$ at time t_c, and it is therefore given by equation (6.5). This **near-zone field** varies as $1/R^2$ and dominates over the radiation field close to the charge. At time t, the fictitious charge is at $\mathbf{r}_c(t_c)+(t-t_c)\,\mathbf{v}=\mathbf{r}_c(t_c)+\mathbf{v}R/c$, so that the vector $\mathbf{D}$ in equation (6.5) is $\mathbf{D}=\mathbf{R}-\boldsymbol{\beta}R=R\left(\widehat{\mathbf{R}}-\boldsymbol{\beta}\right)$. Since ψ is the angle between $\mathbf{D}$ and $\boldsymbol{\beta}$, we also have:

$$\beta^2D^2\sin^2\psi=|\boldsymbol{\beta}\times\mathbf{D}|^2=R^2\left|\boldsymbol{\beta}\times\widehat{\mathbf{R}}\right|^2=R^2\left[\beta^2-\left(\boldsymbol{\beta}\cdot\widehat{\mathbf{R}}\right)^2\right],$$

and $D^2=R^2\left(1-2\boldsymbol{\beta}\cdot\widehat{\mathbf{R}}+\beta^2\right)$, so that $D^2\left(1-\beta^2\sin^2\psi\right)=R^2\left(1-\boldsymbol{\beta}\cdot\widehat{\mathbf{R}}\right)^2$.

Substituting these expressions into equation (6.5) yields the total electric field $\mathbf{E} = \mathbf{E}_{\text{rad}} + \mathbf{E}_{\text{nf}}$:

$$\boxed{\mathbf{E}(\mathbf{r},t) = \frac{q}{4\pi\epsilon_0\left(1-\boldsymbol{\beta}\cdot\widehat{\mathbf{R}}\right)^3}\left[\frac{\widehat{\mathbf{R}}-\boldsymbol{\beta}}{\gamma^2R^2}+\frac{\widehat{\mathbf{R}}\times\left[\left(\widehat{\mathbf{R}}-\boldsymbol{\beta}\right)\times\dot{\boldsymbol{\beta}}\right]}{cR}\right].} \tag{6.35}$$

Here, $\boldsymbol{\beta} \equiv \mathbf{v}(t_c)/c$, $\dot{\boldsymbol{\beta}} = \mathrm{d}\boldsymbol{\beta}/\mathrm{d}t_c$ and $R = |\mathbf{r} - \mathbf{r}_c(t_c)|$, with t_c defined implicitly by $t_c = t - |\mathbf{r} - \mathbf{r}_c(t_c)|/c$. Note that, contrary to the field produced by a charge with a uniform velocity, the field produced by an accelerated charge does not, in general, point away from the instantaneous position of the charge. The velocity term is directed along $\widehat{\mathbf{R}} - \boldsymbol{\beta}$, that is to say away from the position $\mathbf{r}_c(t_c) + \mathbf{v}R/c$ that the charge would occupy at time t if it had continued with the velocity it had at the retarded time, while the radiation term is perpendicular to $\widehat{\mathbf{R}}$.

Similarly, the magnetic field can be computed from the vector potential $\mathbf{A}$, which yields:

$$\boxed{\mathbf{B}(\mathbf{r},t) = \frac{\widehat{\mathbf{R}}\times\mathbf{E}}{c}.} \tag{6.36}$$

In the instantaneous rest frame of the particle, where $\boldsymbol{\beta} = \mathbf{0}$ at the observation time, the electric field does not reduce to Coulomb's law. The reason is that the retarded fields depend on the velocity and acceleration at the retarded time, not on the instantaneous state at the observation time. Thus, **the fields in the instantaneous rest frame are not the same as the Coulomb fields of a uniformly moving charge in its own rest frame**. Even in the frame in which the charge is at rest at the retarded time ($\boldsymbol{\beta} = \mathbf{0}$), equation (6.35) gives the Coulomb field of the charge at its retarded position plus the radiation term $q\,\widehat{\mathbf{R}}\times\left(\widehat{\mathbf{R}}\times\dot{\boldsymbol{\beta}}\right)/(4\pi\epsilon_0 cR)$, which will be the starting point of section 6.2.1.

6.2 Radiated power

In section 6.1.1, it was demonstrated that the electric and magnetic fields of a point charge moving with constant velocity fall off as $1/R^2$, implying that the Poynting vector scales as $1/R^4$. Consequently, no energy is transported to infinity, indicating that a charge moving at constant velocity does not emit radiation.

In contrast, equations (6.35) and (6.36) show that when a charge is accelerating, both $\mathbf{E}$ and $\mathbf{B}$ include terms proportional to $1/R$. This implies that an accelerated charge emits radiation. We now proceed to calculate the radiated power.

6.2.1 Radiation from a non-relativistic accelerated point charge

We start by examining the case where the velocity of the charge is much smaller than the speed of light.

Larmor's formula

As mentioned earlier, at distances far from the charge, the radiation field dominates over the near-zone field contributions. Specifically, from equation (6.35), this condition is satisfied when $R \gg c/\left(\gamma^2\dot{\beta}\right)$. This regime is the focus of our discussion here.

Retaining only the radiation field and considering the limit $\beta \ll 1$, equation (6.34) simplifies to:

$$\mathbf{E}_{\text{rad}}(\mathbf{r},t) = \frac{q}{4\pi\epsilon_0 c}\frac{\widehat{\mathbf{R}}\times\left(\widehat{\mathbf{R}}\times\dot{\boldsymbol{\beta}}\right)}{R}. \tag{6.37}$$

Using spherical coordinates (R,θ,φ), with the z-axis aligned along the acceleration of the particle at the retarded time t_c, and denoting unit vectors with a hat, this can be rewritten as:

$$\mathbf{E}_{\text{rad}}(\mathbf{r},t) = \frac{qa\sin\theta}{4\pi\epsilon_0 c^2 R}\hat{\boldsymbol{\theta}}, \tag{6.38}$$

where $a = c\dot{\beta} = \dot{v}$ is the acceleration. The corresponding magnetic radiation field, given by $\mathbf{B}_{\text{rad}} = \widehat{\mathbf{R}}\times\mathbf{E}_{\text{rad}}/c$ (eq. [6.36]) becomes:

$$\mathbf{B}_{\text{rad}}(\mathbf{r},t) = \frac{qa\sin\theta}{4\pi\epsilon_0 c^3 R}\hat{\boldsymbol{\varphi}}, \tag{6.39}$$

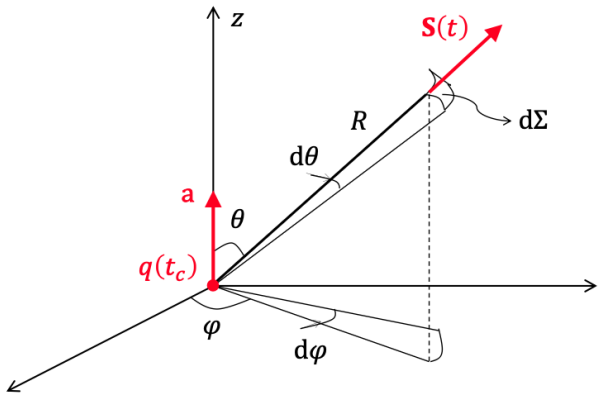
yielding the Poynting vector:

$$\mathbf{S}(\mathbf{r},t) = \frac{1}{\mu_0}\mathbf{E}_{\text{rad}}\times\mathbf{B}_{\text{rad}} = \frac{1}{\mu_0 c}E_{\text{rad}}^2\widehat{\mathbf{R}} = \frac{q^2a^2\sin^2\theta}{16\pi^2\epsilon_0 c^3 R^2}\widehat{\mathbf{R}}. \tag{6.40}$$

Here, $R = |\mathbf{r} - \mathbf{r}_c(t_c)|$, where $\mathbf{r}_c(t_c)$ is the position of the charge at the retarded time t_c and $t_c = t - |\mathbf{r} - \mathbf{r}_c(t_c)|/c$.

The radiation field can also be derived by analysing the motion of the charge and determining how far the information about its velocity, influenced by the acceleration, propagates within a finite amount of time. As this approach offers additional insight into the structure of the radiation field, we have included it in Appendix A.

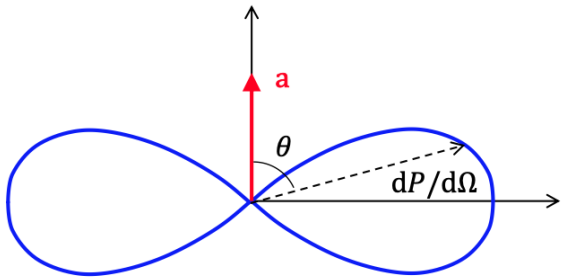
We now apply energy conservation to relate the flux of the Poynting vector to the energy emitted by the charge.



The diagram shows a particle located at the origin at time $t_c = t - R/c$, with its acceleration vector $\mathbf{a}$ oriented vertically. We consider an infinitesimal area $d\Sigma$ on the surface of a sphere of radius R, which subtends a solid angle $\mathrm{d}\Omega = \mathrm{d}\Sigma/R^2$.

Now, consider an observer positioned at the surface element $d\Sigma$. Energy conservation, as expressed by equation (5.104) and integrated over the spherical cone, gives $\mathbf{S} \cdot \widehat{\mathbf{R}} d\Sigma = \mathrm{d}U/\mathrm{d}t$, where $\mathrm{d}U$ represents the energy emitted *into* the spherical cone during the time interval $\mathrm{d}t$, as measured by the observer. For a non-relativistic charge, equation (6.33) yields $\partial_t t_c \simeq 1$, which implies that $\mathrm{d}t$ is approximately equal to the time interval $\mathrm{d}t_c$ during which the radiation is emitted. This can also be seen by noting that $r_c(t_c) \simeq r_c(t)$ for non-relativistic motion, and therefore $t_c \simeq t - |\mathbf{r} - \mathbf{r}_c(t)|/c$, which leads to $\mathrm{d}t_c \simeq \mathrm{d}t$. Thus, the power received by the observer in the infinitesimal solid angle $\mathrm{d}\Omega$, given by $\mathrm{d}P = \mathrm{d}U/\mathrm{d}t$, is equal to the power radiated by the charge into that solid angle. This is expressed as:

$$\boxed{\mathrm{d}P = \frac{q^2 a^2 \sin^2\theta}{4\pi\epsilon_0 c^3} \frac{\mathrm{d}\Omega}{4\pi}.} \tag{6.41}$$



The figure illustrates the radiation pattern, represented as a polar plot of (r, θ), where $r = dP/d\Omega \propto \sin^2\theta$. Note that **no energy is radiated along the direction of acceleration**.

To calculate the total power P_{rad} radiated through the sphere of radius R, we integrate $\mathrm{d}P$ over the surface area $\mathrm{d}\Sigma = R^2 \sin\theta \mathrm{d}\theta \mathrm{d}\varphi$, or equivalently over the solid angle $\mathrm{d}\Omega = \sin\theta \mathrm{d}\theta \mathrm{d}\varphi$:

$$P_{\text{rad}} = \frac{q^2 a^2}{8\pi\epsilon_0 c^3} \int_0^\pi \sin^3\theta d\theta.$$

This leads to **Larmor's formula**:

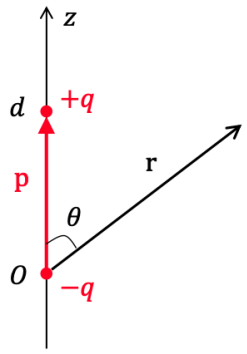
$$\boxed{P_{\text{rad}} = \frac{q^2 a^2}{6\pi\epsilon_0 c^3},} \tag{6.42}$$

which applies to a non-relativistic accelerated charge. This formula demonstrates that **any accelerated charged particle radiates power proportional to the square of its acceleration**.

Notably, P_{rad} is independent of R: it depends only on the acceleration of the charge at the retarded time t_c. This means that the power crossing a spherical surface of radius R at time t is equal to the power crossing another spherical surface of radius R' at time $t + (R' - R)/c$, since both correspond to the radiation emitted at the same retarded time. This result reflects the principle of energy conservation.

Before extending this result to relativistic charges, we first apply it to the cases of an oscillating dipole and an antenna.

Radiation from an oscillating dipole



Consider a physical dipole aligned along the z-axis, where the negative charge $-q$ is fixed at the origin, and the positive charge $+q$ oscillates along the axis. The position of the oscillating charge is described by $z(t) = d\sin\omega t$, giving the dipole moment as $\mathbf{p}(t) = qz(t)\ \hat{\mathbf{z}} = p_0 \sin\omega t\ \hat{\mathbf{z}}$, where $p_0 \equiv qd$. The total field is the sum of the fields produced by both charges, but only the positive, accelerated charge contributes to the radiation field.

The radiated power is therefore given by equation (6.42), with the acceleration $a = -d\omega^2 \sin\omega t$. Averaging over an oscillation period, this results in:

$$\langle P_{\text{rad}} \rangle = \frac{\omega^4 p_0^2}{12\pi\epsilon_0 c^3}. \tag{6.43}$$

Here, it is assumed that the motion is non-relativistic, meaning $v \sim d\omega \ll c$, which implies $d \ll \lambda$, where $\lambda = 2\pi c/\omega$ is the wavelength of the electric field, which oscillates at the same frequency as the charge's motion.

The condition $R \gg c^2/a$ derived earlier compares the radiation field with the near-zone field of a single charge, of order $q/(4\pi\epsilon_0 R^2)$. For a dipole, however, the near-zone fields of the two charges largely cancel at distances $R \gg d$, leaving the static dipole field, of order $p_0/(4\pi\epsilon_0 R^3)$. The radiation field, of order $\omega^2 p_0/(4\pi\epsilon_0 c^2 R)$, therefore dominates when $R \gg c/\omega \sim \lambda$. With $d \ll \lambda$, these conditions can be summarised as $\boldsymbol{R \gg \lambda \gg d}$. When $R \gg d$, we have $R \simeq r$. An oscillating dipole observed from distances much greater than its dimensions is referred to as a *Hertzian dipole*.

In the calculation above, we have computed only the electric dipole contribution to the radiation, arising from $\ddot{\mathbf{p}}$. This is the first term of the multipole expansion of the radiation field. The next terms, the magnetic dipole and electric quadrupole contributions, are smaller by a factor of order $(v/c)^2$. For example, as will be seen in Problem Set 5, a current loop with magnetic dipole moment amplitude M_0 radiates a time-averaged power $\langle P_{\rm mag}\rangle = \mu_0 M_0^2 \omega^4/(12\pi c^3)$. For charges q moving with velocities $v \sim d\omega$ over a region of size d, the magnetic dipole moment is of order $M_0 \sim q d^2 \omega$, so that:

$$\frac{\langle P_{\rm mag}\rangle}{\langle P_{\rm rad}\rangle} \sim \left(\frac{M_0}{p_0 c}\right)^2 \sim \left(\frac{d\omega}{c}\right)^2 = \left(\frac{v}{c}\right)^2 \ll 1.$$

For the dipole considered here, the positive charge moves along a straight line through the origin, so that its position and velocity are parallel and its magnetic dipole moment, $\frac{1}{2} q\, \mathbf{r}\times\mathbf{v}$, vanishes. The leading correction then comes from the electric quadrupole moment, $Q_0 \sim q d^2$, which is suppressed by the same factor $(v/c)^2$. In either case, the correction is negligible for non-relativistic motion, consistent with the assumption $d \ll \lambda$ stated above.

Thomson and Rayleigh scattering

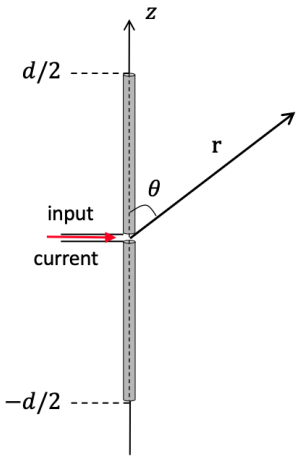
When an electromagnetic wave interacts with a charged particle, it exerts a force that causes the particle to oscillate like a dipole. Consequently, the particle radiates energy in directions other than that of the incident wave, leading to the **scattering of the electromagnetic wave**. If the particle's motion is non-relativistic, the emitted radiation has the same frequency as the incident wave. For a free charged particle, such as a free electron, this phenomenon is known as **Thomson scattering**. In this case, the amplitude of the oscillation, and therefore the dipole moment, varies as $1/\omega^2$, so that the radiated power, which is proportional to $\omega^4 p_0^2$, does not depend on the frequency. The energy carried by this reradiated field is supplied entirely by the incident wave, which therefore loses a corresponding amount of energy, since the incident electric field does work on the charge.

In the atmosphere, electrons bound in molecules are set into oscillation by the electric field of electromagnetic waves originating from the Sun. Since the size of the molecules is much smaller than the wavelength of visible light, the radiation can be analysed using the formula derived earlier for an oscillating dipole. Furthermore, the frequency of visible light is much smaller than the resonance frequency ω_0 of the bound electrons, so that the restoring force dominates over the inertia of the electrons. The amplitude of the oscillation is then of order $eE_0/(m_e\omega_0^2)$, independent of the frequency ω of the incident wave. The scattering of light by particles much smaller than the wavelength under these conditions is known as **Rayleigh scattering**. In this case, the radiated power is proportional to $\omega^4 \propto 1/\lambda^4$, meaning that blue light is scattered several times more effectively than red light (about five times for wavelengths of 450 nm and 700 nm). This scattered radiation draws its energy from the incident sunlight, which is weakened accordingly.

As a result, light observed away from the direction of the incident beam is enriched in high-frequency (blue) components compared to the spectral distribution of the incident sunlight, while the transmitted beam becomes progressively redder. This explains why the sky appears blue.

At sunset, light travels through up to 40 times more air in the atmosphere than at midday, leading to significant scattering of blue light out of the observer's line of sight. Consequently, the remaining light is dominated by red wavelengths, making sunsets appear red.

Radiation from an antenna



An antenna is a device that converts electric power into electromagnetic radiation, or vice versa. Here, we consider a linear antenna extending from $-d/2$ to $+d/2$ along the z-axis, with a small gap at $z = 0$ where a transmitter supplies current. The current in the antenna is described by $I(z,t) = I(z)\cos\omega t$, with:

$$I(z) = I_0\left(1 - \frac{2|z|}{d}\right),$$

and we denote the volume charge density by $\rho(z,t)$.

The charge conservation equation gives $\partial\rho/\partial t = -\boldsymbol{\nabla}\cdot\mathbf{j}$, where $\mathbf{j} = (I/s)\hat{\mathbf{z}}$ is the current density and s is the cross-sectional area of the wire. This yields $\partial\rho_\ell/\partial t = -\partial I/\partial z$, where $\rho_\ell(z,t) = \rho(z,t)s$ is the charge per unit length in the antenna. Substituting the expression for I, we find:

$$\rho_\ell(z,t) = \pm\frac{2I_0}{\omega d}\sin\omega t,$$

where the upper sign applies for $z > 0$ and the lower sign for $z < 0$.

The charge contained in an infinitesimal length $\mathrm{d}z$ at position z is $\mathrm{d}q = \rho_\ell(z,t)\mathrm{d}z$. The dipole moment of this charge distribution is:

$$\mathbf{p}(t) = \hat{\mathbf{z}} \int_{-d/2}^{d/2} z \mathrm{d}q = \frac{2I_0}{\omega d} \sin\omega t \left(-\int_{-d/2}^{0} z\mathrm{d}z + \int_{0}^{d/2} z\mathrm{d}z \right) \hat{\mathbf{z}} = \frac{I_0 d}{2\omega} \sin\omega t \, \hat{\mathbf{z}}.$$

The antenna can therefore be modelled as an oscillating dipole. Assuming $d \ll \lambda$, where λ is the wavelength of the electromagnetic field produced by the antenna, the time-averaged total power radiated is given by equation (6.43) with $p_0 = I_0 d/(2\omega)$:

$$\langle P_{\mathrm{rad}} \rangle = \frac{I_0^2}{48\pi\epsilon_0 c}(kd)^2. \tag{6.44}$$

Note that antennas designed to emit radio waves typically have lengths comparable to the wavelength of the electromagnetic field they produce. As a result, the radiated power for such waves cannot be calculated using the above formulae.

6.2.2 Radiation from a relativistic accelerated point charge

We now consider the general case where the velocity of the charge may be relativistic.

Lorentz invariance of radiated power

In principle, the radiated power P_{rad} can be calculated directly by constructing the Poynting vector from the electric and magnetic fields given by equations (6.35) and (6.36), retaining only the radiation components. However, it is simpler to extend the result obtained for a non-relativistic charge, noting that P_{rad} is Lorentz invariant.

To demonstrate this, let us consider the charge in its instantaneous rest frame (S_t) at time t. Denote $\mathrm{d}U'$ the energy radiated by the particle during a time interval $\mathrm{d}t'$, as measured in (S_t). During this interval, the charge's velocity increases by $\mathrm{d}v'$. Since $\mathrm{d}t'$ is infinitesimally small, $\mathrm{d}v' \ll c$, meaning the charge is non-relativistic in (S_t). Consequently, the radiated power, $\mathrm{d}U'/\mathrm{d}t'$, is given by equation (6.42).

In frame (S), where the observer is at rest, let $\mathrm{d}t$ and $\mathrm{d}U$ denote the corresponding time interval and emitted energy. We have $c\mathrm{d}t = \gamma\left(c\mathrm{d}t' + (\mathbf{v}/c)\cdot\mathrm{d}\mathbf{s}'\right)$, where γ is the Lorentz factor associated with the velocity $v(t)$ and $\mathrm{d}\mathbf{s}'$ is the displacement of the particle in (S_t) during $\mathrm{d}t'$. Since the particle's velocity increases from zero to $\mathrm{d}v'$ during $\mathrm{d}t'$, it follows that $\mathrm{d}s' < \mathrm{d}v'\mathrm{d}t'$. To first order in $\mathrm{d}t'$, we may neglect this term and obtain $\mathrm{d}t = \gamma\mathrm{d}t'$. This expresses the fact that the infinitesimal interval $\mathrm{d}t'$ in (S_t) is a *proper time* interval, since the particle is at rest in this frame to first order.

To calculate $\mathrm{d}U$, consider the field as being composed of photons. A photon emitted during $\mathrm{d}t'$ in the direction of the unit vector $\widehat{\mathbf{n}}'$, making an angle θ' with the acceleration, has a 4-momentum $(E'/c, \mathbf{p}')$, with $\mathbf{p}' = (E'/c)\,\widehat{\mathbf{n}}'$. In frame (S), this photon's energy transforms to $\gamma\,(E' + \mathbf{v}\cdot\mathbf{p}')$. Because the power radiated in (S_t) is proportional to $\sin^2\theta'$, which is unchanged when $\widehat{\mathbf{n}}'$ is replaced by $-\widehat{\mathbf{n}}'$ (that is to say θ' by $\pi - \theta'$), for every photon emitted in the direction $\widehat{\mathbf{n}}'$, there is another photon of the same energy emitted in the opposite direction, $-\widehat{\mathbf{n}}'$, with momentum $-\mathbf{p}'$. The energy of these two photons in (S) is then $2\gamma E'$, since their momenta cancel. Summing over all photons emitted during $\mathrm{d}t'$, the total emitted energy is $\mathrm{d}U = \gamma \mathrm{d}U'$.

Consequently, $\mathrm{d}U/\mathrm{d}t = \mathrm{d}U'/\mathrm{d}t'$, confirming that **the radiated power P_{rad} is Lorentz invariant. Since P_{rad} is given by equation (6.42) in the instantaneous rest frame, it has the same value in any frame, regardless of whether the charge is relativistic or not.**

Here, $P_{\mathrm{rad}} = \mathrm{d}U/\mathrm{d}t$ is the energy emitted per unit time *of emission*. For a relativistic charge, the power received by a distant observer per unit time of reception differs from it since, according to equation (6.33), $\mathrm{d}t_{\mathrm{reception}} = \left(1 - \boldsymbol{\beta}\cdot\widehat{\mathbf{R}}\right)\mathrm{d}t_c$, which depends on the direction of observation.

However, although expression (6.42) is valid in the instantaneous rest frame, where a is the proper (rest frame) acceleration, it is not valid in a general frame. The acceleration $\mathbf{a}$ is not Lorentz invariant, and the radiated power is not in general proportional to a^2. A covariant expression must therefore be constructed from Lorentz invariants.

Covariant representation of radiated power

In the instantaneous rest frame of the charge, the 4-acceleration is $A'^{\mu} = (0, \mathbf{a}')$, where $\mathbf{a}'$ is the proper acceleration. In this frame, the radiated power is given by equation (6.42) as $P_{\mathrm{rad}} = q^2 A'^{\mu} A'_{\mu} / \left(6\pi\epsilon_0 c^3\right)$.

The quantity $A'^{\mu}A'_{\mu}$ is a Lorentz invariant. Since the radiated power is itself Lorentz invariant, its functional dependence on the motion of the charge must also be expressible in terms of Lorentz scalars. Moreover, the radiation field of an accelerated charge (eq. [6.35]) depends only on its velocity and acceleration, so any covariant expression for the power can involve only the 4-velocity U^{μ} and the 4-acceleration A^{μ}. Because $U^{\mu}A_{\mu} = 0$, the only independent scalar one can form from these is $A^{\mu}A_{\mu}$. Therefore the covariant generalisation of the Larmor formula is uniquely determined to be:

$$\boxed{P_{\mathrm{rad}} = \frac{q^2}{6\pi\epsilon_0 c^3} A^{\mu} A_{\mu}.} \tag{6.45}$$

The 4-acceleration is given by equation (4.37), from which we derive:

$$A^\mu A_\mu = \gamma^4 \left[-\gamma^4 \left(\frac{\mathbf{v}\cdot\mathbf{a}}{c} \right)^2 + \left(\gamma^2 \frac{\mathbf{v}\cdot\mathbf{a}}{c^2}\mathbf{v} + \mathbf{a} \right)^2 \right] = \gamma^6 \left[a^2 + \left(\frac{\mathbf{v}\cdot\mathbf{a}}{c} \right)^2 - \frac{a^2 v^2}{c^2} \right], \tag{6.46}$$

where $\mathbf{a} = \mathrm{d}\mathbf{v}/\mathrm{d}t_c$ is the spatial acceleration. The last two terms in the brackets combine to yield $-(\mathbf{v}\times\mathbf{a})^2/c^2$, so the radiated power P_{rad} can be expressed as:

$$\boxed{P_{\mathrm{rad}} = \frac{q^2}{6\pi\epsilon_0 c^3}\gamma^6 \left[a^2 - \left(\frac{\mathbf{v}\times\mathbf{a}}{c} \right)^2 \right].} \tag{6.47}$$

This relativistic extension of Larmor's formula is known as the **Liénard formula**. For linear motion ($\mathbf{v} \parallel \mathbf{a}$), it gives $P_{\mathrm{rad}} \propto \gamma^6 a^2$ while, for circular motion ($\mathbf{v} \perp \mathbf{a}$), it gives $P_{\mathrm{rad}} \propto \gamma^4 a^2$. The strong dependence on γ explains why highly relativistic charged particles in circular accelerators lose large amounts of energy through *synchrotron radiation*.

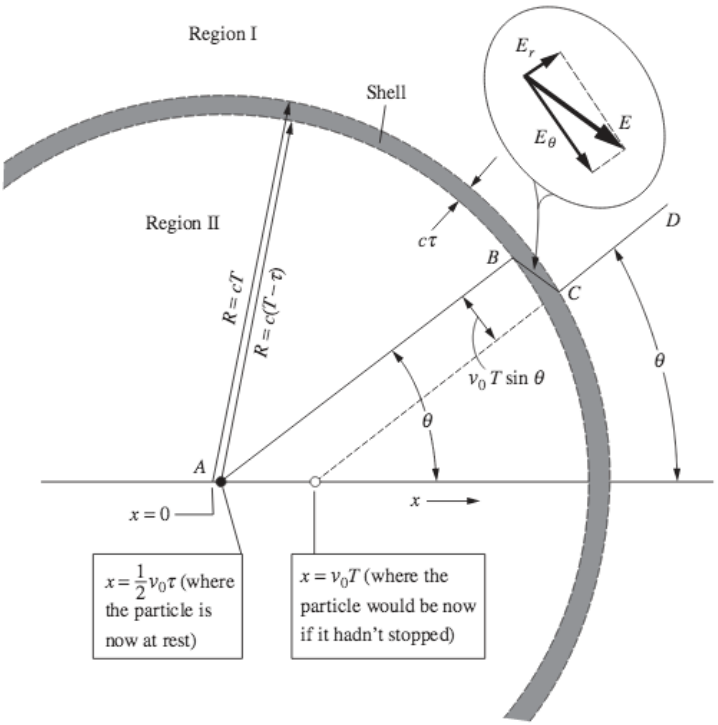
Appendix A

Radiation from moving charges

We derive the power radiated by a non-relativistic accelerated point charge, following the approach given by Purcell (originally proposed by J. J. Thomson). This derivation complements the calculation presented in section 6.2.1 and provides additional insight into the structure of the radiation field.

We consider a particle with charge q that has been travelling at constant velocity $v_0 \ll c$ along the x–axis during the time interval $t = -\infty$ to $t = 0$. At $t = 0$, the particle is located at $x = 0$ and starts to decelerate at a constant rate a, coming to rest at time $t = \tau$. The velocity of the particle during this deceleration is given by $v = -at + v_0$, leading to $\tau = v_0/a$. The position of the particle between $t = 0$ and $t = \tau$ is $x = -at^2/2 + v_0 t$. At $t = \tau$, when the particle comes to rest, its position is $x = v_0\tau/2$.

We now analyse the electric field produced by this point charge at a time $T \gg \tau$.

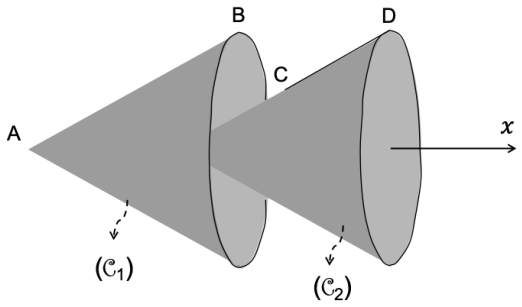


(From E. Purcell & D. Morin, *Electricity and Magnetism*)

The information that the particle begins to decelerate propagates outwards from $x = 0$ at the speed c. By time $t = T$, this information has reached a distance cT from $x = 0$. Consequently, only observers within a sphere of radius cT, centered at $x = 0$, are aware that the particle has decelerated. Observers beyond this sphere (region I in the figure) still perceive the electric field of a charge moving at a constant velocity v_0. If the charge had not decelerated, it would have reached $x = v_0T$ at time $t = T$. Observers in region I thus observe the electric field produced by a charge located at $x = v_0T$, moving with velocity v_0. This is because the particle had been travelling at constant velocity since $t = -\infty$, resulting in the whole space filled with field lines that follow the motion of the charge. As shown in section 6.1.1, the electric field of a charge moving at constant velocity points radially away from the *instantaneous* position of the charge. For instance, at point C in the figure, the electric field is directed along the line segment CD.

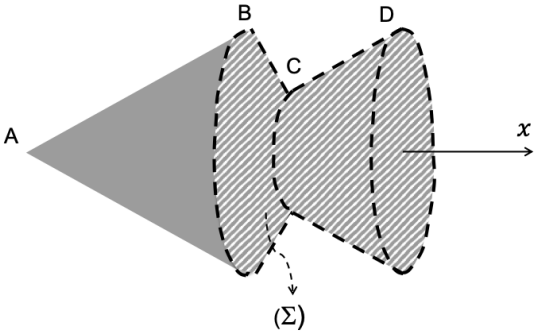
The information that the particle has come to rest propagates outwards from $x = v_0\tau/2$ at the speed c. By time $t = T$, this information has travelled a distance $c(T - \tau)$ from $x = v_0\tau/2$. Consequently, for observers within a sphere of radius $c(T - \tau)$ centered at $x = v_0\tau/2$ (region II in the figure), the electric field corresponds to that of a stationary charge located at $x = v_0\tau/2$. At point B in the figure, the electric field points along the line segment AB and is given by $E_B = q/[4\pi\epsilon_0 c^2(T - \tau)^2]$. Point B is chosen such that the angle θ between AB and the x–axis is the same as that between CD and the x–axis. Since $v_0 \ll c$ and $\tau \ll T$, the position $x = v_0\tau/2 \ll cT$. This implies that the separation between the points at $x = 0$ and at $x = v_0\tau/2$ is negligible compared to the other distances in the problem. Thus, $E_B \simeq q/(4\pi\epsilon_0 R^2)$, where $R = cT$.

We now proceed to calculate the electric field in the transition region, which corresponds to the spherical shell of thickness $c\tau$.



The field line CD is produced by a moving particle, and thus the density of field lines in the cone $(\mathcal{C}_2)$, which surrounds the x–axis, decreases as we approach the axis (refer to section 6.1.1). However, under the assumption that $v_0 \ll c$, this effect can be neglected.

Consequently, since the two cones $(\mathcal{C}_1)$ and $(\mathcal{C}_2)$ have the same opening angle, they carry the same amount of flux originating from the charge q.



This observation implies that the flux of **E** through the surface (Σ), bounded by the bases of $(\mathcal{C}_1)$ and $(\mathcal{C}_2)$, is zero. For this to be true, the line segments AB and CD must belong to the same field line, connected by the segment BC.

From the geometry of the problem, it follows that $E_\theta/E_r = v_0 T \sin\theta/(c\tau)$, where E_r and E_θ are the radial and transverse components of the electric field within the shell, respectively. Since the perpendicular component of **E** must remain continuous, we have $E_r = E_B$, which yields:

$$E_\theta = \frac{q v_0 \sin\theta}{4\pi\epsilon_0 c^2 \tau R},$$

where we have used $T = R/c$. Substituting $v_0/\tau = a$, we obtain:

$$\boxed{E_\theta = \frac{qa\sin\theta}{4\pi\epsilon_0 c^2 R}}. \tag{A.1}$$

We arrive at the important result that E_θ scales as $1/R$, unlike the field of a static charge, which scales as $1/R^2$. Since $E_r \propto 1/R^2$, it follows that as $R \to \infty$, or equivalently $T \to \infty$, $E \to E_\theta$, meaning the electric field becomes purely transverse at large distances.

Accompanying this electric field is a magnetic field given by $B_\varphi = E_\theta/c$, which is perpendicular to both $\hat{\mathbf{R}}$, the unit vector radial vector, and **E**. Since $E_\theta B_\varphi \propto 1/R^2$, the Poynting vector integrated over the surface of a sphere of radius R remains finite, indicating that energy is radiated away. For this reason, E_θ and B_φ are referred to as the **radiation fields**, denoted by E_{rad} and B_{rad}. In contrast, the components of the fields that decay faster than $1/R$ dominate near the charge and are known as *near–zone* fields.