611: Electromagnetic Theory II

CONTENTS

• Special relativity; Lorentz covariance of Maxwell equations

• Scalar and vector potentials, and gauge invariance

• Relativistic motion of charged particles

• Action principle for electromagnetism; energy-momentum tensor

• Electromagnetic waves; waveguides

• Fields due to moving charges

• Radiation from accelerating charges

• Antennae

• Radiation reaction

• Magnetic monopoles, duality, Yang-Mills theory

Suggested textbooks:

• L.D. Landau and E.M. Lifshitz, *The Classical Theory of Fields*

• J.D. Jackson, *Classical Electrodynamics*
## Contents

1 Electrodynamics and Special Relativity ................................. 4
  1.1 Introduction ........................................................................ 4
  1.2 The Lorentz Transformation .............................................. 9
  1.3 An interlude on 3-vectors and suffix notation ...................... 12
  1.4 4-vectors and 4-tensors ..................................................... 17
  1.5 Lorentz tensors ................................................................. 25
  1.6 Proper time and 4-velocity ............................................... 27

2 Electrodynamics and Maxwell’s Equations ............................. 29
  2.1 Natural units ...................................................................... 29
  2.2 Gauge potentials and gauge invariance ............................... 30
  2.3 Maxwell’s equations in 4-tensor notation ......................... 32
  2.4 Lorentz transformation of $\vec{E}$ and $\vec{B}$ ....................... 40
  2.5 The Lorentz force ............................................................ 41
  2.6 Action principle for charged particles ............................... 43
  2.7 Gauge invariance of the action ......................................... 46
  2.8 Canonical momentum, and Hamiltonian ............................ 47

3 Particle Motion in Static Electromagnetic Fields ..................... 48
  3.1 Description in terms of potentials ...................................... 49
  3.2 Particle motion in static uniform $\vec{E}$ and $\vec{B}$ fields ............ 50

4 Action Principle for Electrodynamics ...................................... 55
  4.1 Invariants of the electromagnetic field ............................... 55
  4.2 Action for electrodynamics .............................................. 58
  4.3 Inclusion of sources ........................................................ 61
  4.4 Energy density and energy flux ....................................... 64
  4.5 Energy-momentum tensor .............................................. 66
  4.6 Energy-momentum tensor for the electromagnetic field ......... 72
  4.7 Inclusion of massive charged particles ............................. 76

5 Coulomb’s Law .................................................................... 78
  5.1 Potential of a point charge ............................................... 78
  5.2 Electrostatic energy ........................................................ 80
5.3 Field of a uniformly moving charge ........................................... 81
5.4 Motion of a charge in a Coulomb potential ................................. 84

6 Electromagnetic Waves ............................................................... 90
6.1 Wave equation ................................................................. 90
6.2 Monochromatic plane waves .................................................. 94
6.3 Motion of a point charge in a linearly-polarised E.M. wave ............. 96
6.4 Circular and elliptical polarisation .......................................... 98
6.5 General superposition of plane waves ...................................... 100
6.6 Gauge invariance and electromagnetic fields .............................. 107
6.7 Fourier decomposition of electrostatic fields ............................... 110
6.8 Waveguides ........................................................................ 111
6.9 Resonant cavities .................................................................. 119

7 Fields Due to Accelerating Charges .............................................. 120
7.1 Retarded potentials .............................................................. 120
7.2 Lienard-Wiechert potentials ................................................... 124
7.3 Electric and magnetic fields of a moving charge ......................... 129
7.4 Radiation by accelerated charges ............................................ 132
7.5 Applications of Larmor formula .............................................. 134
7.6 Angular distribution of the radiated power ................................. 137
7.7 Frequency distribution of radiated energy ................................ 142
7.8 Frequency spectrum for relativistic circular motion ..................... 146
7.9 Frequency spectrum for periodic motion .................................. 149
7.10 Cerenkov radiation .............................................................. 151
7.11 Thompson scattering ........................................................... 155

8 Radiating Systems ....................................................................... 157
8.1 Fields due to localised oscillating sources .................................. 157
8.2 Electric dipole radiation .......................................................... 161
8.3 Higher multipoles ................................................................... 165
8.4 Linear antenna ...................................................................... 171

9 Electromagnetism and Quantum Mechanics .................................. 174
9.1 The Schrödinger equation and gauge transformations .................. 174
9.2 Magnetic monopoles .............................................................. 177
9.3 Dirac quantisation condition ........................................... 181

10 Local Gauge Invariance and Yang-Mills Theory .................. 182
  10.1 Relativistic quantum mechanics ..................................... 182
  10.2 Yang-Mills theory ..................................................... 186
1 Electrodynamics and Special Relativity

1.1 Introduction

The final form of Maxwell’s equations describing the electromagnetic field had been established by 1865. Although this was forty years before Einstein formulated his theory of special relativity, the Maxwell equations are, remarkably, fully consistent with the special relativity. The Maxwell theory of electromagnetism is the first, and in many ways the most important, example of what is known as a classical relativistic field theory.

Our emphasis in this course will be on establishing the formalism within which the relativistic invariance of electrodynamics is made manifest, and thereafter exploring the relativistic features of the theory.

In Newtonian mechanics, the fundamental laws of physics, such as the dynamics of moving objects, are valid in all inertial frames (i.e. all non-accelerating frames). If $S$ is an inertial frame, then the set of all inertial frames comprises all frames that are in uniform motion relative to $S$. Suppose that two inertial frames $S$ and $S'$, are parallel, and that their origins coincide at at $t = 0$. If $S'$ is moving with uniform velocity $\vec{v}$ relative to $S$, then a point $P$ with position vector $\vec{r}$ with respect to $S$ will have position vector $\vec{r}'$ with respect to $S'$, where

$$\vec{r}' = \vec{r} - \vec{v} t.$$  \hspace{1cm} (1.1)
Of course, it is always understood in Newtonian mechanics that time is absolute, and so the times $t$ and $t'$ measured by observers in the frames $S$ and $S'$ are the same:

$$ t' = t. \tag{1.2} $$

The transformations (1.1) and (1.2) form part of what is called the Galilean Group. The full Galilean group includes also rotations of the spatial Cartesian coordinate system, so that we can define

$$ \vec{r}' = M \cdot \vec{r} - \vec{v} t, \quad t' = t, \tag{1.3} $$

where $M$ is an orthogonal $3 \times 3$ constant matrix acting by matrix multiplication on the components of the position vector:

$$ \vec{r} \leftrightarrow \begin{pmatrix} x \\ y \\ z \end{pmatrix}, \quad M \cdot \vec{r} \leftrightarrow M \begin{pmatrix} x \\ y \\ z \end{pmatrix}, \tag{1.4} $$

where $M^T M = 1$, and $M^T$ denotes the transpose of the matrix $M$.

Returning to our simplifying assumption that the two frames are parallel, i.e. that $M = 1$, it follows that if a particle having position vector $\vec{r}$ in $S$ moves with velocity $\vec{u} = d\vec{r}/dt$, then its velocity $\vec{u}' = d\vec{r}'/dt$ as measured with respect to the frame $S'$ is given by

$$ \vec{u}' = \vec{u} - \vec{v}. \tag{1.5} $$

Suppose, for example, that $\vec{v}$ lies along the $x$ axis of $S$; i.e. that $S'$ is moving along the $x$ axis of $S$ with speed $v = |\vec{v}|$. If a beam of light were moving along the $x$ axis of $S$ with speed $c$, then the prediction of Newtonian mechanics and the Galilean transformation would therefore be that in the frame $S'$, the speed $c'$ of the light beam would be

$$ c' = c - v. \tag{1.6} $$

Of course, as is well known, this contradicts experiment. As far as we can tell, with experiments of ever-increasing accuracy, the true state of affairs is that the speed of the light beam is the same in all inertial frames. Thus the predictions of Newtonian mechanics and the Galilean transformation are falsified by experiment.

Of course, it should be emphasised that the discrepancies between experiment and the Galilean transformations are rather negligible if the relative speed $v$ between the two inertial frames is of a typical “everyday” magnitude, such as the speed of a car or a plane. But if
v begins to become appreciable in comparison to the speed of light, then the discrepancy becomes appreciable too.

By contrast, it turns out that Maxwell’s equations of electromagnetism do predict a constant speed of light, independent of the choice of inertial frame. To be precise, let us begin with the free-space Maxwell’s equations,

\[ \nabla \cdot \vec{E} = \frac{1}{\epsilon_0} \rho, \quad \nabla \times \vec{B} - \mu_0 \epsilon_0 \frac{\partial \vec{E}}{\partial t} = \mu_0 \vec{J}, \]
\[ \nabla \cdot \vec{B} = 0, \quad \nabla \times \vec{E} + \frac{\partial \vec{B}}{\partial t} = 0, \quad (1.7) \]

where \( \vec{E} \) and \( \vec{B} \) are the electric and magnetic fields, \( \rho \) and \( \vec{J} \) are the charge density and current density, and \( \epsilon_0 \) and \( \mu_0 \) are the permittivity and permeability of free space.\(^1\)

To see the electromagnetic wave solutions, we can consider a region of space where there are no sources, i.e. where \( \rho = 0 \) and \( \vec{J} = 0 \). Then we shall have

\[ \nabla \times (\nabla \times \vec{E}) = -\frac{\partial}{\partial t} \nabla \times \vec{B} = -\mu_0 \epsilon_0 \frac{\partial^2 \vec{E}}{\partial t^2}. \quad (1.8) \]

But using the vector identity \( \nabla \times (\nabla \times \vec{E}) = \nabla (\nabla \cdot \vec{E}) - \nabla^2 \vec{E} \), it follows from \( \nabla \cdot \vec{E} = 0 \) that the electric field satisfies the wave equation

\[ \nabla^2 \vec{E} - \mu_0 \epsilon_0 \frac{\partial^2 \vec{E}}{\partial t^2} = 0. \quad (1.9) \]

This admits plane-wave solutions of the form

\[ \vec{E} = \vec{E}_0 e^{i(\vec{k} \cdot \vec{r} - \omega t)}, \quad (1.10) \]

where \( \vec{E}_0 \) and \( \vec{k} \) are constant vectors, and \( \omega \) is also a constant, where

\[ k^2 = \mu_0 \epsilon_0 \omega^2. \quad (1.11) \]

Here \( k \) means \( |\vec{k}| \), the magnitude of the wave-vector \( \vec{k} \). Thus we see that the waves travel at speed \( c \) given by

\[ c = \frac{\omega}{k} = \frac{1}{\sqrt{\mu_0 \epsilon_0}}. \quad (1.12) \]

Putting in the numbers, this gives \( c \approx 3 \times 10^8 \) metres per second, i.e. the familiar speed of light.

\(^1\)The equations here are written using the system of units known as SI, which could be said to stand for “Super Inconvenient.” In these units, the number of unnecessary dimensionful “fundamental constants” is maximised. We shall pass speedily to more convenient units a little bit later.
A similar calculation shows that the magnetic field $\vec{B}$ also satisfies an identical wave equation, and in fact $\vec{B}$ and $\vec{E}$ are related by

$$\vec{B} = \frac{1}{\omega} \vec{k} \times \vec{E}.$$  \hspace{1cm} (1.13)

The situation, then, is that if the Maxwell equations (1.7) hold in a given frame of reference, then they predict that the speed of light will be $c \approx 3 \times 10^8$ metres per second in that frame. Therefore, if we assume that the Maxwell equations hold in all inertial frames, then they predict that the speed of light will have that same value in all inertial frames. Since this prediction is in agreement with experiment, we can reasonably expect that the Maxwell equations will indeed hold in all inertial frames. Since the prediction contradicts the implications of the Galilean transformations, it follows that the Maxwell equations are not invariant under Galilean transformations. This is just as well, since the Galilean transformations are wrong!

In fact, as we shall see, the transformations that correctly describe the relation between observations in different inertial frames in uniform motion are the Lorentz Transformations of Special Relativity. Furthermore, even though the Maxwell equations were written down in the pre-relativity days of the nineteenth century, they are in fact perfectly invariant under the Lorentz transformations. No further modification is required in order to incorporate Maxwell’s theory of electromagnetism into special relativity.

However, the Maxwell equations as they stand, written in the form given in equation (1.7), do not look manifestly covariant with respect to Lorentz transformations. This is because they are written in the language of 3-vectors. To make the Lorentz transformations look nice and simple, we should instead express them in terms of 4-vectors, where the extra component is associated with the time direction.

Actually, before proceeding it is instructive to take a step back and look at what the Maxwell equations actually looked like in Maxwell’s 1865 paper *A Dynamical Theory of the Electromagnetic Field*, published in the Philosophical Transactions of the Royal Society of London. It must be recalled that in 1865 three-dimensional vectors had not yet been invented, and so everything was written out explicitly in terms of the $x$, $y$ and $z$ components.\(^3\) To make matters worse, Maxwell used a different letter of the alphabet for each component of each field. In terms of the now-familiar electric vector fields $\vec{E}$, $\vec{D}$, the magnetic fields $\vec{B}$, $\vec{H}$, the current density $\vec{J}$ and the charge density $\rho$, Maxwell’s chosen names for the

\(^2\)Strictly, as will be explained later, we should say covariant rather than invariant.

\(^3\)Vectors were invented independently by Josiah Willard Gibbs, and Oliver Heaviside, around the end of the 19th century.
components were
\[ \vec{E} = (P, Q, R), \quad \vec{D} = (f, g, h), \]
\[ \vec{B} = (F, G, H), \quad \vec{H} = (\alpha, \beta, \gamma), \]
\[ \vec{J} = (p, q, r), \quad \rho = e. \quad (1.14) \]

Thus the Maxwell equations that we now write rather compactly as
\[ \vec{\nabla} \cdot \vec{D} = 4\pi \rho, \quad \vec{\nabla} \cdot \vec{B} = 0, \quad (1.15) \]
\[ \vec{\nabla} \times \vec{H} - \frac{\partial \vec{D}}{\partial t} = 4\pi \vec{J}, \quad \vec{\nabla} \times \vec{E} + \frac{\partial \vec{B}}{\partial t} = 0, \quad (1.16) \]
took, in 1865, the highly inelegant forms
\[
\frac{\partial f}{\partial x} + \frac{\partial g}{\partial y} + \frac{\partial h}{\partial z} = 4\pi e, \\
\frac{\partial F}{\partial x} + \frac{\partial G}{\partial y} + \frac{\partial H}{\partial z} = 0, \quad (1.17) 
\]
for the two equations in (1.15), and
\[
\frac{\partial \gamma}{\partial y} - \frac{\partial \beta}{\partial z} - \frac{\partial f}{\partial t} = 4\pi p, \\
\frac{\partial \alpha}{\partial z} - \frac{\partial \gamma}{\partial x} - \frac{\partial g}{\partial t} = 4\pi q, \\
\frac{\partial \beta}{\partial x} - \frac{\alpha}{\partial y} - \frac{\partial h}{\partial t} = 4\pi r, \\
\frac{\partial R}{\partial y} - \frac{\partial Q}{\partial z} + \frac{\partial F}{\partial t} = 0, \\
\frac{\partial P}{\partial z} - \frac{\partial R}{\partial x} + \frac{\partial G}{\partial t} = 0, \\
\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} + \frac{\partial H}{\partial t} = 0, \quad (1.18) 
\]
for the two vector-valued equations in (1.16). Not only does Maxwell’s way of writing his equations, in (1.17) and (1.18), look like a complete mess, but it also completely fails to make manifest the familiar fact that the equations are symmetric under arbitrary rotations of the three-dimensional \((x, y, z)\) coordinate system. Of course in the 3-vector notation of (1.15) and (1.16) this rotational symmetry is completely manifest; that is precisely what the vectors notation was invented for, to make manifest the rotational symmetry of three-dimensional equations like the Maxwell equations. The symmetry is, of course, actually there in Maxwell’s equations (1.17) and (1.18), but it is completely obscure and non-obvious.

\footnote{Here, we are writing the equations in the so-called “Natural Units,” which we shall be using throughout this course.}
So the moral of the story is one not only wants equations that have the nice symmetries, but one wants to write them in a notation that makes these symmetries manifest. It is worth bearing this in mind when we pursue our goal of re-writing the Maxwell equations in a notation that does even more, and makes their symmetry under Lorentz transformations manifest. In order to give a nice elegant treatment of the Lorentz transformation properties of the Maxwell equations, we should first therefore reformulate special relativity in terms of 4-vectors and 4-tensors. Since there are many different conventions on offer in the marketplace, we shall begin with a review of special relativity in the notation that we shall be using in this course.

1.2 The Lorentz Transformation

The derivation of the Lorentz transformation follows from Einstein’s two postulates:

- The laws of physics are the same for all inertial observers.
- The speed of light is the same for all inertial observers.

To derive the Lorentz transformation, let us suppose that we have two inertial frames $S$ and $S'$, whose origins coincide at time zero, that is to say, at $t = 0$ in the frame $S$, and at $t' = 0$ in the frame $S'$. If a flash of light is emitted at the origin at time zero, then it will spread out over a spherical wavefront given by

$$x^2 + y^2 + z^2 - c^2 t^2 = 0 \quad (1.19)$$

in the frame $S$, and by

$$x'^2 + y'^2 + z'^2 - c^2 t'^2 = 0 \quad (1.20)$$

in the frame $S'$. Note that, following the second of Einstein’s postulates, we have used the same speed of light $c$ for both inertial frames. Our goal is to derive the relation between the coordinates $(x, y, z, t)$ and $(x', y', z', t')$ in the two inertial frames.

Consider for simplicity the case where $S'$ is parallel to $S$, and moves along the $x$ axis with velocity $v$. Clearly we must have

$$y' = y, \quad z' = z. \quad (1.21)$$

Furthermore, the transformation between $(x, t)$ and $(x', t')$ must be a linear one, since otherwise it would not be translation-invariant or time-translation invariant. Thus we may say that

$$x' = Ax + Bt, \quad t' = Cx + Dt, \quad (1.22)$$
for constants $A$, $B$, $C$ and $D$ to be determined.

Now, if $x' = 0$, this must, by definition, correspond to the equation $x = vt$ in the frame $S$, and so from the first equation in (1.22) we have $B = -Av$, and so we have

$$x' = A(x - vt). \quad (1.23)$$

By the same token, if we exchange the roles of the primed and the unprimed frames, and consider the origin $x = 0$ for the frame $S$, then this will correspond to $x' = -vt'$ in the frame $S'$. (If the origin of the frame $S'$ moves along $x$ in the frame $S$ with velocity $v$, then the origin of the frame $S$ must be moving along $x'$ in the frame $S'$ with velocity $-v$.) It follows that we must have

$$x = A(x' + vt'). \quad (1.24)$$

Note that it must be the same constant $A$ in both these equations, since the two really just correspond to reversing the direction of the $x$ axis, and the physics must be the same for the two cases.

Now we bring in the postulate that the speed of light is the same in the two frames, so if we have $x = ct$ then this must imply $x' = ct'$. Solving the resulting two equations

$$ct' = A(c - v)t, \quad ct = A(c + v)t' \quad (1.25)$$

for $A$, we obtain

$$A = \frac{1}{\sqrt{1 - v^2/c^2}}. \quad (1.26)$$

Solving $x^2 - c^2t^2 = x'^2 - c^2t'^2$ for $t'$, after using (1.23), we find $t'^2 = A^2(t - vx/c^2)^2$ and hence

$$t' = A(t - \frac{v}{c^2}x). \quad (1.27)$$

(We must choose the positive square root since it must reduce to $t' = +t$ if the velocity $v$ goes to zero.) At this point we shall change the name of the constant $A$ to the conventional one $\gamma$, and thus we arrive at the Lorentz transformation

$$x' = \gamma(x - vt), \quad y' = y, \quad z' = z, \quad t' = \gamma(t - \frac{v}{c^2}x), \quad (1.28)$$

where

$$\gamma = \frac{1}{\sqrt{1 - v^2/c^2}}, \quad (1.29)$$

in the special case where $S'$ is moving along the $x$ direction with velocity $v$.

At this point, for notational convenience, we shall introduce the simplification of working in a system of units in which the speed of light is set equal to 1. We can do this because the
speed of light is the same for all inertial observers, and so we may as well choose to measure length in terms of the time it takes for light in vacuo to traverse the distance. In fact, the metre is nowadays defined to be the distance travelled by light in vacuo in \(1/299,792,458\) of a second. By making the small change of taking the light-second as the basic unit of length, rather than the \(1/299,792,458'th of a light-second, we end up with a system of units in which \(c = 1\). Alternatively, we could measure time in “light metres,” where the unit is the time taken for light to travel 1 metre. In these units, the Lorentz transformation (1.28) becomes

\[
x' = \gamma(x - vt), \quad y' = y, \quad z' = z, \quad t' = \gamma(t - vx),
\]

where

\[
\gamma = \frac{1}{\sqrt{1 - v^2}}.
\]

It will be convenient to generalise the Lorentz transformation (1.30) to the case where the frame \(S'\) is moving with (constant) velocity \(\vec{v}\) in an arbitrary direction, rather than specifically along the \(x\) axis. It is rather straightforward to do this. We know that there is a complete rotational symmetry in the three-dimensional space parameterised by the \((x, y, z)\) coordinate system. Therefore, if we can first rewrite the special case described by (1.30) in terms of 3-vectors, where the 3-vector velocity \(\vec{v}\) happens to be simply \(\vec{v} = (v, 0, 0)\), then generalisation will be immediate. It is easy to check that with \(\vec{v}\) taken to be \((v, 0, 0)\), the Lorentz transformation (1.30) can be written as

\[
\vec{r}' = \vec{r} + \gamma\frac{1}{v^2} (\vec{v} \cdot \vec{r}) \vec{v} - \gamma \vec{v} t, \quad t' = \gamma(t - \vec{v} \cdot \vec{r}),
\]

with \(\gamma = (1-v^2)^{-1/2}\) and \(v \equiv |\vec{v}|\), and with \(\vec{r} = (x, y, z)\). Since these equations are manifestly covariant under 3-dimensional spatial rotations (i.e. they are written entirely in a 3-vector notation), it must be that they are the correct form of the Lorentz transformations for an arbitrary direction for the velocity 3-vector \(\vec{v}\).

The Lorentz transformations (1.32) are what are called the pure boosts. It is easy to check that they have the property of preserving the spherical light-front condition, in the sense that points on the expanding spherical shell given by \(r^2 = t^2\) of a light-pulse emitted at the origin at \(t = 0\) in the frame \(S\) will also satisfy the equivalent condition \(r'^2 = t'^2\) in the primed reference frame \(S'\). (Note that \(r^2 = x^2 + y^2 + z^2\).) In fact, a stronger statement is true: The Lorentz transformation (1.32) satisfies the equation

\[
x^2 + y^2 + z^2 - t^2 = x'^2 + y'^2 + z'^2 - t'^2.
\]
1.3 An interlude on 3-vectors and suffix notation

Before describing the 4-dimensional spacetime approach to special relativity, it may be helpful to give a brief review of some analogous properties of 3-dimensional Euclidean space, and Cartesian vector analysis.

Consider a 3-vector $\vec{A}$, with $x$, $y$ and $z$ components denoted by $A_1$, $A_2$ and $A_3$ respectively. Thus we may write

$$\vec{A} = (A_1, A_2, A_3). \quad (1.34)$$

It is convenient then to denote the set of components by $A_i$, for $i = 1, 2, 3$.

The scalar product between two vectors $\vec{A}$ and $\vec{B}$ is given by

$$\vec{A} \cdot \vec{B} = A_1B_1 + A_2B_2 + A_3B_3 = \sum_{i=1}^{3} A_iB_i . \quad (1.35)$$

This expression can be written more succinctly using the Einstein Summation Convention. The idea is that when writing valid expressions using vectors, or more generally tensors, on every occasion that a summation of the form $\sum_{i=1}^{3}$ is performed, the summand is an expression in which the summation index $i$ occurs exactly twice. Furthermore, there will be no occasion when an index occurs exactly twice in a given term and a sum over $i$ is not performed. Therefore, we can abbreviate the writing by simply omitting the explicit summation symbol, since we know as soon as we see an index occurring exactly twice in a term of an equation that it must be accompanied by a summation symbol. Thus we can abbreviate (1.35) and just write the scalar product as

$$\vec{A} \cdot \vec{B} = A_iB_i . \quad (1.36)$$

The index $i$ here is called a “dummy suffix.” It is just like a local summation variable in a computer program; it doesn’t matter if it is called $i$, or $j$ or anything else, as long as it doesn’t clash with any other index that is already in use.

The next concept to introduce is the Kronecker delta tensor $\delta_{ij}$. This is defined by

$$\delta_{ij} = 1 \quad \text{if} \quad i = j , \quad \delta_{ij} = 0 \quad \text{if} \quad i \neq j , \quad (1.37)$$

Thus

$$\delta_{11} = \delta_{22} = \delta_{33} = 1 , \quad \delta_{12} = \delta_{13} = \cdots = 0 . \quad (1.38)$$

Note that $\delta_{ij}$ is a symmetric tensor: $\delta_{ij} = \delta_{ji}$. The Kronecker delta clearly has the replacement property

$$A_i = \delta_{ij}A_j , \quad (1.39)$$

12
since by (1.37) the only non-zero term in the summation over \( j \) is the term when \( j = i \).

Now consider the vector product \( \vec{A} \times \vec{B} \). We have

\[
\vec{A} \times \vec{B} = (A_2B_3 - A_3B_2, A_3B_1 - A_1B_3, A_1B_2 - A_2B_1) .
\]  

(1.40)

To write this using index notation, we first define the 3-index totally-antisymmetric tensor \( \epsilon_{ijk} \). Total antisymmetry means that the tensor changes sign if any pair of indices is swapped. For example

\[
\epsilon_{ijk} = -\epsilon_{ikj} = -\epsilon_{jik} = -\epsilon_{kji} .
\]  

(1.41)

Given this total antisymmetry, we actually only need to specify the value of one non-zero component in order to pin down the definition completely. We shall define \( \epsilon_{123} = +1 \). From the total antisymmetry, it then follows that

\[
\epsilon_{123} = \epsilon_{231} = \epsilon_{312} = +1, \quad \epsilon_{132} = \epsilon_{321} = \epsilon_{213} = -1,
\]  

(1.42)

with all other components vanishing.

It is now evident that in index notation, the \( i \)'th component of the vector product \( \vec{A} \times \vec{B} \) can be written as

\[
(\vec{A} \times \vec{B})_i = \epsilon_{ijk}A_jB_k .
\]  

(1.43)

For example, the \( i = 1 \) component (the \( x \) component) is given by

\[
(\vec{A} \times \vec{B})_1 = \epsilon_{1jk}A_jB_k = \epsilon_{123}A_2B_3 + \epsilon_{132}A_3B_2 = A_2B_3 - A_3B_2 ,
\]  

(1.44)

in agreement with the \( x \)-component given in (1.40).

Now, let us consider the vector triple product \( \vec{A} \times (\vec{B} \times \vec{C}) \). The \( i \) component is therefore given by

\[
[\vec{A} \times (\vec{B} \times \vec{C})]_i = \epsilon_{ijk}A_j(\vec{B} \times \vec{C})_k = \epsilon_{ijk}\epsilon_{k\ell m}A_jB_\ell C_m .
\]  

(1.45)

For convenience, we may cycle the indices on the second \( \epsilon \) tensor around and write this as

\[
[\vec{A} \times (\vec{B} \times \vec{C})]_i = \epsilon_{ijk}\epsilon_{\ell mk}A_jB_\ell C_m .
\]  

(1.46)

There is an extremely useful identity, which can be proved simply by considering all possible values of the free indices \( i, j, \ell, m \):

\[
\epsilon_{ijk}\epsilon_{\ell mk} = \delta_{i\ell}\delta_{jm} - \delta_{im}\delta_{j\ell} .
\]  

(1.47)
Using this in (1.46), we have
\[
[\vec{A} \times (\vec{B} \times \vec{C})]_i = (\delta_{ij}\delta_{jm} - \delta_{im}\delta_{ji})A_j B_t C_m,
\]
\[
= \delta_{ij}\delta_{jm}A_j B_t C_m - \delta_{im}\delta_{ji}A_j B_t C_m,
\]
\[
= B_i A_j C_j - C_i A_j B_j,
\]
\[
= (\vec{A} \cdot \vec{C}) B_i - (\vec{A} \cdot \vec{B}) C_i. \tag{1.48}
\]
In other words, we have proven that
\[
\vec{A} \times (\vec{B} \times \vec{C}) = (\vec{A} \cdot \vec{C}) \vec{B} - (\vec{A} \cdot \vec{B}) \vec{C}. \tag{1.49}
\]

It is useful also to apply the index notation to the gradient operator \(\vec{\nabla}\). This is a vector-valued differential operator, whose components are given by
\[
\vec{\nabla} = \left( \frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z} \right).
\]
In terms of the index notation, we may therefore say that the \(i\)’th component \((\vec{\nabla})_i\) of the vector \(\vec{\nabla}\) is given by \(\partial/\partial x_i\). In order to make the writing a little less clumsy, it is useful to rewrite this as
\[
\partial_i = \frac{\partial}{\partial x_i}. \tag{1.51}
\]
Thus, the \(i\)’th component of \(\vec{\nabla}\) is \(\partial_i\).

It is now evident that the divergence and the curl of a vector \(\vec{A}\) can be written in index notation as
\[
\text{div}\vec{A} = \vec{\nabla} \cdot \vec{A} = \partial_i A_i, \quad (\text{curl}\vec{A})_i = (\vec{\nabla} \times \vec{A})_i = \epsilon_{ijk}\partial_j A_k. \tag{1.52}
\]
The Laplacian, \(\nabla^2 = \vec{\nabla} \cdot \vec{\nabla} = \partial^2/\partial x^2 + \partial^2/\partial y^2 + \partial^2/\partial z^2\), is given by
\[
\nabla^2 = \partial_i \partial_i. \tag{1.53}
\]

By the rules of partial differentiation, we have \(\partial_i x_j = \delta_{ij}\). If we consider the position vector \(\vec{r} = (x, y, z)\), then we have \(r^2 = x^2 + y^2 + z^2\), which can be written as
\[
r^2 = x_j x_j. \tag{1.54}
\]
If we now act with \(\partial_i\) on both sides, we get
\[
2r \partial_i r = 2x_j \partial_i x_j = 2x_j \delta_{ij} = 2x_i. \tag{1.55}
\]
Thus we have the very useful result that
\[
\partial_i r = \frac{x_i}{r}. \tag{1.56}
\]
So far, we have not given any definition of what a 3-vector actually is, and now is the
time to remedy this. We may define a 3-vector \( \vec{A} \) as an ordered triplet of real quantities,
\( \vec{A} = (A_1, A_2, A_3) \), which transforms under rigid rotations of the Cartesian axes in the same
way as does the position vector \( \vec{r} = (x, y, z) \). Now, any rigid rotation of the Cartesian
coordinate axes can be expressed as a constant 3 \( \times \) 3 orthogonal matrix \( M \) acting on the
column vector whose components are \( x, y \) and \( z \):

\[
\begin{pmatrix}
  x' \\
  y' \\
  z'
\end{pmatrix} = M
\begin{pmatrix}
  x \\
  y \\
  z
\end{pmatrix},
\]

where

\[
M^T M = 1.
\]

An example would be the matrix describing a rotation by a (constant) angle \( \theta \) around the
z axis, for which we would have

\[
M = \begin{pmatrix}
  \cos \theta & \sin \theta & 0 \\
  -\sin \theta & \cos \theta & 0 \\
  0 & 0 & 1
\end{pmatrix}.
\]

Matrices satisfying the equation (1.58) are called *orthogonal* matrices. If they are of
dimension \( n \times n \), they are called \( O(n) \) matrices. Thus the 3-dimensional rotation matrices
are called \( O(3) \) matrices.\(^5\)

In index notation, we can write \( M \) as \( M_{ij} \), where \( i \) labels the rows and \( j \) labels the
columns:

\[
M = \begin{pmatrix}
  M_{11} & M_{12} & M_{13} \\
  M_{21} & M_{22} & M_{23} \\
  M_{31} & M_{32} & M_{33}
\end{pmatrix}.
\]

The rotation (1.57) can then be expressed as

\[
x_i' = M_{ij} x_j,
\]

\(^5\)There is a little subtlety that we have glossed over, here. If we take the determinant of (1.58), and
use the facts that \( \det(AB) = (\det A)(\det B) \) and \( \det(A^T) = \det A \), we see that \( (\det M)^2 = 1 \) and hence
\( \det M = \pm 1 \). The matrices with \( \det M = +1 \) are called \( SO(n) \) matrices in \( n \) dimensions, where the “S”
stands for “special,” meaning unit determinant. It is actually \( SO(n) \) matrices that are pure rotations. The
transformations with \( \det M = -1 \) are actually rotations combined with a *reflection* of the coordinates (such
as \( x \to -x \)). Thus, the pure rotation group in 3 dimensions is \( SO(3) \).
and the orthogonality condition (1.58) is
\[ M_{ki} M_{kj} = \delta_{ij}. \]  
(1.62)

(Note that if \( M \) has components \( M_{ij} \) then its transpose \( M^T \) has components \( M_{ji} \).) One can directly see in the index notation that the orthogonality condition (1.62) implies, together with (1.61), that the quadratic form \( x_i x_i = x^2 + y^2 + z^2 \) is invariant under these rotations and reflections:
\[ x'_i x'_i = M_{ij} x_j x_k = \delta_{jk} x_j x_k = x_j x_j = x_i x_i. \]  
(1.63)

As stated above, the components of any 3-vector transform the same way under rotations as do the components of the position vector \( \vec{r} \). Thus, if \( \vec{A} \) and \( \vec{B} \) are 3-vectors, then after a rotation by the matrix \( M \) we shall have
\[ A'_i = M_{ij} A_j, \quad B'_i = M_{ij} B_j. \]  
(1.64)

If we calculate the scalar product of \( \vec{A} \) and \( \vec{B} \) after the rotation, we shall therefore have
\[ A'_i B'_i = M_{ij} A_j M_{ik} B_k. \]  
(1.65)

(Note the choice of a different dummy suffix in the expression for \( B'_i \)!) Using the orthogonality condition (1.62), we therefore have that
\[ A'_i B'_i = A_j B_k \delta_{jk} = A_j B_j. \]  
(1.66)

Thus the scalar product of any two 3-vectors is invariant under rotations of the coordinate axes. That is to say, \( A_i B_i \) is a scalar quantity, and by definition a scalar is invariant under rotations.

It is useful to count up how many independent parameters are needed to specify the most general possible rotation matrix \( M \). Looking at (1.60), we can see that a general \( 3 \times 3 \) matrix has 9 components. But our matrix \( M \) is required to be orthogonal, i.e. it must satisfy \( M^T M - 1 = 0 \). How many equations does this amount to? Naively, it is a \( 3 \times 3 \) matrix equation, and so implies 9 conditions. But this is not correct, since the left-hand side of \( M^T M - 1 = 0 \) is in fact a symmetric matrix. (Take the transpose, and verify this.) A \( 3 \times 3 \) symmetric matrix has \( (3 \times 4)/2 = 6 \) independent components, and so setting a symmetric \( 3 \times 3 \) matrix to zero implies only 6 independent equations rather than 9. Thus the orthogonality condition imposes 6 constraints on the 9 components of a general \( 3 \times 3 \) matrix, and so that leaves
\[ 9 - 6 = 3 \]  
(1.67)
as the number of independent components of a $3 \times 3$ orthogonal matrix, it is easy to see that this is the correct counting; to specify a general rotation in 3-dimensional space, we need two angles to specify an axis (for example, the latitude and longitude), and a third angle to specify the rotation around that axis.

The above are just a few simple examples of the use of index notation in order to write 3-vector and 3-tensor expressions in Cartesian 3-tensor analysis. It is a very useful notation when one needs to deal with complicated expressions. As we shall now see, there is a very natural generalisation to the case of vector and tensor analysis in 4-dimensional Minkowski spacetime.

### 1.4 4-vectors and 4-tensors

The Lorentz transformations given in (1.32) are linear in the space and time coordinates. They can be written more succinctly if we first define the set of four spacetime coordinates denoted by $x^\mu$, where $\mu$ is an index, or label, that ranges over the values 0, 1, 2 and 3. The case $\mu = 0$ corresponds to the time coordinate $t$, while $\mu = 1, 2$ and 3 corresponds to the space coordinates $x, y$ and $z$ respectively. Thus we have

$$ (x^0, x^1, x^2, x^3) = (t, x, y, z). $$

(1.68)

Of course, once the abstract index label $\mu$ is replaced, as here, by the specific index values 0, 1, 2 and 3, one has to be very careful when reading a formula to distinguish between, for example, $x^2$ meaning the symbol $x$ carrying the spacetime index $\mu = 2$, and $x^2$ meaning the square of $x$. It should generally be obvious from the context which is meant.

The invariant quadratic form appearing on the left-hand side of (1.33) can now be written in a nice way, if we first introduce the 2-index quantity $\eta_{\mu\nu}$, defined by

$$ \eta_{00} = -1, \quad \eta_{11} = \eta_{22} = \eta_{33} = 1, $$

(1.69)

with $\eta_{\mu\nu} = 0$ if $\mu \neq \nu$. Note that $\eta_{\mu\nu}$ is symmetric:

$$ \eta_{\mu\nu} = \eta_{\nu\mu}. $$

(1.70)

Using $\eta_{\mu\nu}$, the quadratic form on the left-hand side of (1.33) can be rewritten as

$$ x^2 + y^2 + z^2 - t^2 = \sum_{\mu=0}^{3} \sum_{\nu=0}^{3} \eta_{\mu\nu} x^\mu x^\nu. $$

(1.71)

---

6 The choice to put the index label $\mu$ as a superscript, rather than a subscript, is purely conventional. But, unlike the situation with many arbitrary conventions, in this case the coordinate index is placed upstairs in all modern literature.
In the same way as we previously associated 2-index objects in 3-dimensional Euclidean space with $3 \times 3$ matrices, so here too we can associate $\eta_{\mu\nu}$ with a $4 \times 4$ matrix $\eta$:

\[
\eta = \begin{pmatrix}
\eta_{00} & \eta_{01} & \eta_{02} & \eta_{03} \\
\eta_{10} & \eta_{11} & \eta_{12} & \eta_{13} \\
\eta_{20} & \eta_{21} & \eta_{22} & \eta_{23} \\
\eta_{30} & \eta_{31} & \eta_{32} & \eta_{33}
\end{pmatrix} = \begin{pmatrix}
-1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1
\end{pmatrix}.
\] (1.72)

Thus one can think of the rows of the matrix on the right as being labelled by the index $\mu$ and the columns being labelled by the index $\nu$.

At this point, it is convenient again to introduce the Einstein Summation Convention, now for four-dimensional spacetime indices. This makes the writing of expressions such as (1.71) much less cumbersome. The summation convention works as follows:

In an expression such as (1.71), if an index appears exactly twice in a term, then it will be understood that the index is summed over the natural index range (0, 1, 2, 3 in our present case), and the explicit summation symbol will be omitted. An index that occurs twice in a term, thus is understood to be summed over, is called a Dummy Index.

Since in (1.71) both $\mu$ and $\nu$ occur exactly twice, we can rewrite the expression, using the Einstein summation convention, as simply

\[
x^2 + y^2 + z^2 - t^2 = \eta_{\mu\nu} x^\mu x^\nu.
\] (1.73)

On might at first think there would be a great potential for ambiguity, but this is not the case. The point is that in any valid vectorial (or, more generally, tensorial) expression, the only time that a particular index can ever occur exactly twice in a term is when it is summed over. Thus, there is no ambiguity resulting from agreeing to omit the explicit summation symbol, since it is logically inevitable that a summation is intended.\footnote{As a side remark, it should be noted that in a valid vectorial or tensorial expression, a specific index can NEVER appear more than twice in a given term. If you have written down a term where a given index occurs 3, 4 or more times then there is no need to look further at it; it is WRONG. Thus, for example, it is totally meaningless to write $\eta_{\mu\mu} x^\mu x^\mu$. If you ever find such an expression in a calculation then you must stop, and go back to find the place where an error was made.} Note that the pair of dummy indices will always occur with one index upstairs and the other downstairs, in any valid expression.

Now let us return to the Lorentz transformations. The pure boosts written in (1.32), being linear in the space and time coordinates, can be written in the form

\[
x^\mu = \Lambda^\mu_{\nu} x^\nu,
\] (1.74)
where $\Lambda_{\mu\nu}$ are constants, and the Einstein summation convention is operative for the dummy index $\nu$. By comparing (1.74) carefully with (1.32), we can see that the components $\Lambda_{\mu\nu}$ are given by

$$
\begin{align*}
\Lambda^{0\,0} &= \gamma, \\
\Lambda^{0\,i} &= -\gamma v_i, \\
\Lambda^{i\,0} &= -\gamma v_i, \\
\Lambda^{ij} &= \delta^{ij} + \frac{\gamma - 1}{v^2} v_i v_j,
\end{align*}
\tag{1.75}
$$

where $\delta_{ij}$ is the Kronecker delta symbol,

$$
\delta_{ij} = 1 \text{ if } i = j, \quad \delta_{ij} = 0 \text{ if } i \neq j.
\tag{1.76}
$$

A couple of points need to be explained here. Firstly, we are introducing Latin indices here, namely the $i$ and $j$ indices, which range only over the three spatial index values, $i = 1$, 2 and 3. Thus the 4-index $\mu$ can be viewed as $\mu = (0, i)$, where $i = 1$, 2 and 3. This piece of notation is useful because the three spatial index values always occur on a completely symmetric footing, whereas the time index value $\mu = 0$ is a bit different. This can be seen, for example, in the definition of $\eta_{\mu\nu}$ in (1.72) or (1.69).

The second point is that when we consider spatial indices (for example when $\mu$ takes the values $i = 1$, 2 or 3), it actually makes no difference whether we write the index $i$ upstairs or downstairs. Sometimes, as in (1.75), it will be convenient to be rather relaxed about whether we put spatial indices upstairs or downstairs. By contrast, when the index takes the value 0, it is very important to be careful about whether it is upstairs or downstairs. The reason why we can be cavalier about the Latin indices, but not the Greek, will become clearer as we proceed.

We already saw that the Lorentz boost transformations (1.32), re-expressed in terms of $\Lambda_{\mu\nu}$ in (1.75), have the property that $\eta_{\mu\nu} x^\mu x^\nu = \eta_{\mu\nu} (x^\mu x^\nu)$. Thus from (1.74) we have

$$
\eta_{\mu\nu} x^\mu x^\nu = \eta_{\mu\nu} \Lambda^\rho_\mu \Lambda^\sigma_\nu x^\rho x^\sigma.
\tag{1.77}
$$

(Note that we have been careful to choose two different dummy indices for the two implicit summations over $\rho$ and $\sigma$!) On the left-hand side, we can replace the dummy indices $\mu$ and $\nu$ by $\rho$ and $\sigma$, and thus write

$$
\eta_{\rho\sigma} x^\rho x^\sigma = \eta_{\mu\nu} \Lambda^\rho_\mu \Lambda^\nu_\sigma x^\rho x^\sigma.
\tag{1.78}
$$

This can be grouped together as

$$
(\eta_{\rho\sigma} - \eta_{\mu\nu} \Lambda^\rho_\mu \Lambda^\nu_\sigma)x^\rho x^\sigma = 0,
\tag{1.79}
$$
and, since it is true for any $x^\mu$, we must have that

$$\eta_{\mu\nu} \Lambda^\mu_\rho \Lambda^\nu_\sigma = \eta_{\rho\sigma}. \tag{1.80}$$

(This can also be verified directly from (1.75).) The full set of $\Lambda$'s that satisfy (1.80) are the Lorentz Transformations. The Lorentz Boosts, given by (1.75), are examples, but they are just a subset of the full set of Lorentz transformations that satisfy (1.80). Essentially, the additional Lorentz transformations consist of rotations of the three-dimensional spatial coordinates. Thus, one can really say that the Lorentz boosts (1.75) are the “interesting” Lorentz transformations, i.e. the ones that rotate space and time into one another. The remainder are just rotations of our familiar old 3-dimensional Euclidean space.

We can count the number of independent parameters in a general Lorentz transformation in the same way we did for the 3-dimensional rotations in the previous section. We start with $\Lambda^\mu_\nu$, which can be thought of as a $4 \times 4$ matrix with rows labelled by $\mu$ and columns labelled by $\nu$. Thus

$$\Lambda^\mu_\nu \rightarrow \Lambda = \begin{pmatrix} A^0_0 & A^0_1 & A^0_2 & A^0_3 \\ A^1_0 & A^1_1 & A^1_2 & A^1_3 \\ A^2_0 & A^2_1 & A^2_2 & A^2_3 \\ A^3_0 & A^3_1 & A^3_2 & A^3_3 \end{pmatrix}. \tag{1.81}$$

These $4 \times 4 = 16$ components are subject to the conditions (1.80). In matrix notation, (1.80) clearly translates into

$$\Lambda^T \eta \Lambda - \eta = 0. \tag{1.82}$$

This is itself a $4 \times 4$ matrix equation, but not all its components are independent since the left-hand side is a symmetric matrix. (Verify this by taking its transpose.) Thus (1.82) contains $(4 \times 5)/2 = 10$ independent conditions, implying that the most general Lorentz transformation has

$$16 - 10 = 6 \quad \tag{1.83}$$

independent parameters.

Notice that if $\eta$ had been simply the $4 \times 4$ unit matrix, then (1.82) would have been a direct 4-dimensional analogue of the 3-dimensional orthogonality condition (1.58). In other words, were it not for the minus sign in the 00 component of $\eta$, the Lorentz transformations would just be spatial rotations in 4 dimensions, and they would be elements of the group $O(4)$. The counting of the number of independent such transformations would be identical to the one given above, and so the group $O(4)$ of orthogonal $4 \times 4$ matrices is characterised by 6 independent parameters.
Because of the minus sign in $\eta$, the group of $4 \times 4$ matrices satisfying (1.82) is called $O(1, 3)$, with the numbers 1 and 3 indicating the number of time and space dimensions respectively. Thus the four-dimensional Lorentz Group is $O(1, 3)$.

Obviously, the subset of $\Lambda$ matrices of the form

$$\Lambda = \begin{pmatrix} 1 & 0 \\ 0 & M \end{pmatrix},$$

which is shorthand for $\Lambda = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & M_{11} & M_{12} & M_{13} \\ 0 & M_{21} & M_{22} & M_{23} \\ 0 & M_{31} & M_{32} & M_{33} \end{pmatrix}$, (1.84)

where $M$ is any $3 \times 3$ orthogonal matrix, satisfies (1.82). This $O(3)$ subgroup of the $O(1, 3)$ Lorentz group describes the pure rotations (and reflections) in the 3-dimensional spatial directions. The 3 parameters characterising these transformations, together with the 3 parameters of the velocity vector characterising the pure boost Lorentz transformations (1.75), comprise the total set of $3 + 3 = 6$ parameters of the general Lorentz transformations.

It is useful to note that any Lorentz transformation $\Lambda^{\mu\nu}$ can be decomposed into a product of a pure Lorentz boost and a pure spatial rotation. Thus we can write a general Lorentz transformation $\Lambda^{\mu\nu}$ in the form

$$\Lambda^{\mu\nu} = \Lambda^{\mu\rho}(R) \Lambda^{\rho\nu}(B),$$

where $\Lambda^{\rho\nu}(B)$ denotes a pure boost, of the form (1.75), and $\Lambda^{\mu\rho}(R)$ denotes a pure spatial rotation, of the form (1.84).

The decomposition given in (1.85) has been organised in the form of a pure Lorentz boost, followed by a pure spatial rotation. One could instead make a decomposition of $\Lambda^{\mu\nu}$ in the opposite order, as a pure spatial rotation followed by a pure Lorentz boost:

$$\Lambda^{\mu\nu} = \tilde{\Lambda}^{\mu\rho}(B) \tilde{\Lambda}^{\rho\nu}(R),$$

(1.86)

Note that the pure boost and pure rotation transformations will, in general, differ from those in the previous decomposition (1.85), which is why they are written with tildes in (1.86). In other words, the pure boost and the pure spatial rotation matrices do not commute in general.

An example of a decomposition into boost times rotation appears in homework 1, where you are asked to re-express the composition of a pure boost along $x$ followed by a pure boost along $y$ in the form (1.85).

The coordinates $x^\mu = (x^0, x^1)$ live in a four-dimensional spacetime, known as Minkowski Spacetime. This is the four-dimensional analogue of the three-dimensional Euclidean Space.
described by the Cartesian coordinates \( x^i = (x, y, z) \). The quantity \( \eta_{\mu\nu} \) is called the \textit{Minkowski Metric}, and for reasons that we shall see presently, it is called a \textit{tensor}. It is called a metric because it provides the rule for measuring distances in the four-dimensional Minkowski spacetime. The distance, or to be more precise, the \textit{interval}, between two infinitesimally-separated points \((x^0, x^1, x^2, x^3)\) and \((x^0 + dx^0, x^1 + dx^1, x^2 + dx^2, x^3 + dx^3)\) in spacetime is written as \( ds \), and is given by

\[
ds^2 = \eta_{\mu\nu} dx^\mu dx^\nu.
\] (1.87)

Clearly, this is the Minkowskian generalisation of the three-dimensional distance \( ds_E \) between neighbouring points \((x, y, z)\) and \((x + dx, y + dy, z + dz)\) in Euclidean space, which, by Pythagoras’ theorem, is given by

\[
ds_E^2 = dx^2 + dy^2 + dz^2 = \delta_{ij} dx^i dx^j.
\] (1.88)

The Euclidean metric (1.88) is invariant under arbitrary constant rotations of the \((x, y, z)\) coordinate system. (This is clearly true because the distance between the neighbouring points must obviously be independent of how the axes of the Cartesian coordinate system are oriented.) By the same token, the Minkowski metric (1.87) is invariant under arbitrary Lorentz transformations. In other words, as can be seen to follow immediately from (1.80), the spacetime interval \( ds'^2 = \eta_{\mu\nu} dx'^\mu dx'^\nu \) calculated in the primed frame is identical to the interval \( ds^2 \) calculated in the unprimed frame

\[
ds'^2 = \eta_{\mu\nu} dx'^\mu dx'^\nu = \eta_{\mu\nu} \Lambda^\mu_\rho \Lambda^\nu_\sigma dx^\rho dx^\sigma,
\]

\[
= \eta_{\rho\sigma} dx^\rho dx^\sigma = ds^2.
\] (1.89)

For this reason, we do not need to distinguish between \( ds^2 \) and \( ds'^2 \), since it is the same in all inertial frames. It is what is called a \textit{Lorentz Scalar}.

The Lorentz transformation rule of the coordinate differential \( dx^\mu \), i.e.

\[
dx^\mu = \Lambda^\mu_\nu dx^\nu,
\] (1.90)

can be taken as the prototype for more general 4-vectors. Thus, we may define any set of four quantities \( U^\mu \), for \( \mu = 0, 1, 2 \) and \( 3 \), to be the components of a Lorentz 4-vector (often, we shall just abbreviate this to simply a 4-vector) if they transform, under Lorentz transformations, according to the rule

\[
U'^\mu = \Lambda^\mu_\nu U^\nu.
\] (1.91)
The Minkowski metric $\eta_{\mu \nu}$ may be thought of as a $4 \times 4$ matrix, whose rows are labelled by $\mu$ and columns labelled by $\nu$, as in (1.72). Clearly, the inverse of this matrix takes the same form as the matrix itself. We denote the components of the inverse matrix by $\eta^{\mu \nu}$. This is called, not surprisingly, the inverse Minkowski metric. Clearly it satisfies the relation

$$ \eta_{\mu \nu} \eta^{\nu \rho} = \delta_\rho^\mu, $$

(1.92)

where the 4-dimensional Kronecker delta is defined to equal 1 if $\mu = \rho$, and to equal 0 if $\mu \neq \rho$. Note that like $\eta_{\mu \nu}$, the inverse $\eta^{\mu \nu}$ is symmetric also: $\eta^{\mu \nu} = \eta^{\nu \mu}$.

The Minkowski metric and its inverse may be used to lower or raise the indices on other quantities. Thus, for example, if $U^\mu$ are the components of a 4-vector, then we may define

$$ U_\mu = \eta_{\mu \nu} U^\nu. $$

(1.93)

This is another type of 4-vector. Two distinguish the two, we call a 4-vector with an upstairs index a contravariant 4-vector, while one with a downstairs index is called a covariant 4-vector. Note that if we raise the lowered index in (1.93) again using $\eta^{\mu \nu}$, then we get back to the starting point:

$$ \eta^{\mu \nu} U_\nu = \eta^{\mu \nu} \eta_{\rho \nu} U^\rho = \delta_\rho^\mu U^\rho = U^\mu. $$

(1.94)

It is for this reason that we can use the same symbol $U$ for the covariant 4-vector $U_\mu = \eta_{\mu \nu} U^\nu$ as we used for the contravariant 4-vector $U^\mu$.

In a similar fashion, we may define the quantities $\Lambda_\mu^\nu$ by

$$ \Lambda_\mu^\nu = \eta_{\mu \rho} \eta^{\nu \sigma} \Lambda_\rho^\sigma. $$

(1.95)

It is then clear that (1.80) can be restated as

$$ \Lambda_\mu^\nu \Lambda_\mu^\rho = \delta_\nu^\rho. $$

(1.96)

Notice two points concerning raising and lowering indices with $\eta$. The first is that if we have a vector-valued or tensor-valued equation, such as $A^\mu = B^\mu$, or $S_{\mu \nu} = T_{\mu \nu}$ or whatever, we can raise or lower these free indices at will, as long as we raise them on both sides of the equation at the same time. Thus, for example,

$$ S_{\mu \nu} = T_{\mu \nu} \iff S_\mu^\nu = T_\mu^\nu \iff S^\mu_\nu = T^\mu_\nu \iff S^{\mu \nu} = T^{\mu \nu}. $$

(1.97)

The second point is that if there is an index contraction in a term, we can freely “see-saw” the pair of dummy indices, moving the upper index down and simultaneously the lower index up. Thus, for example,

$$ A^\mu B_\mu = A_\mu B^\mu, $$

(1.98)
and so on. The reason for emphasising these two points is just to make clear that one does not need to make a song and dance about raising or lowering free indices, or see-sawing dummy index positions. After any such operations, a valid covariant equation will always have the properties that every matching free index will be in the same location (upstairs or downstairs) in every term in the equation. Furthermore, every dummy index pair will always have one occurrence of that index upstairs, and the other downstairs.\(^8\)

We can invert the Lorentz transformation \(x'\mu = \Lambda^\mu_\nu x^\nu\), by multiplying both sides by \(\Lambda_\mu^\rho\) and using (1.96) to give \(x^\mu \Lambda_\mu^\rho = \delta_\nu^\rho x^\nu = x^\rho\), and hence, after relabelling,

\[
x^\mu = \Lambda^\nu_\mu x'^\nu.
\]

(1.99)

It now follows from (1.91) that the components of the covariant 4-vector \(U_\mu\) defined by (1.93) transform under Lorentz transformations according to the rule

\[
U'_\mu = \Lambda_\mu^\nu U_\nu.
\]

(1.100)

Any set of 4 quantities \(U_\mu\) which transform in this way under Lorentz transformations will be called a covariant 4-vector.

Using (1.99), we can see that the gradient operator \(\partial / \partial x^\mu\) transforms as a covariant 4-vector. Using the chain rule for partial differentiation we have

\[
\frac{\partial}{\partial x'^\mu} = \frac{\partial x'^\nu}{\partial x^\mu} \frac{\partial}{\partial x^\nu}.
\]

(1.101)

But from (1.99) we have (after a relabelling of indices) that

\[
\frac{\partial x'^\nu}{\partial x^\mu} = \Lambda^\nu_\mu,
\]

(1.102)

and hence (1.101) gives

\[
\frac{\partial}{\partial x'^\mu} = \Lambda^\nu_\mu \frac{\partial}{\partial x^\nu}.
\]

(1.103)

As can be seen from (1.100), this is precisely the transformation rule for a a covariant 4-vector. The gradient operator arises sufficiently often that it is useful to use a special\(^8\)These statements apply to equations written in the four-dimensionally covariant language, with Greek indices \(\mu, \nu, \ldots\) ranging over 0, 1, 2 and 3. As has already been emphasised, if one decomposes a four-dimensionally covariant expression into the 1+3 language of time plus three spatial diections (denoted by the Latin spatial indices \(i, j, \ldots\)), then one is completely free to write the Latin indices upstairs or downstairs, unmatched between different terms. In the context of our Minkowski spacetime discussions in this course, the only reason for caring about the distinction between upstairs and downstairs indices is because of the time direction. Thus we must respect the upstairs/downstairs rules for four-dimensional covariance, but it is unimportant in the 1+3 language we use for purely three-dimensional rotational covariance.
symbol to denote it. We therefore define

\[ \partial_\mu \equiv \frac{\partial}{\partial x^\mu}. \]  

(1.104)

Thus the Lorentz transformation rule (1.103) is now written as

\[ \partial'_\mu = \Lambda_\mu^\nu \partial_\nu. \]  

(1.105)

### 1.5 Lorentz Tensors

Having seen how contravariant and covariant 4-vectors transform under Lorentz transformations (as given in (1.91) and (1.100) respectively), we can now define the transformation rules for more general objects called tensors. These objects carry multiple indices, and each one transforms with a \( \Lambda \) factor, of either the (1.91) type if the index is upstairs, or of the (1.100) type if the index is downstairs. Thus, for example, a tensor \( T_{\mu\nu} \) transforms under Lorentz transformations according to the rule

\[ T'_{\mu\nu} = \Lambda_\mu^\rho \Lambda_\nu^\sigma T_{\rho\sigma}. \]  

(1.106)

More generally, a tensor \( T^{\mu_1 \cdots \mu_m \nu_1 \cdots \nu_n} \) will transform according to the rule

\[ T'^{\mu_1 \cdots \mu_m \nu_1 \cdots \nu_n} = \Lambda^{\mu_1}_{\rho_1} \cdots \Lambda^{\mu_m}_{\rho_m} \Lambda_{\nu_1}^{\sigma_1} \cdots \Lambda_{\nu_n}^{\sigma_n} T_{\rho_1 \cdots \rho_m \sigma_1 \cdots \sigma_n}. \]  

(1.107)

Note that scalars are just special cases of tensors with no indices, while vectors are special cases with just one index.

It is easy to see that products of tensors give rise again to tensors. For example, if \( U^\mu \) and \( V^\nu \) are two contravariant vectors then \( T^{\mu\nu} \equiv U^\mu V^\nu \) is a tensor, since, using the known transformation rules for \( U \) and \( V \) we have

\[ T'^{\mu\nu} = U'^{\mu} V'^{\nu} = \Lambda^{\mu}_{\rho} U^{\rho} \Lambda^{\nu}_{\sigma} V^{\sigma}, \]

\[ = \Lambda^{\mu}_{\rho} \Lambda^{\nu}_{\sigma} T^{\rho\sigma}. \]  

(1.108)

Note that the gradient operator \( \partial_\mu \) can also be used to map a tensor into another tensor. For example, if \( U_\mu \) is a vector field (i.e. a vector that changes from place to place in spacetime) then \( S_{\mu\nu} \equiv \partial_\mu U_\nu \) is a tensor field. As always, the way to check that something is a tensor is to check that it transforms in the proper way under Lorentz transformations. So in this case, one needs to check that it transforms in the way an \((m, n) = (0, 2)\) tensor in eqn (1.107) does.

We make also define the operation of **Contraction**, which reduces a tensor to one with a smaller number of indices. A contraction is performed by setting an upstairs index on a
tensor equal to a downstairs index. The Einstein summation convention then automatically comes into play, and the result is that one has an object with one fewer upstairs indices and one fewer downstairs indices. Furthermore, a simple calculation shows that the new object is itself a tensor. Consider, for example, a tensor $T^\mu{}_{\nu}$. This, of course, transforms as

$$
T'{}^\mu{}_{\mu} = \Lambda^\mu{}_{\rho} \Lambda\nu{}^{\sigma} T^{\rho}{}_{\sigma}
$$

(1.109)

under Lorentz transformations. If we form the contraction and define $\phi \equiv T^\mu{}_{\mu}$, then we see that under Lorentz transformations we shall have

$$
\phi' \equiv T'{}^{\mu}{}_{\mu} = \Lambda^\mu{}_{\rho} \Lambda\nu{}^{\sigma} T^{\rho}{}_{\sigma},
$$

$$
= \delta^{\sigma}_{\rho} T^{\rho}{}_{\sigma} = \phi.
$$

(1.110)

Since $\phi' = \phi$, it follows, by definition, that $\phi$ is a scalar.

An essentially identical calculation shows that for a tensor with arbitrary numbers of upstairs and downstairs indices, if one makes an index contraction of one upstairs with one downstairs index, the result is a tensor with the corresponding reduced numbers of indices. Of course multiple contractions work in the same way.

The Minkowski metric $\eta_{\mu\nu}$ is itself a tensor, but of a rather special type, known as an invariant tensor. This is because, unlike a generic 2-index tensor, the Minkowski metric is identical in all Lorentz frames. To see this, let us first write out how it would transform under Lorentz transformations, using the usual transformation rules in (1.107):

$$
\eta'{}_{\mu\nu} = \Lambda^\mu{}_{\rho} \Lambda\nu{}^{\sigma} \eta_{\rho\sigma}
$$

(1.111)

Our goal is to show that in fact $\eta'{}_{\mu\nu} = \eta_{\mu\nu}$, i.e. that it is actually invariant under Lorentz transformations. Now although the right-hand side of (1.111) looks reminiscent of what one has in (1.80), which is the defining property of the $\Lambda^\mu{}_{\nu}$ Lorentz transformations, it is not the same. Specifically, in (1.111) the indices of the two $\Lambda$ transformations are contracted with $\eta$ on their second indices, rather than on the first indices as in (1.80). We can easily work out what the right-hand side of (1.111) is by going through the following steps. First, we rewrite (1.80) in matrix language as $\Lambda^T \eta \Lambda = \eta$. Then right-multiply by $\Lambda^{-1}$ and left-multiply by $\eta^{-1}$; this gives $\eta^{-1} \Lambda^T \eta = \Lambda^{-1}$. Next left-multiply by $\Lambda$ and right-multiply by $\eta^{-1}$, which gives $\Lambda \eta^{-1} \Lambda^T = \eta^{-1}$. (This is the analogue for the Lorentz transformations of the proof, for ordinary orthogonal matrices, that $M^T M = 1$ implies $M M^T = 1$.) Converting back to index notation gives $\Lambda^\mu{}_{\rho} \Lambda^\nu{}_{\sigma} \eta^{\rho\sigma} = \eta^{\mu\nu}$. After some index raising and lowering, this gives

$$
\Lambda^\mu{}_{\rho} \Lambda^\nu{}_{\sigma} \eta_{\rho\sigma} = \eta_{\mu\nu}.
$$

(1.112)
Applying this result to the right-hand side of (1.111) therefore gives the desired result,

\[ \eta'_{\mu\nu} = \eta_{\mu\nu}. \]  

(1.113)

Thus we have shown that the tensor \( \eta_{\mu\nu} \) is actually invariant under Lorentz transformations. The same is also true for the inverse metric \( \eta^{\mu\nu} \).

We already saw that the gradient operator \( \partial_{\mu} \equiv \partial / \partial x^{\mu} \) transforms as a covariant vector. If we define, in the standard way, \( \partial^{\mu} \equiv \eta^{\mu\nu} \partial_{\nu} \), then it is evident from what we have seen above that the operator

\[ \square \equiv \partial^{\mu} \partial_{\mu} = \eta^{\mu\nu} \partial_{\mu} \partial_{\nu} \]  

(1.114)

transforms as a scalar under Lorentz transformations. This is a very important operator, which is otherwise known as the wave operator, or d’Alembertian:

\[ \square = -\partial_{0} \partial_{0} + \partial_{i} \partial_{i} = -\partial^{2} / \partial t^{2} + \partial^{2} / \partial x^{2} + \partial^{2} / \partial y^{2} + \partial^{2} / \partial z^{2}. \]  

(1.115)

It is worth commenting further at this stage about a remark that was made earlier. Notice that in (1.115) we have been cavalier about the location of the Latin indices, which of course range only over the three spatial directions \( i = 1, 2, 3 \). We can get away with this because the metric that is used to raise or lower the Latin indices is just the Minkowski metric restricted to the index values 1, 2, and 3. But since we have

\[ \eta_{00} = -1, \quad \eta_{ij} = \delta_{ij}, \quad \eta_{0i} = \eta_{i0} = 0, \]  

(1.116)

this means that Latin indices are lowered and raised using the Kronecker delta \( \delta_{ij} \) and its inverse \( \delta^{ij} \). But these are just the components of the unit matrix, and so raising or lowering Latin indices has no effect. It is because of the minus sign associated with the \( \eta_{00} \) component of the Minkowski metric that we have to pay careful attention to the process of raising and lowering Greek indices. Thus, we can get away with writing \( \partial_{i} \partial_{i} \), but we cannot write \( \partial_{\mu} \partial_{\mu} \).

### 1.6 Proper time and 4-velocity

We defined the Lorentz-invariant interval \( ds \) between infinitesimally-separated spacetime events by

\[ ds^{2} = \eta_{\mu\nu} dx^{\mu} dx^{\nu} = -dt^{2} + dx^{2} + dy^{2} + dz^{2}. \]  

(1.117)

This is the Minkowskian generalisation of the spatial interval in Euclidean space. Note that \( ds^{2} \) can be positive, negative or zero. These cases correspond to what are called spacelike, timelike or null separations, respectively.
On occasion, it is useful to define the negative of $ds^2$, and write

$$d\tau^2 = -ds^2 = -\eta_{\mu\nu} dx^\mu dx^\nu = dt^2 - dx^2 - dy^2 - dz^2. \quad (1.118)$$

This is called the *Proper Time* interval, and $\tau$ is the proper time. Since $ds$ is a Lorentz scalar, it is obvious that $d\tau$ is a scalar too.

We know that $dx^\mu$ transforms as a contravariant 4-vector. Since $d\tau$ is a scalar, it follows that

$$U^\mu \equiv \frac{dx^\mu}{d\tau} \quad (1.119)$$

is a contravariant 4-vector also. If we think of a particle following a path, or *worldline* in spacetime parameterised by the proper time $\tau$, i.e. it follows the path $x^\mu = x^\mu(\tau)$, then $U^\mu$ defined in (1.119) is called the 4-velocity of the particle.

It is useful to see how the 4-velocity is related to the usual notion of 3-velocity of a particle. By definition, the 3-velocity $\vec{u}$ is a 3-vector with components $u^i$ given by

$$u^i = \frac{dx^i}{dt}. \quad (1.120)$$

From (1.118), it follows that

$$d\tau^2 = dt^2[1 - (dx/dt)^2 - (dy/dt)^2 - (dz/dt)^2] = dt^2(1 - u^2), \quad (1.121)$$

where $u = |\vec{u}|$, or in other words, $u = \sqrt{u^i u^i}$. In view of the definition of the $\gamma$ factor in (1.31), it is natural to define

$$\gamma \equiv \frac{1}{\sqrt{1 - u^2}}. \quad (1.122)$$

Thus we have $d\tau = dt/\gamma$, and so from (1.119) the 4-velocity can be written as

$$U^\mu = \frac{dt}{d\tau} \frac{dx^\mu}{dt} = \gamma \frac{dx^\mu}{dt}. \quad (1.123)$$

Since $dx^0/dt = 1$ and $dx^i/dt = u^i$, we therefore have that

$$U^0 = \gamma, \quad U^i = \gamma u^i. \quad (1.124)$$

Note that $U^\mu U_\mu = -1$, since, from (1.118), we have

$$U^\mu U_\mu = \eta_{\mu\nu} U^\mu U^\nu = \frac{\eta_{\mu\nu} dx^\mu dx^\nu}{(d\tau)^2} = \frac{-(d\tau)^2}{(d\tau)^2} = -1. \quad (1.125)$$

We shall sometimes find it convenient to rewrite (1.124) as

$$U^\mu = (\gamma, \gamma u^i) \quad \text{or} \quad U^\mu = (\gamma, \gamma \vec{u}). \quad (1.126)$$
Having set up the 4-vector formalism, it is now completely straightforward write down how velocities transform under Lorentz transformations. We know that the 4-velocity \( U^\mu \) will transform according to (1.91), and this is identical to the way that the coordinates \( x^\mu \) transform:

\[
U'^\mu = \Lambda^{\mu \nu} U^\nu, \quad x'^\mu = \Lambda^{\mu \nu} x^\nu.
\] (1.127)

Therefore, if we want to know how the 3-velocity transforms, we need only write down the Lorentz transformations for \((t, x, y, z)\), and then replace \((t, x, y, z)\) by \((U^0, U^1, U^2, U^3)\). Finally, using (1.126) to express \((U^0, U^1, U^2, U^3)\) in terms of \(\vec{u}\) will give the result.

Consider, for simplicity, the case where \(S'\) is moving along the \(x\) axis with velocity \(v\). The Lorentz transformation for \(U^\mu\) can therefore be read off from (1.30) and (1.31):

\[
\begin{align*}
U'^0 &= \gamma_v (U^0 - vU^1), \\
U'^1 &= \gamma_v (U^1 - vU^0), \\
U'^2 &= U^2, \\
U'^3 &= U^3,
\end{align*}
\] (1.128)

where we are now using \(\gamma_v \equiv (1 - v^2)^{-1/2}\) to denote the gamma factor of the Lorentz transformation, to distinguish it from the \(\gamma\) constructed from the 3-velocity \(\vec{u}\) of the particle in the frame \(S\), which is defined in (1.122). Thus from (1.126) we have

\[
\begin{align*}
\gamma' &= \gamma \gamma_v (1 - vu_x), \\
\gamma' u'_x &= \gamma \gamma_v (u_x - v), \\
\gamma' u'_y &= \gamma u_y, \\
\gamma' u'_z &= \gamma u_z,
\end{align*}
\] (1.129)

where, of course, \(\gamma' = (1 - u'^2)^{-1/2}\) is the analogue of \(\gamma\) in the frame \(S'\). Thus we find

\[
\begin{align*}
u'_x &= \frac{u_x - v}{1 - v u_x}, \\
u'_y &= \frac{u_y}{\gamma_v (1 - v u_x)}, \\
u'_z &= \frac{u_z}{\gamma_v (1 - v u_x)}.
\end{align*}
\] (1.130)

## 2 Electrodynamics and Maxwell’s Equations

### 2.1 Natural units

We saw earlier that the supposition of the universal validity of Maxwell’s equations in all inertial frames, which in particular would imply that the speed of light should be the same in all frames, is consistent with experiment. It is therefore reasonable to expect that Maxwell’s
equations should be compatible with special relativity. However, written in their standard form (1.7), this compatibility is by no means apparent. Our next task will be to re-express the Maxwell equations, in terms of 4-tensors, in a way that makes their Lorentz covariance manifest.

We shall begin by changing units from the S.I. system in which the Maxwell equations are given in (1.7). The first step is to change to Gaussian units, by performing the rescalings

\[ E \rightarrow \frac{1}{\sqrt{4\pi\varepsilon_0}} \tilde{E}, \quad B \rightarrow \sqrt{\frac{\mu_0}{4\pi}} \tilde{B}, \]
\[ \rho \rightarrow \sqrt{4\pi\varepsilon_0} \rho, \quad J \rightarrow \sqrt{4\pi\varepsilon_0} \tilde{J}. \] (2.1)

Bearing in mind that the speed of light is given by \( c = \frac{1}{\sqrt{\mu_0\varepsilon_0}} \), we see that the Maxwell equations (1.7) become

\[ \nabla \cdot \tilde{E} = 4\pi \rho, \quad \nabla \times \tilde{B} - \frac{1}{c} \frac{\partial \tilde{E}}{\partial t} = \frac{4\pi}{c} \tilde{J}, \]
\[ \nabla \cdot \tilde{B} = 0, \quad \nabla \times \tilde{E} + \frac{1}{c} \frac{\partial \tilde{B}}{\partial t} = 0, \] (2.2)

Finally, we pass from Gaussian units to Natural units, by choosing our units of length and time so that \( c = 1 \), as we did in our discussion of special relativity. Thus, in natural units, the Maxwell equations become

\[ \nabla \cdot \tilde{E} = 4\pi \rho, \quad \nabla \times \tilde{B} - \frac{\partial \tilde{E}}{\partial t} = 4\pi \tilde{J}, \] (2.3)
\[ \nabla \cdot \tilde{B} = 0, \quad \nabla \times \tilde{E} + \frac{\partial \tilde{B}}{\partial t} = 0, \] (2.4)

The equations (2.3), which have sources on the right-hand side, are called the Field Equations. The equations (2.4) are called Bianchi Identities. We shall elaborate on this a little later.

### 2.2 Gauge potentials and gauge invariance

We already remarked that the two Maxwell equations (2.4) are known as Bianchi identities. They are not field equations, since there are no sources; rather, they impose constraints on the electric and magnetic fields. The first equation in (2.4), i.e. \( \nabla \cdot \tilde{B} = 0 \), can be solved by writing

\[ \tilde{B} = \nabla \times \tilde{A}, \] (2.5)

where \( \tilde{A} \) is the magnetic 3-vector potential. Note that (2.5) \textit{identically} solves \( \nabla \cdot \tilde{B} = 0 \), because of the vector identity that \( \text{div curl} \equiv 0 \). Substituting (2.5) into the second equation...
in (2.4), we obtain
\[ \vec{\nabla} \times \left( \vec{E} + \frac{\partial \vec{A}}{\partial t} \right) = 0. \tag{2.6} \]
This can be solved, again identically, by writing
\[ \vec{E} + \frac{\partial \vec{A}}{\partial t} = -\vec{\nabla} \phi, \tag{2.7} \]
where \( \phi \) is the electric scalar potential. Thus we can solve the Bianchi identities (2.4) by writing \( \vec{E} \) and \( \vec{B} \) in terms of scalar and 3-vector potentials \( \phi \) and \( \vec{A} \):
\[ \vec{E} = -\vec{\nabla} \phi - \frac{\partial \vec{A}}{\partial t}, \quad \vec{B} = \vec{\nabla} \times \vec{A}. \tag{2.8} \]

Although we have now “disposed of” the two Maxwell equations in (2.4), it has been achieved at a price, in that there is a redundancy in the choice of gauge potentials \( \phi \) and \( \vec{A} \). First, we may note that that \( \vec{B} \) in (2.8) is unchanged if we make the replacement
\[ \vec{A} \rightarrow \vec{A} + \vec{\nabla} \lambda, \tag{2.9} \]
where \( \lambda \) is an arbitrary function of position and time. The expression for \( \vec{E} \) will also be invariant, if we simultaneously make the replacement
\[ \phi \rightarrow \phi - \frac{\partial \lambda}{\partial t}. \tag{2.10} \]

To summarise, if a given set of electric and magnetic fields \( \vec{E} \) and \( \vec{B} \) are described by a scalar potential \( \phi \) and 3-vector potential \( \vec{A} \) according to (2.8), then the identical physical situation (i.e. identical electric and magnetic fields) is equally well described by a new pair of scalar and 3-vector potentials, related to the original pair by the Gauge Transformations given in (2.9) and (2.10), where \( \lambda \) is an arbitrary function of position and time.

We can in fact use the gauge invariance to our advantage, by making a convenient and simplifying gauge choice for the scalar and 3-vector potentials. We have one arbitrary function (i.e. \( \lambda(t, \vec{r}) \)) at our disposal, and so this allows us to impose one functional relation on the potentials \( \phi \) and \( \vec{A} \). For our present purposes, the most useful gauge choice is to use this freedom to impose the Lorenz gauge condition,
\[ \vec{\nabla} \cdot \vec{A} + \frac{\partial \phi}{\partial t} = 0. \tag{2.11} \]
Note that, contrary to the belief of many physicists, this gauge choice was introduced by the Danish physicist Ludvig Valentin Lorenz, and not the Dutch physicist Hendrik Antoon Lorentz who is responsible for the Lorentz transformation. Adding to the confusion is that
Unlike many other gauge choices that one encounters, the Lorenz gauge condition is, as we shall see later, Lorentz invariant.

Substituting (2.8) into the remaining Maxwell equations (i.e. (2.3), and using the Lorenz gauge condition (2.11), we therefore find

\[
\nabla^2 \phi - \frac{\partial^2 \phi}{\partial t^2} = -4\pi \rho,
\]

\[
\nabla^2 \vec{A} - \frac{\partial^2 \vec{A}}{\partial t^2} = -4\pi \vec{J}.
\]

(2.12)

The important thing, which we shall make use of shortly, is that in each case we have on the left-hand side the d’Alembertian operator \( \Box = \partial^\mu \partial_\mu \), which we discussed earlier.

2.3 Maxwell’s equations in 4-tensor notation

The next step is to write the Maxwell equations in terms of four-dimensional quantities. Since the 3-vectors \( \vec{E} \) and \( \vec{B} \) describing the electric and magnetic fields have three components each, there is clearly no way in which they can be “assembled” into 4-vectors. However, we may note that in four dimensions a two-index antisymmetric tensor has \( (4 \times 3)/2 = 6 \) independent components. Since this is equal to \( 3 + 3 \), it suggests that perhaps we should be grouping the electric and magnetic fields together into a single 2-index antisymmetric
tensor. This is in fact exactly what is needed. Thus we introduce a tensor $F_{\mu\nu}$, satisfying
\[ F_{\mu\nu} = -F_{\nu\mu}. \] (2.13)

It turns out that we should define its components in terms of $\vec{E}$ and $\vec{B}$ as follows:
\[ F_{0i} = -E_i, \quad \text{(which implies} \quad F_{i0} = E_i), \quad F_{ij} = \epsilon_{ijk} B_k. \] (2.14)

Here $\epsilon_{ijk}$ is the usual totally-antisymmetric tensor of 3-dimensional vector calculus. It is equal to $+1$ if $(ijk)$ is an even permutation of $(123)$, to $-1$ if it is an odd permutation, and to zero if it is no permutation (i.e. if two or more of the indices $(ijk)$ are equal). In other words, we have
\[ F_{23} = B_1, \quad F_{31} = B_2, \quad F_{12} = B_3, \]
\[ F_{32} = -B_1, \quad F_{13} = -B_2, \quad F_{21} = -B_3. \] (2.15)

Viewing $F_{\mu\nu}$ as a matrix $F$ with rows labelled by $\mu$ and columns labelled by $\nu$, we shall have
\[
F = \begin{pmatrix}
F_{00} & F_{01} & F_{02} & F_{03} \\
F_{10} & F_{11} & F_{12} & F_{13} \\
F_{20} & F_{21} & F_{22} & F_{23} \\
F_{30} & F_{31} & F_{32} & F_{33}
\end{pmatrix} = \begin{pmatrix}
0 & -E_1 & -E_2 & -E_3 \\
E_1 & 0 & B_3 & -B_2 \\
E_2 & -B_3 & 0 & B_1 \\
E_3 & B_2 & -B_1 & 0
\end{pmatrix}. \] (2.16)

We also need to combine the charge density $\rho$ and the 3-vector current density $\vec{J}$ into a four-dimensional quantity. This is easy; we just define a 4-vector $J^\mu$, whose spatial components $J^i$ are just the usual 3-vector current components, and whose time component $J^0$ is equal to the charge density $\rho$:
\[ J^0 = \rho, \quad J^i = J^i. \] (2.17)

A word of caution is in order here. Although we have defined objects $F_{\mu\nu}$ and $J^\mu$ that have the appearance of a 4-tensor and a 4-vector, we are only entitled to call them such if we have verified that they transform in the proper way under Lorentz transformations. In fact they do, and we shall justify this a little later.

For now, we shall proceed to see how the Maxwell equations look when expressed in terms of $F_{\mu\nu}$ and $J^\mu$. The answer is that they become
\[
\partial_\mu F^{\mu\nu} = -4\pi J^\nu, \quad \partial_\mu F_{\nu\rho} + \partial_\nu F_{\rho\mu} + \partial_\rho F_{\mu\nu} = 0.
\] (2.18) (2.19)
Two very nice things have happened. First of all, the original four Maxwell equations (2.3) and (2.4) have become just two four-dimensional equations; (2.18) is the field equation, and (2.19) is the Bianchi identity. Secondly, the equations are manifestly Lorentz covariant; i.e. they transform tensorially under Lorentz transformations. This means that they keep exactly the same form in all Lorentz frames. If we start with (2.18) and (2.19) in the unprimed frame \( S \), then we know that in the frame \( S' \), related to \( S \) by the Lorentz transformation (1.74), the equations will look identical, except that they will now have primes on all the quantities. Furthermore, we know precisely how the primed quantities are related to the unprimed:

\[
F'_{\mu\nu} = \Lambda_{\mu}^{\rho} \Lambda_{\nu}^{\sigma} F_{\rho\sigma}, \quad J'^{\mu} = \Lambda^{\mu}_{\nu} J^{\nu},
\]  

(2.20)

etc., where \( \Lambda^{\mu}_{\nu} \) describes the Lorentz transformation from the frame \( S \) to the frame \( S' \).

We should first verify that indeed (2.18) and (2.19) are equivalent to the Maxwell equations (2.3) and (2.4). Consider first (2.18). This equation is vector-valued, since it has the free index \( \nu \). Therefore, to reduce it down to three-dimensional equations, we have two cases to consider, namely \( \nu = 0 \) or \( \nu = j \). For \( \nu = 0 \) we have

\[
\partial_{\mu} F^{\mu 0} = \partial_0 F^{00} + \partial_i F^{i0} = \partial_i F^{i0} = -4\pi J^0,
\]

(2.21)

which therefore corresponds (see (2.14) and (2.17)) to

\[
-\partial_i E_i = -4\pi \rho, \quad \text{i.e.} \quad \vec{\nabla} \cdot \vec{E} = 4\pi \rho.
\]

(2.22)

For \( \nu = j \), we shall have

\[
\partial_{\mu} F^{\mu j} = \partial_0 F^{0j} + \partial_i F^{ij} = -4\pi J^j,
\]

(2.23)

which gives

\[
\partial_0 E_j + \epsilon_{ijk} \partial_i B_k = -4\pi J^j.
\]

(2.24)

This is just\(^9\)

\[
-\frac{\partial \vec{E}}{\partial t} + \vec{\nabla} \times \vec{B} = 4\pi \vec{J}.
\]

(2.25)

Thus (2.18) is equivalent to the two Maxwell field equations in (2.3).

Turning now to (2.19), it follows from the antisymmetry (2.13) of \( F_{\mu\nu} \) that the left-hand side is totally antisymmetric in \( (\mu\nu\rho) \) (i.e. it changes sign under any exchange of a pair of indices). Therefore there are two distinct inequivalent assignments of indices, after we make

\(^9\)Recall that the \( i \)’th component of \( \vec{\nabla} \times \vec{V} \) is given by \( (\vec{\nabla} \times \vec{V})_i = \epsilon_{ijk} \partial_j V_k \) for any 3-vector \( \vec{V} \).
the $1 + 3$ decomposition $\mu = (0, i)$ etc.: Either one of the indices is a 0 with the other two Latin, or else all three are Latin. Consider first $(\mu, \nu, \rho) = (0, i, j)$:

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

(2.26)

which, from (2.14), means

$$
\epsilon_{ijk} \frac{\partial B_k}{\partial t} + \partial_i E_j - \partial_j E_i = 0 .
$$

(2.27)

Since this is antisymmetric in $ij$ there is no loss of generality involved in contracting with $\epsilon_{ij\ell}$, which gives$^{10}$

$$
2 \frac{\partial B_\ell}{\partial t} + 2 \epsilon_{ij\ell} \partial_i E_j = 0 .
$$

(2.28)

This is just the statement that

$$
\nabla \times \vec{E} + \frac{\partial \vec{B}}{\partial t} = 0 ,
$$

(2.29)

which is the second of the Maxwell equations in (2.4).

The other distinct possibility for assigning decomposed indices in (2.19) is to take $(\mu, \nu, \rho) = (i, j, k)$, giving

$$
\partial_i F_{jk} + \partial_j F_{ki} + \partial_k F_{ij} = 0 .
$$

(2.30)

Since this is totally antisymmetric in $(i, j, k)$, no generality is lost by contracting it with $\epsilon_{ijk}$, giving

$$
3 \epsilon_{ijk} \partial_i E_j = 0 .
$$

(2.31)

From (2.14), this implies

$$
3 \epsilon_{ijk} \epsilon_{jk\ell} \partial_i B_\ell = 0 , \quad \text{and hence} \quad 6 \partial_i B_i = 0 .
$$

(2.32)

This has just reproduced the first Maxwell equation in (2.4), i.e. $\nabla \cdot \vec{B} = 0$.

We have now demonstrated that the equations (2.18) and (2.19) are equivalent to the four Maxwell equations (2.3) and (2.4). Since (2.18) and (2.19) are written in a four-dimensional notation, it is highly suggestive that they are indeed Lorentz covariant. However, we should be a little more careful, in order to be sure about this point. Not every set of objects $V^\mu$ can be viewed as a Lorentz 4-vector, after all. The test is whether they transform properly, as in (1.91), under Lorentz transformations.

We may begin by considering the quantities $J^\mu = (\rho, J^i)$. Note first that by applying $\partial_\nu$ to the Maxwell field equation (2.18), we get identically zero on the left-hand side, since

$^{10}$Recall that $\epsilon_{ijm} \epsilon_{k\ell m} = \delta_{ik} \delta_{j\ell} - \delta_{ij} \delta_{k\ell}$, and hence $\epsilon_{ijm} \epsilon_{kjm} = 2 \delta_{ik}$. 

35
partial derivatives commute and \( F^{\mu \nu} \) is antisymmetric. Thus from the right-hand side we get

\[
\partial_\mu J^\mu = 0 .
\]  
(2.33)

This is the equation of charge conservation. Decomposed into the \( 3 + 1 \) language, it takes the familiar form

\[
\frac{\partial \rho}{\partial t} + \vec{\nabla} \cdot \vec{J} = 0 .
\]  
(2.34)

By integrating over a closed 3-volume \( V \) and using the divergence theorem on the second term, we learn that the rate of change of charge inside \( V \) is balanced by the flow of charge through its boundary \( S \):

\[
\frac{\partial}{\partial t} \int_V \rho \, d^3x = - \int_S \vec{J} \cdot d\vec{S} ,
\]  
(2.35)

where \( d^3x = dx dy dz \) is the spatial 3-volume element, and \( d\vec{S} = (dydz, dzdx, dx dy) \) is the 2-area element.

Now we are in a position to show that \( J^\mu = (\rho, \vec{J}) \) is indeed a 4-vector. Considering \( J^0 = \rho \) first, we may note that

\[
dQ \equiv \rho \, dx dy dz
\]  
(2.36)
is clearly Lorentz invariant, since it is an electric charge. Clearly, all Lorentz observers will agree on the number of electrons in a specified closed spatial region, and so they will agree on the amount of charge. Another quantity that is Lorentz invariant is

\[
dv = dt dx dy dz ,
\]  
(2.37)

the 4-volume element of an infinitesimal volume in spacetime. This can be seen from the fact that the Jacobian \( J \) of the transformation from \( dv \) to \( dv' = dt' dx' dy' dz' \) is given by

\[
J = \det \left( \frac{\partial x'^\mu}{\partial x^\nu} \right) = \det(\Lambda^\mu_\nu) .
\]  
(2.38)

Now the defining property (1.80) of the Lorentz transformation can be written in a matrix notation as

\[
\Lambda^T \eta \Lambda = \eta ,
\]  
(2.39)

and hence taking the determinant, we get \( (\det \Lambda)^2 = 1 \) and hence

\[
\det \Lambda = \pm 1 .
\]  
(2.40)

Assuming that we restrict attention to Lorentz transformations without reflections, then they will be connected to the identity (we can take the boost velocity \( \vec{v} \) to zero and/or
the rotation angle to zero and continuously approach the identity transformation), and so
\text{det} \Lambda = 1. \text{ Thus it follows from (2.38) that for Lorentz transformations without reflections, the 4-volume element } dv = dt dx dy dz \text{ is invariant.}

Comparing \(dQ = \rho dx dy dz\) and \(dv = dt dx dy dz\), both of which we have argued are Lorentz invariant, we can conclude that, just as \(dt\) transforms as the 0 component of a 4-vector, so the charge density \(\rho\) must transform as the 0 component of a 4-vector under Lorentz transformations. Thus writing, as we did, that \(J^0 = \rho\), is justified.

In the same way, we may consider the spatial components \(J^i\) of the putative 4-vector \(J^\mu\). Considering \(J^1\), for example, we know that \(J^1 dy dz\) is the current flowing through the area element \(dy dz\). Therefore in time \(dt\), there will have been a flow of charge \(J^1 dt dy dz\). Being a charge, this must be Lorentz invariant, and so it follows from the known Lorentz invariance of \(dv = dt dx dy dz\) that \(J^1\) must transform the same way as \(dx\) under Lorentz transformations. That is, \(J^1\) must transform as the 1 component of a 4-vector. Similar arguments apply to \(J^2\) and \(J^3\). (It is important in this argument that, because of the charge-conservation equation (2.33) or (2.35), the flow of charges we are discussing when considering the \(J^i\) components are the same charges we discussed when considering the \(J^0\) component.)

We have now argued that \(J^\mu = (\rho, J^i)\) is indeed a Lorentz 4-vector, where \(\rho\) is the charge density and \(J^i\) the 3-vector current density.

Actually, the argument we have presented for showing that \(J^\mu\) is a 4-vector is a little sketchy. One should really be rather more careful about getting the orientations of the 3-area elements orthogonal to the 0, 1, 2 and 3 directions right. This involves defining the infinitesimal 3-area 4-vector
\[
d\Sigma_\mu = (dx dy dz, -dt dy dz, -dt dz dx, -dt dx dy). \quad (2.41)
\]
Charge conservation can then be described in 4-dimensional terms, by considering an arbitrary 4-volume \(V_4\) in spacetime, which is bounded by a 3-dimensional surface \(S_3\). Integrating \(\partial_\mu J^\mu = 0\) over \(V_4\) and then using the 4-dimensional analogue of the divergence theorem gives\(^{11}\)
\[
0 = \int_{V_4} \partial_\mu J^\mu \, d^4x = \int_{S_3} J^\mu \, d\Sigma_\mu, \quad (2.42)
\]
\(^{11}\)One can verify, by carefully generalising the usual proof of the 3-dimensional divergence theorem to four dimensions, that the signs given in the definition of \(d\Sigma_\mu\) in (2.41) are correct, where the convention is that a 3-area element such as \(dt dx dy\) is positively oriented, in the sense that \(dt dx dy\) would give a positive number when integrated over a 3-area in the \((t, x, y)\) hyperplane.
where \( d^4x = dt dx dy dz \) is the infinitesimal 4-volume element.

Now consider a 4-volume \( V_4 \) comprising the entire infinite spatial 3-volume sandwiched between an initial timelike surface at \( t = t_1 \) and a final timelike surface at \( t = t_2 \). Thus the bounding 3-surface \( S_3 \) is like an infinite-radius hyper-cylinder, comprising the two end-caps given by the infinite spatial hyperplane at \( t = t_1 \) with \(-\infty \leq x \leq \infty, -\infty \leq y \leq \infty \) and \(-\infty \leq z \leq \infty \), and the infinite spatial hyperplane at \( t = t_2 \), again with \(-\infty \leq x \leq \infty, -\infty \leq y \leq \infty \) and \(-\infty \leq z \leq \infty \); and then finally the the side of the cylinder, with \( t_1 \leq t \leq t_2 \) and \( x, y \) and \( z \) all being at infinity (the “sphere at infinity”). We shall assume that the charge and current densities are localised, and that they fall off at spatial infinity. Thus from (2.42) we then have

\[
0 = -\int_{\text{Cap 1}} J^0 dx dy dz + \int_{\text{Cap 2}} J^0 dx dy dz + \int_{\text{Sphere at infinity}} J^i d\Sigma_i. \tag{2.43}
\]

(The minus sign on the first term is because in the divergence theorem the 3-area element \( d\Sigma_i \) points outwards from the 4-volume \( V_4 \) on all of the boundary surface.) The fall-off assumptions imply the final integral vanishes, and so we have

\[
\int_{\text{Cap 1}} J^0 dx dy dz = \int_{\text{Cap 2}} J^0 dx dy dz, \quad \text{i.e.} \quad \int_{t=t_1} \rho dx dy dz = \int_{t=t_2} \rho dx dy dz, \tag{2.44}
\]

and hence we conclude that the total charge at time \( t_1 \) is the same as the total charge at time \( t_2 \). That is to say, charge is conserved.

The conclusion about charge conservation would be the same for any choice of inertial observer. That is to say, the conclusion must be independent of the choice of Lorentz frame. This implies that \( J^\mu \, d\Sigma_\mu \) must be Lorentz invariant. Since one could consider all possible timelike slicings for the “sandwich” in (2.42), one is essentially saying that \( J^\mu \, d\Sigma_\mu \) must be a Lorentz scalar where the 4-vector \( d\Sigma_\mu \) can be oriented arbitrarily. By the quotient theorem (see homework 2), it therefore follows that \( J^\mu = (\rho, J^i) \) must be a 4-vector.

At this point, we recall that by choosing the Lorenz gauge (2.11), we were able to reduce the Maxwell field equations (2.3) to (2.12). Furthermore, we can write these equations together as

\[
\Box A^\mu = -4\pi J^\mu, \tag{2.45}
\]

where

\[
A^\mu = (\phi, \vec{A}), \tag{2.46}
\]

where the d’Alembertian, or wave operator, \( \Box = \partial^\mu \partial_\mu = \partial_i \partial_i - \partial_0^2 \) was introduced in (1.115). We saw that it is manifestly a Lorentz scalar operator, since it is built from the contraction
of indices on the two Lorentz-vector gradient operators. Since we have already established that $J^\mu$ is a 4-vector, it therefore follows that $A^\mu$ is a 4-vector. Note, *en passant*, that the Lorenz gauge condition (2.11) that we imposed earlier translates, in the four-dimensional language, into

$$\partial_\mu A^\mu = 0,$$  \hspace{1cm} (2.47)

which is nicely Lorentz invariant.

The final step is to note that our definition (2.14) is precisely consistent with (2.46) and (2.8), if we write

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu.$$

(2.48)

First, we note from (2.46) that because of the $\eta_{00} = -1$ needed when lowering the 0 index, we shall have

$$A_\mu = (-\phi, \vec{A}).$$

(2.49)

Therefore we find

$$F_{0i} = \partial_0 A_i - \partial_i A_0 = \frac{\partial A_i}{\partial t} + \partial_i \phi = -E_i,$$

$$F_{ij} = \partial_i A_j - \partial_j A_i = \epsilon_{ijk}(\vec{\nabla} \times \vec{A})_k = \epsilon_{ijk} B_k.$$  \hspace{1cm} (2.50)

In summary, we have shown that $J^\mu$ is a 4-vector, and hence, using (2.45), that $A^\mu$ is a 4-vector. Then, it is manifest from (2.48) that $F_{\mu\nu}$ is a 4-tensor. Hence, we have established that the Maxwell equations, written in the form (2.18) and (2.19), are indeed expressed in terms of 4-tensors and 4-vectors, and so the manifest Lorentz covariance of the Maxwell equations is established.

Finally, it is worth remarking that in the 4-tensor description, the way in which the gauge invariance arises is very straightforward. First, it is manifest that the Bianchi identity (2.19) is solved identically by writing

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu,$$

(2.51)

for some 4-vector $A_\mu$. This is because (2.19) is totally antisymmetric in $\mu \nu \rho$, and so, when (2.51) is substituted into it, one gets identically zero since partial derivatives commute. (Try making the substitution and verify this explicitly. The vanishing because of the commutativity of partial derivatives is essentially the same as the reason why $\text{curl grad} \equiv 0$ and $\text{div curl} \equiv 0$.) It is also clear from (2.51) that $F_{\mu\nu}$ will be unchanged if we make the replacement

$$A_\mu \rightarrow A_\mu + \partial_\mu \lambda,$$

(2.52)
where $\lambda$ is an arbitrary function of position and time. Again, the reason is that partial derivatives commute. Comparing (2.52) with (2.49), we see that (2.52) implies

$$\phi \rightarrow \phi - \frac{\partial \lambda}{\partial t}, \quad A_i \rightarrow A_i + \partial_i \lambda,$$

(2.53)

and so we have reproduced the gauge transformations (2.9) and (2.10).

It should have become clear by now that all the familiar features of the Maxwell equations are equivalently described in the spacetime formulation in terms of 4-vectors and 4-tensors. The only difference is that everything is described much more simply and elegantly in the four-dimensional language.

2.4 Lorentz transformation of $\vec{E}$ and $\vec{B}$

Although for many purposes the four-dimensional description of the Maxwell equations is the most convenient, it is sometimes useful to revert to the original description in terms of $\vec{E}$ and $\vec{B}$. For example, we may easily derive the Lorentz transformation properties of $\vec{E}$ and $\vec{B}$, making use of the four-dimensional formulation. In terms of $F_{\mu\nu}$, there is no work needed to write down its behaviour under Lorentz transformations. Raising the indices for convenience, we shall have

$$F'_{\mu\nu} = \Lambda_\mu^\rho \Lambda^\nu_\sigma F_{\rho\sigma}.$$  

(2.54)

From this, and the fact (see (2.14) that $F_{0i} = E_i$, $F^{ij} = \epsilon_{ijk} B_k$, we can then immediately read of the Lorentz transformations for $\vec{E}$ and $\vec{B}$.

From the expressions (1.75) for the most general Lorentz boost transformation, we may first calculate $E'_i$, calculated from

$$E'_i = F'^{0i} = \Lambda^0_\rho \Lambda^i_\sigma F^{\rho\sigma},$$

$$= \Lambda^0_0 \Lambda^i_k F^{0k} + \Lambda^0_k \Lambda^i_0 F^{k0} + \Lambda^0_k \Lambda^i_\ell F^{k\ell},$$

$$= \gamma \left( \delta_{ik} + \frac{\gamma - 1}{v^2} v_i v_k \right) E_k - \gamma^2 v_i v_k E_k - \gamma v_k \left( \delta_{i\ell} + \frac{\gamma - 1}{v^2} v_i v_\ell \right) \epsilon_{k\ell m} B_m,$$

$$= \gamma E_i + \gamma \epsilon_{ijk} v_j B_k - \frac{\gamma - 1}{v^2} v_i v_k E_k.$$  

(2.55)

(Note that because $F^{\mu\nu}$ is antisymmetric, there is no $F^{00}$ term on the right-hand side on the second line.) Thus, in terms of 3-vector notation, the Lorentz boost transformation of the electric field is given by

$$\vec{E}' = \gamma (\vec{E} + \vec{v} \times \vec{B}) - \frac{\gamma - 1}{v^2} (\vec{v} \cdot \vec{E}) \vec{v}.$$  

(2.56)
An analogous calculation shows that the Lorentz boost transformation of the magnetic field is given by
\[
\vec{B}' = \gamma (\vec{B} - \vec{v} \times \vec{E}) - \frac{\gamma - 1}{v^2} (\vec{v} \cdot \vec{B}) \vec{v}.
\] (2.57)

Suppose, for example, that in the frame \(S\) there is just a magnetic field \(\vec{B}\), while \(\vec{E} = 0\). An observer in a frame \(S'\) moving with uniform velocity \(\vec{v}\) relative to \(S\) will therefore observe not only a magnetic field, given by
\[
\vec{B}' = \gamma \vec{B} - \frac{\gamma - 1}{v^2} (\vec{v} \cdot \vec{B}) \vec{v},
\] (2.58)
but also an electric field, given by
\[
\vec{E}' = \gamma \vec{v} \times \vec{B}.
\] (2.59)

This, of course, is the principle of the dynamo.\(^\text{12}\)

It is instructive to write out the Lorentz transformations explicitly in the case when the boost is along the \(x\) direction, \(\vec{v} = (v, 0, 0)\). Equations (2.56) and (2.57) become
\[
E'_x = E_x, \quad E'_y = \gamma (E_y - vB_z), \quad E'_z = \gamma (E_z + vB_y),
\]
\[
B'_x = B_x, \quad B'_y = \gamma (B_y + vE_z), \quad B'_z = \gamma (B_z - vE_y).
\] (2.60)

### 2.5 The Lorentz force

Consider a point particle following the path, or worldline, \(x^\mu = x^\mu(\tau)\) in Minkowski space-time. As we saw earlier, its 4-velocity is given by
\[
U^\mu = \frac{dx^\mu(\tau)}{d\tau} = (\gamma, \gamma \vec{u}), \quad \text{where} \quad \gamma = \frac{1}{\sqrt{1 - u^2}},
\] (2.61)
where \(u^i = dx^i/dt\) is its 3-velocity. Multiplying (2.61) by the rest mass \(m\) of the particle gives another 4-vector, namely the 4-momentum
\[
p^\mu = mU^\mu = (m\gamma, m\gamma \vec{u}).
\] (2.62)

The 0 component \(p^0 = m\gamma\) is called the relativistic energy \(E\), and the spatial components \(p^i = m\gamma u^i\) are called the relativistic 3-momentum. Note that since \(U^\mu U_\mu = -1\), we shall have
\[
p^\mu p_\mu = -m^2.
\] (2.63)

\(^\text{12}\)In a practical dynamo the rotor is moving with a velocity \(\vec{v}\) which is much less than the speed of light, i.e. \(|\vec{v}| \ll 1\) in natural units. This means that the gamma factor \(\gamma = (1 - v^2)^{-1/2}\) is approximately equal to unity in such cases.
We now define the relativistic 4-force \( f^\mu \) acting on the particle to be

\[
f^\mu = \frac{dp^\mu}{d\tau},
\]

where \( \tau \) is the proper time. Clearly \( f^\mu \) is indeed a 4-vector, since it is the 4-vector \( dp^\mu \) divided by the scalar \( d\tau \).

Using (2.62), we can write the 4-force as

\[
f^\mu = \left( m\gamma^2 \vec{u} \cdot \frac{d\vec{u}}{d\tau}, m\gamma^2 \vec{u} \cdot \frac{d\vec{u}}{d\tau} \vec{u} + m\gamma \frac{d\vec{u}}{d\tau} \right).
\]

(2.65)

It follows that if we move to the instantaneous rest frame of the particle, i.e. the frame in which \( \vec{u} = 0 \) at the particular moment we are considering, then \( f^\mu \) reduces to

\[
f^\mu \bigg|_{\text{rest frame}} = (0, \vec{F}),
\]

(2.66)

where

\[
\vec{F} = m \frac{d\vec{u}}{dt}
\]

(2.67)

is the Newtonian force measured in the rest frame of the particle.\(^{13}\) Thus, we should interpret the 4-force physically as describing the Newtonian 3-force when measured in the instantaneous rest frame of the accelerating particle.

If we now suppose that the particle has electric charge \( e \), and that it is moving under the influence of an electromagnetic field \( F_{\mu\nu} \), then its motion is given by the Lorentz force equation

\[
f^\mu = eF^{\mu\nu} U_\nu.
\]

(2.68)

One can more or less justify this equation on the grounds of “what else could it be?”, since we know that there must exist a relativistic equation (i.e. a Lorentz covariant equation) that describes the motion. In fact it is easy to see that (2.68) is correct. We calculate the spatial components:

\[
\begin{align*}
    f^i &= eF^{i\nu} U_\nu = eF^{i0} U_0 + eF^{ij} U_j, \\
    &= e(-E_i)(-\gamma) + e\epsilon_{ijk} B_k \gamma u_j,
\end{align*}
\]

(2.69)

and thus

\[
\vec{f} = e\gamma (\vec{E} + \vec{u} \times \vec{B}).
\]

(2.70)

\(^{13}\)Note that we can replace the proper time \( \tau \) by the coordinate time \( t \) in the instantaneous rest frame, since \( d\tau = dt/\gamma \), and \( \gamma = 1 \) when \( \vec{u} = 0 \).
But \( f^\mu = dp^\mu/d\tau \), and so \( \vec{f} = d\vec{p}/d\tau = \gamma d\vec{p}/dt \) (recall from section 1.6 that \( d\tau = dt/\gamma \)) and so we have

\[
\frac{d\vec{p}}{dt} = e(\vec{E} + \vec{u} \times \vec{B}),
\]

(2.71)

where \( d\vec{p}/dt \) is the rate of change of the relativistic 3-momentum \( \vec{p} = m\gamma \vec{u} \). This is indeed the standard expression for the motion of a charged particle under the Lorentz force.

### 2.6 Action principle for charged particles

In this section, we shall show how the equations of motion for a charged particle moving in an electromagnetic field can be derived from an action principle. To begin, we shall consider an uncharged particle of mass \( m \), with no forces acting on it. It will, of course, move in a straight line. It turns out that its equation of motion can be derived from the Lorentz-invariant action

\[
S = -m \int_{\tau_1}^{\tau_2} d\tau,
\]

(2.72)

where \( \tau \) is the proper time along the trajectory \( x^\mu(\tau) \) of the particle, starting at proper time \( \tau = \tau_1 \) and ending at \( \tau = \tau_2 \). The action principle then states that if we consider all possible paths between the initial and final spacetime points on the path, then the actual path followed by the particle will be such that the action \( S \) is stationary. In other words, if we consider small variations of the path around the actual path, then to first order in the variations we shall have \( \delta S = 0 \).

To see how this works, we note that \( d\tau^2 = dt^2 - dx^i dx^i = dt^2(1 - v_i v_i) = dt^2(1 - v^2) \), where \( v_i = dx^i/dt \) is the 3-velocity of the particle. Thus \( d\tau = dt/\gamma \) where \( \gamma = (1 - v^2)^{-1/2} \) and so

\[
S = -m \int_{t_1}^{t_2} \frac{dt}{\gamma} = -m \int_{t_1}^{t_2} (1 - v^2)^{1/2} dt = -m \int_{t_1}^{t_2} (1 - \dot{x}^i \dot{x}^i)^{1/2} dt. \quad (2.73)
\]

In other words, the Lagrangian \( L \), for which \( S = \int_{t_1}^{t_2} L dt \), is given by

\[
L = -\frac{m}{\gamma} = -m(1 - \dot{x}^i \dot{x}^i)^{1/2}. \quad (2.74)
\]

As a check, if we expand (2.74) for small velocities (i.e. small compared with the speed of light, so \( |\dot{x}^i| << 1 \)), we shall have

\[
L = -m + \frac{1}{2} m v^2 + \cdots. \quad (2.75)
\]

Since the Lagrangian is given by \( L = T - V \) we see that \( T \) is just the usual kinetic energy \( \frac{1}{2} m v^2 \) for a non-relativistic particle of mass \( m \), while the potential energy is just \( m \). Of course if we were not using units where the speed of light were unity, this energy would be
$mc^2$. Since it is just a constant, this rest-mass energy of the particle does not affect the equations of motion that will follow from the action principle.

Now let us consider small variations $\delta x^i(t)$ around the path $x^i(t)$ followed by the particle. The action will vary according to

$$
\delta S = m \int_{t_1}^{t_2} (1 - \dot{x}^i \dot{x}^i)^{-1/2} \dot{x}^i \delta \dot{x}^i dt.
$$

(2.76)

Integrating by parts then gives

$$
\delta S = -m \int_{t_1}^{t_2} \frac{d}{dt} \left( (1 - \dot{x}^i \dot{x}^i)^{-1/2} \dot{x}^i \right) \delta \dot{x}^i dt + m \left[ (1 - \dot{x}^i \dot{x}^i)^{-1/2} \dot{x}^i \delta \dot{x}^i \right]_{t_1}^{t_2}.
$$

(2.77)

As usual in an action principle, we restrict to variations of the path that vanish at the endpoints, so $\delta x^i(t_1) = \delta x^i(t_2) = 0$ and the boundary term can be dropped. The variation $\delta x^i$ is allowed to be otherwise arbitrary in the time interval $t_1 < t < t_2$, and so we conclude from the requirement of stationary action $\delta S = 0$ that

$$
\frac{d}{dt} \left( m(1 - \dot{x}^i \dot{x}^i)^{-1/2} \dot{x}^i \right) = 0.
$$

(2.78)

Now, recalling that we define $\gamma = (1 - v^2)^{-1/2}$, we see that

$$
\frac{d(m\gamma \vec{v})}{dt} = 0,
$$

(2.79)

or, in other words,

$$
\frac{d\vec{p}}{dt} = 0,
$$

(2.80)

where $\vec{p} = m\gamma \vec{v}$ is the relativistic 3-momentum. We have, of course, derived the equation for straight-line motion in the absence of any forces acting.

Now we extend the discussion to the case of a particle of mass $m$ and charge $e$, moving under the influence of an electromagnetic field $F_{\mu\nu}$. This field will be written in terms of a 4-vector potential:

$$
F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu.
$$

(2.81)

The action will now be the sum of the free-particle action (2.73) above plus a term describing the interaction of the particle with the electromagnetic field. The total action turns out to be

$$
S = \int_{\tau_1}^{\tau_2} (-md\tau + eA_\mu dx^\mu).
$$

(2.82)

Note that it is again Lorentz invariant.

From (2.49) we have $A^\mu = (\phi, \vec{A})$ and hence $A_\mu = (-\phi, \vec{A})$, and so

$$
A_\mu dx^\mu = A_\mu \frac{dx^\mu}{dt} dt = (A_0 + A_i \dot{x}^i) dt = (-\phi + A_i \dot{x}^i) dt.
$$

(2.83)
Thus we have $S = \int_{t_1}^{t_2} L \, dt$ with the Lagrangian $L$ given by

$$L = -m(1 - \dot{x}^j \dot{x}^j)^{1/2} - e\phi + eA_i \dot{x}^i,$$

(2.84)

where potentials $\phi$ and $A_i$ depend on $t$ and $x$. The first-order variation of the action under a variation $\delta x^i$ in the path gives

$$\delta S = \int_{t_1}^{t_2} \left[ m(1 - \dot{x}^j \dot{x}^j)^{-1/2} \dot{x}^i \delta \dot{x}^i - e\partial_i \phi \delta x^i + eA_i \delta \dot{x}^i + e\partial_j A_i \dot{x}^i \delta x^j \right] dt,$$

$$= \int_{t_1}^{t_2} \left[ -\frac{d}{dt}(m\gamma \dot{x}^i) - e\partial_i \phi - e\frac{dA_i}{dt} + e\partial_i A_j \dot{x}^j \right] \delta x^i dt.$$  

(2.85)

(We have dropped the boundary terms immediately, since $\delta x^i$ is again assumed to vanish at the endpoints.) Thus the principle of stationary action $\delta S = 0$ implies

$$\frac{d(m\gamma \dot{x}^i)}{dt} = -e\partial_i \phi - \frac{dA_i}{dt} + e\partial_i A_j \dot{x}^j.$$  

(2.86)

Now, the total time derivative $dA_i/dt$ has two contributions, and we may write it as

$$\frac{dA_i}{dt} = \partial A_i + \partial_j A_i \frac{dx^j}{dt} = \partial A_i + \partial_j A_i \dot{x}^j.$$  

(2.87)

This arises because first of all, $A_i$ can depend explicitly on the time coordinate; this contribution is $\partial A_i/\partial t$. Additionally, $A_i$ depends on the spatial coordinates $x^i$, and along the path followed by the particle, $x^i$ depends on $t$ because the path is $x^i = x^i(t)$. This accounts for the second term.

Putting all this together, we have

$$\frac{d(m\gamma \dot{x}^i)}{dt} = e \left( -\partial_i \phi - \partial A_i \right) + e(\partial_i A_j - \partial_j A_i) \dot{x}^j,$$

$$= e(E_i + \epsilon_{ijk} \dot{x}^j B_k).$$  

(2.88)

In other words, we have

$$\frac{d\vec{p}}{dt} = e(\vec{E} + \vec{v} \times \vec{B}),$$  

(2.89)

which is the Lorentz force equation (2.71).

It is worth noting that although we gave a “three-dimensional” derivation of the equations of motion following from the action (2.82), we can also instead directly derive the four-dimensional equation $dp^\mu/d\tau = eF^{\mu\nu}U_\nu$. To begin, we note that since $d\tau^2 = -\eta_{\mu\nu}dx^\mu dx^\nu$, its variation variation under a variation of the path $x^\mu(\tau)$ in spacetime gives $2d\tau \delta(d\tau) = -2\eta_{\mu\nu}d(\delta x^\mu) \, dx^\nu$, and so dividing by $2d\tau$ gives

$$\delta(d\tau) = -\eta_{\mu\nu} \frac{dx^\nu}{d\tau} d(\delta x^\mu),$$

$$= -U_\mu d(\delta x^\mu),$$  

(2.90)
where $U_\mu$ is the 4-velocity. Thus the variation of the action (2.82) gives

$$
\delta S = \int_{\tau_1}^{\tau_2} \left( m U_\mu \delta x^\mu + e A_\mu \delta x^\mu + e \partial_\nu A_\mu \delta x^\mu d\tau \right),
$$

where we have dropped the boundary terms $\int_{\tau_1}^{\tau_2} d(m U_\mu \delta x^\mu + e A_\mu \delta x^\mu)$ in getting to the third line, since they integrate to give $[m U_\mu \delta x^\mu + e A_\mu \delta x^\mu]_{\tau_1}^{\tau_2}$ and therefore vanish, as we assume $\delta x^\mu = 0$ at the initial and final proper times $\tau_1$ and $\tau_2$. Now by the chain rule

$$
\frac{dA_\mu}{d\tau} = \partial_\nu A_\mu \frac{dx^\nu}{d\tau} = \partial_\nu A_\mu U^\nu,
$$

and so

$$
\delta S = \int_{\tau_1}^{\tau_2} \left( - m \frac{dU_\mu}{d\tau} - e \partial_\nu A_\mu U^\nu + e \partial_\mu A_\nu U^\nu \right) \delta x^\mu d\tau,
$$

Requiring $\delta S = 0$ for all variations (that vanish at the endpoints) we therefore obtain the equation of motion

$$
m \frac{dU_\mu}{d\tau} = e F^\mu_{\nu} U^\nu.
$$

Thus we have reproduced the Lorentz force equation in its four-dimensionally covariant form

$$
\frac{dp^\mu}{d\tau} = e F^\mu_{\nu} U^\nu,
$$

where $p^\mu = mU^\mu$ is the 4-momentum.

### 2.7 Gauge invariance of the action

In writing down the relativistic action (2.82) for a charged particle we had to make use of the 4-vector potential $A_\mu$. This is itself not physically observable, since, as we noted earlier, $A_\mu$ and $A'_\mu = A_\mu + \partial \lambda$ describe the same physics, where $\lambda$ is any arbitrary function in spacetime, since $A_\mu$ and $A'_\mu$ give rise to the same electromagnetic field $F_{\mu\nu}$. One might worry, therefore, that the action itself would be gauge dependent, and therefore might not properly describe the required physical situation. However, all is in fact well. This already
can be seen from the fact that, as we demonstrated, the variational principle for the action (2.82) does in fact produce the correct gauge-invariant Lorentz force equation (2.71).

It is instructive also to examine the effects of a gauge transformation directly at the level of the action. If we make the gauge transformation $A_\mu \to A'_\mu = A_\mu + \partial_\mu \lambda$, we see from (2.82) that the action $S$ transforms to $S'$ given by

$$S' = \int_{\tau_1}^{\tau_2} \left( -md\tau + eA_\mu dx^\mu + e\partial_\mu \lambda dx^\mu \right),$$

or

$$S' = S + e \int_{\tau_1}^{\tau_2} \partial_\mu \lambda dx^\mu = e \int_{\tau_1}^{\tau_2} d\lambda,$$

(2.96)

and so

$$S' = S + e[\lambda(\tau_2) - \lambda(\tau_1)].$$

(2.97)

The simplest situation to consider is where we restrict ourselves to gauge transformations that vanish at the endpoints, in which case the action will be gauge invariant, $S' = S$. Even if $\lambda$ is non-vanishing at the endpoints, we see from (2.97) that $S$ and $S'$ merely differ by a constant that depends solely on the values of $\lambda$ at $\tau_1$ and $\tau_2$. Clearly, the addition of this constant has no effect on the equations of motion that one derives from $S'$.

### 2.8 Canonical momentum, and Hamiltonian

Given any Lagrangian $L(x^i, \dot{x}^i, t)$ one defines the canonical momentum $\pi_i$ as

$$\pi_i = \frac{\partial L}{\partial \dot{x}^i}.$$  

(2.98)

The relativistic Lagrangian for the charged particle is given by (2.84), and so we have

$$\pi_i = m(1 - \dot{x}^j \dot{x}^j)^{-1/2} \dot{x}^i + eA_i,$$

(2.99)

or, in other words,

$$\pi_i = m\gamma \dot{x}^i + eA_i,$$

(2.100)

$$\pi_i = p_i + eA_i,$$

(2.101)

where $p_i$ as usual is the standard mechanical relativistic 3-momentum of the particle.

As usual, the Hamiltonian for the system is given by

$$H = \pi_i \dot{x}^i - L,$$

(2.102)

and so we find

$$H = m\gamma \dot{x}^i \dot{x}^i + \frac{m}{\gamma} + e\phi,$$

$$= m\gamma v^2 + \frac{m}{\gamma} + e\phi.$$

(2.103)
Now, \( m \gamma v^2 + m / \gamma = m \gamma (v^2 + (1 - v^2)) = m \gamma, \) so we have

\[
H = m \gamma + e \phi.
\]  
(2.104)

As always, the Hamiltonian is to be viewed as a function of the coordinates \( x^i \) and the canonical momenta \( \pi_i \). To express \( \gamma \) in terms of \( \pi_i \), we note from (2.100) that \( m \gamma \dot{x}^i = \pi_i - eA_i \), and so squaring, we get \( m^2 \gamma^2 v^2 = m^2 v^2 / (1 - v^2) = (\pi_i - eA_i)^2 \). Solving for \( v^2 \), and hence for \( \gamma \), we find that \( m^2 \gamma^2 = (\pi_i - eA_i)^2 + m^2 \), and so finally, from (2.104), we arrive at the Hamiltonian

\[
H = \sqrt{(\pi_i - eA_i)^2 + m^2} + e \phi,
\]  
(2.105)

with \( H \) expressed as a function of the coordinates \( x^i \) and the canonical momenta \( \pi_i \).

Note that Hamilton’s equations, which will necessarily give rise to the same Lorentz force equations of motion we encountered previously, are given by

\[
\frac{\partial H}{\partial \pi_i} = \dot{x}^i, \quad \frac{\partial H}{\partial x^i} = -\dot{\pi}_i.
\]  
(2.106)

As a check of the correctness of the Hamiltonian (2.105) we may examine it in the non-relativistic limit when \((\pi_i - eA_i)^2\) is much less than \(m^2\). We then extract an \( m^2 \) factor from inside the square root in \( \sqrt{(\pi_i - eA_i)^2 + m^2} \) and expand to get

\[
H = m \sqrt{1 + (\pi_i - eA_i)^2 / m^2} + e \phi,
\]

\[
= m + \frac{1}{2m} (\pi_i - eA_i)^2 + e \phi + \cdots.
\]  
(2.107)

The first term is the rest-mass energy, which is just a constant, and the remaining terms presented explicitly in (2.107) give the standard non-relativistic Hamiltonian for a charged particle

\[
H_{\text{non-rel.}} = \frac{1}{2m} (\pi_i - eA_i)^2 + e \phi.
\]  
(2.108)

This should be familiar from quantum mechanics, when one writes down the Schrödinger equation for the wave function for a charged particle in an electromagnetic field.

### 3 Particle Motion in Static Electromagnetic Fields

In this chapter, we discuss the motion of a charged particle in static (i.e. time-independent) electromagnetic fields.
3.1 Description in terms of potentials

If we are describing static electric and magnetic fields, \( \vec{E} = \vec{E}(\vec{r}) \) and \( \vec{B} = \vec{B}(\vec{r}) \), it is natural (and always possible) to describe them in terms of scalar and 3-vector potentials that are also static, \( \phi = \phi(\vec{r}) \), \( \vec{A} = \vec{A}(\vec{r}) \). Thus we write

\[
\vec{E} = -\nabla \phi - \frac{\partial \vec{A}}{\partial t} = -\nabla \phi(\vec{r}),
\]
\[
\vec{B} = \nabla \times \vec{A}(\vec{r}).
\] (3.1)

We can still perform gauge transformations, as given in (2.9) and (2.10). The most general gauge transformation that preserves the time-independence of the potentials is therefore given by taking the parameter \( \lambda \) to be of the form

\[
\lambda(\vec{r}, t) = \lambda(\vec{r}) + k t,
\] (3.2)

where \( k \) is an arbitrary constant. This implies that \( \phi \) and \( \vec{A} \) will transform according to

\[
\phi \rightarrow \phi - k, \quad \vec{A} \rightarrow \vec{A} + \nabla \lambda(\vec{r}).
\] (3.3)

Note, in particular, that the electrostatic potential \( \phi \) can just be shifted by an arbitrary constant. This is the familiar freedom that one typically uses to set \( \phi = 0 \) at infinity.

Recall that the Hamiltonian for a particle of mass \( m \) and charge \( e \) in an electromagnetic field is given by (2.104)

\[
H = m\gamma + e \phi,
\] (3.4)

where \( \gamma = (1 - v^2)^{-1/2} \). In the present situation with static fields, the Hamiltonian does not depend explicitly on time, i.e. \( \partial H/\partial t = 0 \). It then follows that the Hamiltonian is conserved (i.e. it is the same at all times) since we have (by the chain rule)

\[
\frac{dH}{dt} = \frac{\partial H}{\partial t} + \frac{\partial H}{\partial x^i} \dot{x}^i + \frac{\partial H}{\partial \pi_i} \dot{\pi}_i = 0.
\] (3.5)

(We used the Hamilton equations (2.106) in getting to the second line.) This time-independent quantity \( H \) is then just the energy \( \mathcal{E} \) of the system:

\[
\mathcal{E} \equiv H = m\gamma + e \phi.
\] (3.6)

We may think of the first term in \( \mathcal{E} \) as being the *mechanical* term,

\[
\mathcal{E}_{\text{mech}} = m\gamma,
\] (3.7)
since this is just the total energy of a particle of rest mass $m$ moving with velocity $\vec{v}$. The second term, $e\phi$, is the contribution to the total energy from the electric field. Note that the magnetic field, described by the 3-vector potential $\vec{A}$, does not contribute to the conserved energy. This is because the magnetic field $\vec{B}$ does no work on the charge:

Recall that the Lorentz force equation can be written as

$$\frac{d(m\gamma v^i)}{dt} = e(E_i + \epsilon_{ijk} v^j B_k).$$

(3.8)

Multiplying by $v^i$ we therefore have

$$m\gamma v^i \frac{dv^i}{dt} + mv^i v^i \frac{d\gamma}{dt} = ev^i E_i.$$  

(3.9)

Now $\gamma = (1 - v^2)^{-1/2}$, so

$$\frac{d\gamma}{dt} = (1 - v^2)^{-3/2} v^i \frac{dv^i}{dt} = \gamma^3 v^i \frac{dv^i}{dt},$$

(3.10)

and so (3.9) gives

$$m \frac{d\gamma}{dt} = ev^i E_i.$$  

(3.11)

Since $E_{\text{mech}} = m\gamma$, and $m$ is a constant, we therefore have

$$\frac{dE_{\text{mech}}}{dt} = e\vec{v} \cdot \vec{E}.$$  

(3.12)

Thus, the mechanical energy of the particle is changed only by the electric field, and not by the magnetic field.

Note that another (and equivalent) derivation of the constancy of $E = m\gamma + e\phi$ is as follows:

$$\frac{dE}{dt} = \frac{d(m\gamma)}{dt} + e \frac{d\phi}{dt}$$

$$= \frac{dE_{\text{mech}}}{dt} + e \partial_i \phi \frac{dx^i}{dt},$$

$$= e\vec{v} \cdot \vec{E} - e\vec{v} \cdot \vec{E} = 0.$$  

(3.13)

### 3.2 Particle motion in static uniform $\vec{E}$ and $\vec{B}$ fields

Let us consider the case where a charged particle is moving in static (i.e. time-independent) uniform $\vec{E}$ and $\vec{B}$ fields. In other words, $\vec{E}$ and $\vec{B}$ are constant vectors, independent of time and of position. In this situation, it is easy to write down explicit expressions for the corresponding scalar and 3-vector potentials. For the scalar potential, we can take

$$\phi = -\vec{E} \cdot \vec{r} = -E_i x^i.$$  

(3.14)
Clearly this gives the correct electric field, since

$$-\partial_i \phi = \partial_i(E_j x^j) = E_j \partial_i x^j = E_j \delta_{ij} = E_i. \quad (3.15)$$

(It is, of course, essential that $E_j$ is constant for this calculation to be valid.)

Turning now to the uniform $\vec{B}$ field, it is easily seen that this can be written as $\vec{B} = \vec{\nabla} \times \vec{A}$, with the 3-vector potential given by

$$\vec{A} = \frac{1}{2} \vec{B} \times \vec{r}. \quad (3.16)$$

It is easiest to check this using index notation. We have

$$\begin{align*}
(\vec{\nabla} \times \vec{A})_i &= \epsilon_{ijk} \partial_j A_k = \epsilon_{ijk} \epsilon_{jk\ell m} B_\ell x^m, \\
&= \frac{1}{2} \epsilon_{ijk} \epsilon_{j\ell m} B_\ell \partial_j x^m = \frac{1}{2} \epsilon_{ijk} \epsilon_{\ell jk} B_\ell, \\
&= \delta_{i\ell} B_\ell = B_i. \quad (3.17)
\end{align*}$$

Of course the potentials we have written above are not unique, since we can still perform gauge transformations. If we restrict attention to transformations that maintain the time-independence of $\phi$ and $\vec{A}$, then for $\phi$ the only remaining freedom is to add an arbitrary constant to $\phi$. For the 3-vector potential, we can still add $\vec{\nabla} \lambda(\vec{r})$ to $\vec{A}$, where $\lambda(\vec{r})$ is an arbitrary function of position. It is sometimes helpful, for calculational reasons, to do this. Suppose, for example, that the uniform $\vec{B}$ field lies along the $z$ axis: $\vec{B} = (0, 0, B)$. From (3.16), we may therefore write the 3-vector potential

$$\vec{A} = \left(-\frac{1}{2} B y, \frac{1}{2} B x, 0\right). \quad (3.18)$$

Another choice is to take $\vec{A}' = \vec{A} + \vec{\nabla} \lambda(\vec{r})$, with $\lambda = -\frac{1}{2} B x y$. This gives

$$\vec{A}' = (-B y, 0, 0). \quad (3.19)$$

One easily verifies that indeed $\vec{\nabla} \times \vec{A}' = (0, 0, B)$.

### 3.2.1 Motion in a static uniform electric field

From the Lorentz force equation, we shall have

$$\frac{d\vec{p}}{dt} = e \vec{E}, \quad (3.20)$$

where $\vec{p} = m \gamma \vec{v}$ is the relativistic 3-momentum. Without loss of generality, we may take the electric field to lie along the $x$ axis, and so we will have

$$\frac{dp_x}{dt} = e E, \quad \frac{dp_y}{dt} = 0, \quad \frac{dp_z}{dt} = 0. \quad (3.21)$$
Since \( p_y \) and \( p_z \) are therefore constants, we can without loss of generality rotate the coordinate system around the \( x \) axis so that \( p_z = 0 \). Thus we may integrate (3.21 to give

\[
p_x = eEt, \quad p_y = \bar{p}, \quad p_z = 0, \quad \tag{3.22}
\]

where \( \bar{p} \) is a constant. Note that when integrating \( dp_x/\ dt \), we have fixed the unimportant constant of integration by choosing the origin for the time coordinate \( t \) such that \( p_x = 0 \) at \( t = 0 \).

Recalling that the 4-momentum is given by \( p^\mu = (m\gamma, \vec{p}) = (E_{\text{mech}}, \vec{p}) \), and that \( p^\mu p_\mu = m^2U^\mu U_\mu = -m^2 \), we see that \(-E_{\text{mech}}^2 + \vec{p} \cdot \vec{p} = -m^2\), and so

\[
E_{\text{mech}} = \sqrt{m^2 + p_x^2 + p_y^2} = \sqrt{m^2 + \bar{p}^2 + (eEt)^2}. \tag{3.23}
\]

Hence we may write

\[
E_{\text{mech}} = \sqrt{E_0^2 + (eEt)^2}, \tag{3.24}
\]

where \( E_0^2 = m^2 + \bar{p}^2 \) is the square of the mechanical energy at time \( t = 0 \).

We have \( \vec{p} = m\gamma \vec{v} = E_{\text{mech}} \vec{v} \), and so \( p_x = E_{\text{mech}} dx/\ dt \) and therefore

\[
\frac{dx}{\ dt} = \frac{p_x}{E_{\text{mech}}} = \frac{eEt}{\sqrt{E_0^2 + (eEt)^2}}, \tag{3.25}
\]

which can be integrated to give

\[
x = \frac{1}{eE} \sqrt{E_0^2 + (eEt)^2}. \tag{3.26}
\]

(The constant of integration has been absorbed into a choice of origin for the \( x \) coordinate.)

Note from (3.25) that the \( x \)-component of the 3-velocity asymptotically approaches 1 as \( t \) goes to infinity. Thus the particle is accelerated closer and closer to the speed of light, but never reaches it.

We also have

\[
\frac{dy}{\ dt} = \frac{p_y}{E_{\text{mech}}} = \frac{\bar{p}}{\sqrt{E_0^2 + (eEt)^2}}. \tag{3.27}
\]

This can be integrated by changing variable from \( t \) to \( u \), defined by

\[
eEt = E_0 \sinh u. \tag{3.28}
\]

This gives \( y = \bar{p} u/(eE) \), and hence

\[
y = \frac{\bar{p}}{eE} \arcsinh\left( \frac{eEt}{E_0} \right). \tag{3.29}
\]

(Again, the constant of integration has been absorbed into the choice of origin for \( y \).)
The solutions (3.26) and (3.29) for $x$ and $y$ as functions of $t$ can be combined to give $x$ as a function of $y$, leading to

$$x = \frac{\mathcal{E}_0}{eE} \cosh \left( \frac{eEy}{\bar{p}} \right). \quad (3.30)$$

This is a catenary.

In the non-relativistic limit when $|v| << 1$, we have $\bar{p} \approx m\bar{v}$ and then, expanding (3.30) we find the standard “Newtonian” parabolic motion

$$x \approx \text{constant} + \frac{eE}{2m\bar{v}^2} y^2. \quad (3.31)$$

### 3.2.2 Motion in a static uniform magnetic field

From the Lorentz force equation we shall have

$$\frac{d\vec{p}}{dt} = e\vec{v} \times \vec{B}. \quad (3.32)$$

Recalling (3.11), we see that in the absence of an electric field we shall have $\gamma = \text{constant}$, and hence $d\vec{p}/dt = d(m\gamma\vec{v})/dt = m\gamma d\vec{v}/dt$, leading to

$$\frac{d\vec{v}}{dt} = \frac{e}{m\gamma} \vec{v} \times \vec{B} = \frac{e}{\mathcal{E}} \vec{v} \times \vec{B}; \quad (3.33)$$

since $\mathcal{E} = m\gamma + e\phi = m\gamma$ (a constant) here.

Without loss of generality we may choose the uniform $\vec{B}$ field to lie along the $z$ axis: $\vec{B} = (0, 0, B)$. Defining

$$\omega \equiv \frac{eB}{\mathcal{E}} = \frac{eB}{m\gamma}, \quad (3.34)$$

we then find

$$\frac{dv_x}{dt} = \omega v_y, \quad \frac{dv_y}{dt} = -\omega v_x, \quad \frac{dv_z}{dt} = 0. \quad (3.35)$$

From this, it follows that

$$\frac{d(v_x + i v_y)}{dt} = -i \omega (v_x + i v_y), \quad (3.36)$$

and so the first two equations in (3.35) can be integrated to give

$$v_x + i v_y = v_0 e^{-i(\omega t + \alpha)}, \quad (3.37)$$

where $v_0$ is a real constant, and $\alpha$ is a constant (real) phase. Thus after further integrations we obtain

$$x = x_0 + r_0 \sin(\omega t + \alpha), \quad y = y_0 + r_0 \cos(\omega t + \alpha), \quad z = z_0 + \bar{v}_z t, \quad (3.38)$$
for constants \( r_0, x_0, y_0, z_0 \) and \( \bar{v}_z \), with

\[ r_0 = \frac{v_0}{\omega} = \frac{m\gamma v_0}{eB} = \frac{\bar{p}}{eB}, \tag{3.39} \]

where \( \bar{p} \) is the magnitude of the relativistic 3-momentum in the \((x, y)\) plane. The particle therefore follows a helical path, of radius \( r_0 \), twisting along the \( z \) axis.

### 3.2.3 Motion in uniform \( \vec{E} \) and \( \vec{B} \) fields

Having considered the case of particle motion in a uniform \( \vec{E} \) field, and in a uniform \( \vec{B} \) field, we may also consider the situation of motion in uniform \( \vec{E} \) and \( \vec{B} \) fields together. To discuss this in detail is quite involved, and we shall not pursue it extensively here. In fact a relatively simple way to study this general case is to work directly in the 4-dimensional language, solving

\[ \frac{dU^\mu}{d\tau} = \frac{e}{m} F^\mu_\nu U^\nu, \tag{3.40} \]

where \( U = \frac{dx^\mu}{d\tau} \) is the 4-velocity of the particle. Since we are assuming \( \vec{E} \) and \( \vec{B} \) are uniform, constant, fields it follows that \( F^\mu_\nu \) is a constant tensor. See Homework 4, where this approach is explored further.

One can, of course, still approach the problem from a 3-dimensional standpoint. The equations can become quite complicated in general. Here, we consider the situation where we take

\[ \vec{B} = (0, 0, B), \quad \vec{E} = (0, E_y, E_z), \tag{3.41} \]

(there is no loss of generality in choosing axes so that this is the case), and we make the simplifying assumption that the motion is non-relativistic, i.e. \( |\vec{v}| \ll 1 \). The equations of motion will therefore be

\[ m \frac{d\vec{v}}{dt} = e(\vec{E} + \vec{v} \times \vec{B}), \tag{3.42} \]

and so

\[ m\ddot{x} = eB\dot{y}, \quad m\ddot{y} = eE_y - eB\dot{x}, \quad m\ddot{z} = eE_z. \tag{3.43} \]

We can immediately solve for \( z \), finding

\[ z = \frac{e}{2m} E_z t^2 + \bar{v} t, \tag{3.44} \]

where we have chosen the \( z \) origin so that \( z = 0 \) at \( t = 0 \). The \( x \) and \( y \) equations can be combined into

\[ \frac{d}{dt}(\dot{x} + i\dot{y}) + i\omega(\dot{x} + i\dot{y}) = \frac{ie}{m} E_y, \tag{3.45} \]
where $\omega = eB/m$. Thus we find

$$\dot{x} + i \dot{y} = ae^{-i\omega t} \frac{e}{m\omega} E_y = ae^{-i\omega t} + \frac{E_y}{B}.$$  \hfill (3.46)

Choosing the origin of time so that $a$ is real, we have

$$\dot{x} = a \cos \omega t + \frac{E_y}{B}, \quad \dot{y} = -a \sin \omega t.$$  \hfill (3.47)

Taking the time averages, we see that

$$\langle \dot{x} \rangle = \frac{E_y}{B}, \quad \langle \dot{y} \rangle = 0.$$  \hfill (3.48)

The averaged velocity along the $x$ direction is called the *drift velocity*. Notice that it is perpendicular to $\vec{E}$ and $\vec{B}$. It can be written in general as

$$\vec{v}_{\text{drift}} = \frac{\vec{E} \times \vec{B}}{B^2}.$$  \hfill (3.49)

For our assumption that $|\vec{v}| \ll 1$ to be valid, we must have $|\vec{E} \times \vec{B}| \ll B^2$, i.e. $|E_y| \ll |B|$.

Integrating (3.47) once more, we find

$$x = \frac{a}{\omega} \sin \omega t + \frac{E_y}{B} t, \quad y = \frac{a}{\omega} (\cos \omega t - 1),$$  \hfill (3.50)

where the origins of $x$ and $y$ have been chosen so that $x = y = 0$ at $t = 0$. These equations describe the projection of the particle’s motion onto the $(x, y)$ plane. The curve is called a *trochoid*. If $|a| > E_y/B$ there will be loops in the motion, and in the special case $a = -E_y/B$ the curve becomes a *cycloid*, with cusps:

$$x = \frac{E_y}{\omega B} (\omega t - \sin \omega t), \quad y = \frac{E_y}{\omega B} (1 - \cos \omega t).$$  \hfill (3.51)

### 4 Action Principle for Electrodynamics

In this section, we shall show how the Maxwell equations themselves can be derived from an action principle. We shall also introduce the notion of the *energy-momentum tensor* for the electromagnetic field. We begin with a discussion of Lorentz invariant quantities that can be built from the Maxwell field strength tensor $F_{\mu\nu}$.

#### 4.1 Invariants of the electromagnetic field

As we shall now show, it is possible to build two independent Lorentz invariants that are quadratic in the electromagnetic field. One of these will turn out to be just what is needed in order to construct an action for electrodynamics.
4.1.1 The first invariant

The first quadratic invariant is very simple; we may write

\[ I_1 \equiv F_{\mu \nu} F^{\mu \nu}. \]  (4.1)

Obviously this is Lorentz invariant, since it is built from the product of two Lorentz tensors, with all indices contracted. It is instructive to see what this looks like in terms of the electric and magnetic fields. From the expressions given in (2.14), we see that

\[ I_1 = F_{0i} F^{0i} + F_{ij} F^{ij}, \]
\[ = 2F_{0i} F^{0i} + F_{ij} F^{ij} = -2E_i E_i + \epsilon _{ijk} B_k \epsilon _{ij\ell} B_\ell, \]
\[ = -2E_i E_i + 2B_i B_i, \]  (4.2)

and so

\[ I_1 \equiv F_{\mu \nu} F^{\mu \nu} = 2(\vec{B}^2 - \vec{E}^2). \]  (4.3)

One could, of course, verify from the Lorentz transformations (2.56) and (2.57) for \( \vec{E} \) and \( \vec{B} \) that indeed \( (\vec{B}^2 - \vec{E}^2) \) was invariant, i.e. \( I'_1 = I_1 \) under Lorentz transformations. This would be quite an involved computation. However, the great beauty of the 4-dimensional language is that there is absolutely no work needed at all; one can see by inspection that \( F_{\mu \nu} F^{\mu \nu} \) is Lorentz invariant.

4.1.2 The second invariant

The second quadratic invariant that we can write down is given by

\[ I_2 \equiv \frac{1}{2} \epsilon ^{\mu \nu \rho \sigma} F_{\mu \nu} F_{\rho \sigma}. \]  (4.4)

First, we need to explain the tensor \( \epsilon ^{\mu \nu \rho \sigma} \). This is the four-dimensional Minkowski spacetime generalisation of the totally-antisymmetric tensor \( \epsilon ^{ijk} \) of three-dimensional Cartesian tensor analysis. The tensor \( \epsilon ^{\mu \nu \rho \sigma} \) is also totally antisymmetric in all its indices. That means that it changes sign if any two indices are exchanged. For example,\(^{14}\)

\[ \epsilon ^{\mu \nu \rho \sigma} = -\epsilon ^{\nu \mu \rho \sigma} = -\epsilon ^{\mu \nu \sigma \rho} = -\epsilon ^{\sigma \nu \rho \mu}. \]  (4.5)

\(^{14}\)Beware that in an odd dimension, such as 3, the process of “cycling” the indices on \( \epsilon _{ijk} \) (for example, pushing one off the right-hand end and bringing it to the front) is an even permutation; \( \epsilon _{kij} = \epsilon _{ijk} \). By contrast, in an even dimension, such as 4, the process of cycling is an odd permutation; \( \epsilon ^{\sigma \mu \nu \rho} = -\epsilon ^{\mu \nu \sigma \rho} \). This is an elementary point, but easily overlooked if one is familiar only with three dimensions!
Since all the non-vanishing components of $\epsilon^{\mu\nu\rho\sigma}$ are related by the antisymmetry, we need only specify one non-vanishing component in order to define the tensor completely. We shall define

$$\epsilon^{0123} = -1,$$

or, equivalently, $\epsilon_{0123} = +1$. (4.6)

Thus $\epsilon^{\mu\nu\rho\sigma}$ is $-1$, $+1$ or $0$ according to whether $(\mu\nu\rho\sigma)$ is an even permutation of $(0123)$, and odd permutation, or no permutation at all. We use this definition of $\epsilon^{\mu\nu\rho\sigma}$ in all frames.

This can be done because, like the Minkowski metric $\eta_{\mu\nu}$, the tensor $\epsilon^{\mu\nu\rho\sigma}$ is an invariant tensor, as we shall now discuss.

Actually, to be more precise, $\epsilon^{\mu\nu\rho\sigma}$ is an invariant pseudo-tensor. This means that under Lorentz transformations that are connected to the identity (pure boosts and/or pure rotations), it is truly an invariant tensor. However, it reverses its sign under Lorentz transformations that involve a reflection. To see this, let us calculate what the transformation of $\epsilon^{\mu\nu\rho\sigma}$ would be if we assume it behaves as an ordinary Lorentz tensor:

$$\epsilon'^{\mu\nu\rho\sigma} \equiv \Lambda^\mu_\alpha \Lambda^\nu_\beta \Lambda^\rho_\gamma \Lambda^\sigma_\delta \epsilon^{\alpha\beta\gamma\delta},$$

$$= (\det \Lambda) \epsilon^{\mu\nu\rho\sigma}. \quad (4.7)$$

The last equality can easily be seen by writing out all the terms. (It is easier to play around with the analogous identity in 2 or 3 dimensions, to convince oneself of it in an example with fewer terms to write down.) Now, we already saw in section 2.3 that $\det \Lambda = \pm 1$, with $\det \Lambda = +1$ for pure boosts and/or rotations, and $\det \Lambda = -1$ if there is a reflection as well. (See the discussion leading up to equation (2.40).) Thus we see from (4.7) that $\epsilon^{\mu\nu\rho\sigma}$ behaves like an invariant tensor, taking the same values in all Lorentz frames, provided there is no reflection. (Lorentz transformations connected to the identity, i.e. where there is no reflection, are sometimes called proper Lorentz transformations.) In practice, we shall almost always be considering only proper Lorentz transformations, and so the distinction between a tensor and a pseudo-tensor will not concern us.

Returning now to the second quadratic invariant, (4.4), we shall have

$$I_2 = \frac{1}{2} \epsilon^{\mu\nu\rho\sigma} F_{\mu\nu} F_{\rho\sigma} = \frac{1}{2} \times 4 \times \epsilon^{0ijk} F_{0i} F_{jk},$$

$$= 2(-\epsilon_{ijk})(-E_i)\epsilon_{jkl}B_\ell,$$

$$= 4E_iB_i = 4\vec{E} \cdot \vec{B}. \quad (4.8)$$

Thus, to summarise, we have the two quadratic invariants

$$I_1 = F_{\mu\nu} F^{\mu\nu} = 2(\vec{B}^2 - \vec{E}^2),$$

$$I_2 = \frac{1}{2} \epsilon^{\mu\nu\rho\sigma} F_{\mu\nu} F_{\rho\sigma} = 4\vec{E} \cdot \vec{B}. \quad (4.9)$$
Since the two quantities $I_1$ and $I_2$ are (manifestly) Lorentz invariant, this means that, even though it is not directly evident in the three-dimensional language without quite a lot of work, the two quantities

$$\vec{B}^2 - \vec{E}^2, \quad \text{and} \quad \vec{E} \cdot \vec{B}$$

are Lorentz invariant; i.e. they take the same values in all Lorentz frames. This has a number of consequences. For example

1. If $\vec{E}$ and $\vec{B}$ are perpendicular in one Lorentz frame, then they are perpendicular in all Lorentz frames.

2. In particular, if there exists a Lorentz frame where the electromagnetic field is purely electric ($\vec{B} = 0$), or purely magnetic ($\vec{E} = 0$), then $\vec{E}$ and $\vec{B}$ are perpendicular in any other frame.

3. If $|\vec{E}| > |\vec{B}|$ in one frame, then it is true in all frames. Conversely, if $|\vec{E}| < |\vec{B}|$ in one frame, then it is true in all frames.

4. By making an appropriate Lorentz transformation, we can, at a given point, make $\vec{E}$ and $\vec{B}$ equal to any values we like, subject only to the conditions that we cannot alter the values of $(\vec{B}^2 - \vec{E}^2)$ and $\vec{E} \cdot \vec{B}$ at that point.

### 4.2 Action for electrodynamics

We have already discussed the action principle for a charged particle moving in an electromagnetic field. In that discussion, the electromagnetic field was just a specified background, which, of course, would be a solution of the Maxwell equations. We can also derive the Maxwell equations themselves from an action principle, as we shall now show.

We begin by introducing the notion of Lagrangian density. This is a quantity that is integrated over a three-dimensional spatial volume (typically, all of 3-space) to give the Lagrangian:

$$L = \int \mathcal{L} d^3 x.$$  \hspace{1cm} (4.11)

Then, the Lagrangian is integrated over a time interval $t_1 \leq t \leq t_2$ to give the action,

$$S = \int_{t_1}^{t_2} L dt = \int_V \mathcal{L} d^4 x.$$  \hspace{1cm} (4.12)

Here, we are defining the 4-volume $V$ to be the entire infinite spatial 3-space in the sandwich between the initial time surface at $t = t_1$ and the final time surface at $t = t_2$. In order to
make the following discussion more precise, it will be helpful to think of the infinite 3-space as being the limit in which one sends the radius $R$ of a 3-dimensional ball to infinity. While $R$ is still finite, the corresponding 4-volume will be of the form of the interior of a “cylinder” that is coaxial with the time coordinate, with end caps at $t = t_1$ and $t = t_2$, an having radius $R$ in the three spatial directions. We cannot visualise such a 4-dimensional cylinder, but it is really like the higher-dimensional analogue of a tin can, with the time coordinate coaxial with the can. A $t = constant$ slice through a familiar tin can would be a circle; in the present case, the $t = constant$ slice will be the surface of a 2-dimensional sphere of radius $R$. Eventually, when we send $R$ to infinity this sphere becomes the “sphere at infinity” in 3-space.

Consider first the vacuum Maxwell equations without sources,

$$\partial_\mu F^{\mu\nu} = 0, \quad \partial_\mu F_{\nu\rho} + \partial_\nu F_{\rho\mu} + \partial_\rho F_{\mu\nu} = 0.$$  (4.13)

We immediately solve the second equation (the Bianchi identity) by writing $F_{\mu\nu}$ in terms of a potential:

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu.$$  (4.14)

Since the Maxwell field equations are linear in the fields, it is natural to expect that the action should be quadratic. In fact, it turns out that the first invariant we considered above provides the appropriate Lagrangian density. We take

$$\mathcal{L} = -\frac{1}{16\pi} F_{\mu\nu} F^{\mu\nu},$$  (4.15)

and so the action will be

$$S = -\frac{1}{16\pi} \int_V F_{\mu\nu} F^{\mu\nu} d^4x.$$  (4.16)

We can now derive the source-free Maxwell equations by requiring that this action be stationary with respect to variations of the gauge field $A_\mu$. It must be emphasised that we treat $A_\mu$ as the fundamental field here.

The derivation goes as follows. We shall have

$$\delta S = -\frac{1}{16\pi} \int_V (\delta F_{\mu\nu} F^{\mu\nu} + F_{\mu\nu} \delta F^{\mu\nu}) d^4x = -\frac{1}{8\pi} \int_V \delta F_{\mu\nu} F^{\mu\nu} d^4x,$$

$$= -\frac{1}{8\pi} \int_V F^{\mu\nu} (\partial_\mu \delta A_\nu - \partial_\nu \delta A_\mu) d^4x = -\frac{1}{4\pi} \int_V F^{\mu\nu} \partial_\mu \delta A_\nu d^4x,$$

$$= -\frac{1}{4\pi} \int_V \partial_\nu (F^{\mu\nu} \delta A_\mu) d^4x + \frac{1}{4\pi} \int_V (\partial_\mu F^{\mu\nu}) \delta A_\nu d^4x,$$

$$= -\frac{1}{4\pi} \int_\Sigma F^{\mu\nu} \delta A_\nu d\Sigma_\mu + \frac{1}{4\pi} \int_V (\partial_\mu F^{\mu\nu}) \delta A_\nu d^4x,$$

$$= \frac{1}{4\pi} \int_V (\partial_\mu F^{\mu\nu}) \delta A_\nu d^4x.$$  (4.17)
Note that in the penultimate step, we have used the 4-dimensional analogue of the divergence theorem to turn the 4-volume integral of the divergence of a vector into a 3-volume integral over the bounding surface \( \Sigma \). The four-dimensional divergence theorem says that for any 4-vector \( W^\mu \), we have

\[
\int_V (\partial_\mu W^\mu) \, d^4x = \int_\Sigma W^\mu d\Sigma_\mu .
\] (4.18)

In our case, as discussed above, the 4-volume \( V \) consists of the “sandwich” of all of 3-space between the two surfaces \( t = t_1 \) and \( t = t_2 \). We think of this as the limit of a sandwich between \( t = t_1 \) and \( t_2 \) extending out to spatial radius \( R \), where eventually \( R \) is sent to infinity. The 3-dimensional boundary \( \Sigma \) of this “tin can” consists of the endcaps at \( t = t_1 \) and \( t = t_2 \), plus the spatial sphere of radius \( R \) along the sides of the can. The contributions from the integration over \( \Sigma \) have been dropped in getting to the final line in eqn (4.17).

This happens for two reasons: Firstly, by definition we shall require our variations \( \delta A_\mu \) to vanish on the end-caps at \( t = t_1 \) and \( t_2 \). This is part of the specification of the variational problem. It is the analogue of what one always does in a variational problem in particle mechanics: the variations in the path are required to vanish at the initial and final times. We also require the electromagnetic field to fall off at spatial infinity. This ensures that the portion of the surface integral in the penultimate line of (4.17) that is evaluated over the sides of the “tin can” goes to zero as we send the radius \( R \) to infinity.

Finally, we argue that if the action \( S \) is to be stationary for all possible infinitesimal variations \( \delta A_\mu \) that vanish at \( t = t_1 \) and \( t = t_2 \), it must be that the cofactor of \( \delta A_\nu \) in the final line of (4.17) must vanish:

\[
\partial_\mu F^{\mu\nu} = 0.
\] (4.19)

Thus we have derived the source-free Maxwell field equation. Of course the Bianchi identity has already been taken care of by writing \( F_{\mu\nu} \) in terms of the 4-vector potential \( A_\mu \).

The action (4.16), whose variation gave the Maxwell field equation, is written in what is called second-order formalism; that is, the action is expressed in terms of the 4-vector potential \( A_\mu \) as the fundamental field, with \( F_{\mu\nu} \) just being a short-hand notation for \( \partial_\mu A_\nu - \partial_\nu A_\mu \). It is sometimes convenient to use instead the first-order formalism, in which one treats \( A_\mu \) and \( F_{\mu\nu} \) as independent fields. In this formalism, the equation of motion coming from demanding that \( S \) be stationary under variations of \( F_{\mu\nu} \) will derive the equation \( F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu \). To do this, we need a different action as our starting point, namely

\[
S_{f.o.} = \frac{1}{4\pi} \int (\frac{1}{4} F^{\mu\nu} F_{\mu\nu} - F^{\mu\nu} \partial_\mu A_\nu) d^4x .
\] (4.20)
First, consider the variation of $F^{\mu \nu}$, now treated as an independent fundamental field. This gives

$$\delta S_{\text{f.o.}} = \frac{1}{4\pi} \int (\frac{1}{2} F_{\mu \nu} \delta F^{\mu \nu} - \delta F^{\mu \nu} \partial_{\mu} A_{\nu}) d^4 x,$$

$$= \frac{1}{4\pi} \int \left[ \frac{1}{2} F_{\mu \nu} \delta F^{\mu \nu} - \frac{1}{2} \delta F^{\mu \nu} (\partial_{\mu} A_{\nu} - \partial_{\nu} A_{\mu}) \right] d^4 x,$$  \hspace{1cm} (4.21)

where, in getting to the second line, we have used the fact that $F^{\mu \nu}$ is antisymmetric. The reason for doing this is that when we vary $F^{\mu \nu}$ we can take $\delta F^{\mu \nu}$ to be arbitrary, but it must still be antisymmetric. Thus it is helpful to force an explicit antisymmetrisation on the $\partial_{\mu} A_{\nu}$ that multiplies it, since the symmetric part automatically gives zero when contracted onto the antisymmetric $\delta F^{\mu \nu}$. Requiring $\delta S_{\text{f.o.}} = 0$ for arbitrary $\delta F^{\mu \nu}$ then implies the integrand must vanish. This gives, as promised, the equation of motion

$$F_{\mu \nu} = \partial_{\mu} A_{\nu} - \partial_{\nu} A_{\mu}.$$  \hspace{1cm} (4.22)

Varying $S_{\text{f.o.}}$ in (4.20) instead with respect to $A_{\mu}$, we get

$$\delta S_{\text{f.o.}} = -\frac{1}{4\pi} \int F_{\mu \nu} \partial_{\mu} \delta A_{\nu} d^4 x,$$

$$= \frac{1}{4\pi} \int (\partial_{\mu} F^{\mu \nu}) \delta A_{\nu} d^4 x,$$  \hspace{1cm} (4.23)

and hence requiring that the variation of $S_{\text{f.o.}}$ with respect to $A_{\mu}$ vanish gives the Maxwell field equation

$$\partial_{\mu} F^{\mu \nu} = 0$$  \hspace{1cm} (4.24)

again. Note that in this calculation, we immediately dropped the boundary term coming from the integration by parts, for the usual reason that we only allow variations that vanish on the boundary.

In practice, we shall usually use the previous, second-order, formalism.

### 4.3 Inclusion of sources

In general, the Maxwell field equation reads

$$\partial_{\mu} F^{\mu \nu} = -4\pi J^{\nu}.$$  \hspace{1cm} (4.25)

So far, we have seen that by varying the second-order action (4.16) with respect to $A_{\mu}$, we obtain

$$\delta S = \frac{1}{4\pi} \int \partial_{\mu} F^{\mu \nu} \delta A_{\nu} d^4 x.$$  \hspace{1cm} (4.26)
To derive the Maxwell field equation with a source current $J^\mu$, we can simply add a term to the action, to give

$$ S = \int \left( -\frac{1}{16\pi} F_{\mu\nu} F^{\mu\nu} + J^\mu A_\mu \right) d^4x. \quad (4.27) $$

Treating $J^\mu$ as independent of $A_\mu$, we therefore find

$$ \delta S = \int \left( \frac{1}{4\pi} \partial_\mu F^{\mu\nu} + J^\nu \right) \delta A_\nu d^4x, \quad (4.28) $$

and so requiring $\delta S = 0$ gives the Maxwell field equation (4.25) with the source on the right-hand side.

The form of the source current $J^\mu$ depends, of course, on the details of the situation one is considering. One might simply have a situation where $J^\mu$ is an externally-supplied source field. Alternatively, the source $J^\mu$ might itself be given dynamically in terms of some charged matter fields, or in terms of a set of moving point charges. Let us consider this possibility in more detail.

If there is a single point charge $q$ at the location $\vec{r}_0$, then it will be described by the charge density

$$ \rho = q \delta^3(\vec{r} - \vec{r}_0), \quad (4.29) $$

where the three-dimensional delta-function $\delta^3(\vec{r})$, with $\vec{r} = (x, y, z)$, means

$$ \delta^3(\vec{r}) = \delta(x) \delta(y) \delta(z). \quad (4.30) $$

If the charge is moving, so that its location at time $t$ is at $\vec{r} = \vec{r}_0(t)$, then of course we shall have

$$ \rho = q \delta^3(\vec{r} - \vec{r}_0(t)). \quad (4.31) $$

The 3-vector current will be given by multiplying this by the 3-velocity $\vec{v} = \frac{d\vec{r}_0(t)}{dt}$, giving

$$ \vec{J} = q \delta^3(\vec{r} - \vec{r}_0(t)) \frac{d\vec{r}_0}{dt}, \quad (4.32) $$

and so the 4-current is

$$ J^\mu = (\rho, \rho \vec{v}), \quad \text{where} \quad \vec{v} = \frac{d\vec{r}_0}{dt}, \quad (4.33) $$

and $\rho$ is given by (4.31). We can verify that this is the correct current vector, by checking that it properly satisfies the charge-conservation equation $\partial_\mu J^\mu = \frac{\partial \rho}{\partial t} + \partial_i J^i = 0$. Thus we have

$$ \frac{\partial \rho}{\partial t} = q \frac{\partial}{\partial t} \delta^3(\vec{r} - \vec{r}_0(t)) = q \frac{\partial}{\partial x_0^i} \delta^3(\vec{r} - \vec{r}_0(t)) \frac{dx_0^i}{dt}, $$

$$ = -q \frac{\partial}{\partial x^i} \delta^3(\vec{r} - \vec{r}_0(t)) \frac{dx_0^i}{dt} = -\partial_i \left( \rho \frac{dx_0^i}{dt} \right), $$

$$ = -\partial_i (\rho \vec{v}^i) = -\partial_i J_i. \quad (4.34) $$
Note that we used the chain rule for differentiation in the first line, and that in getting to the second line we used the result that $\frac{\partial f(x-y)}{\partial x} = f'(x-y) = -\frac{\partial f(x-y)}{\partial y}$ for any function $f$ with argument $(x-y)$ (where $f'$ denotes the derivative of $f$ with respect to its argument).

It is also useful to note that we can write (4.33) as

$$J^\mu = \rho \frac{dx_0^\mu}{dt},$$

where we simply define $x_0^\mu(t)$ with $\mu = 0$ to be $t$.

Note that the integral $\int J^\mu A_\mu$ for the point charge gives a contribution to the action that is precisely of the form we saw in equation (2.82):

$$\int J^\mu A_\mu d^4x = \int q \delta^3(\vec{r} - \vec{r}_0) \frac{dx_0^\mu}{dt} A_\mu d^3x dt,$$

$$= \int_{\text{path}} q \frac{dx_0^\mu}{dt} A_\mu(x_0^\nu) dt = q \int_{\text{path}} A_\mu dx^\mu. \quad (4.36)$$

Suppose now we have $N$ charges $q_a$, following paths $\vec{r}_a(t)$. Then the total charge density will be given by

$$\rho = \sum_{a=1}^{N} q_a \delta^3(\vec{r} - \vec{r}_a(t)). \quad (4.37)$$

Since we have alluded several times to the fact that $\partial_\mu J^\mu = 0$ is the equation of charge conservation, it is appropriate to examine this in a little more detail. The total charge $Q$ at time $t_1$ is given by integrating the charge density over the spatial 3-volume:

$$Q(t_1) = \int_{t=t_1} \rho d^3x = \int_{t=t_1} J^0 d\Sigma_0, \quad \text{where} \quad d\Sigma_0 = dx dy dz. \quad (4.38)$$

This can be written covariantly as

$$Q(t_1) = \int_{t=t_1} J^\mu d\Sigma_\mu, \quad (4.39)$$

where we define also\(^{15}\)

$$d\Sigma_1 = -dt dy dz, \quad d\Sigma_2 = -dt dz dx, \quad d\Sigma_3 = -dt dy dz. \quad (4.40)$$

Because the integral in (4.38) is defined to be over the 3-surface at constant $t$, it follows that the extra terms, for $\mu = 1, 2, 3$, in (4.39) do not contribute.

If we now calculate the charge at a later time $t_2$, and then take the difference between the two charges, we will obtain

$$Q(t_2) - Q(t_1) = \int_{\Sigma} J^\mu d\Sigma_\mu, \quad (4.41)$$

\(^{15}\)As has been mentioned previously, a careful derivation of the 4-dimensional divergence theorem $\int_V \partial_\alpha W^\alpha d^4x = \int_\Sigma W_\alpha d\Sigma_\alpha$ shows that the components of the 3-area element $d\Sigma_\mu$ should be as stated in eqns (4.38) and (4.40).
where Σ is the boundary of the 4-cylinder V, and it consists of the two “end caps” formed by the surfaces t = t_1 and t = t_2, and by the sides at spatial infinity. We are assuming the charges are confined to a finite region, and so the current J^\mu is zero on the sides of the cylinder, as the radius is sent to infinity.

By the 4-dimensional analogue of the divergence theorem we shall have

$$\int_\Sigma J^\mu d\Sigma_\mu = \int_V \partial_\mu J^\mu d^4x,$$

(4.42)

where V is the 4-volume bounded by Σ. Thus we have

$$Q(t_2) - Q(t_1) = \int_V \partial_\mu J^\mu d^4x = 0,$$

(4.43)

since \(\partial_\mu J^\mu = 0\). Thus we see that \(\partial_\mu J^\mu = 0\) implies that the total charge in an isolated finite region is independent of time.

Note that the equation of charge conservation implies the gauge invariance of the action. We have

$$S = \int \left( -\frac{1}{16\pi} F_{\mu\nu} F^{\mu\nu} + J^\mu A_\mu \right) d^4x,$$

(4.44)

and so under a gauge transformation \(A_\mu \rightarrow A_\mu + \partial_\mu \lambda\), we find

$$S \rightarrow \int \left( -\frac{1}{16\pi} F_{\mu\nu} F^{\mu\nu} + J^\mu A_\mu \right) d^4x + \int J^\mu \partial_\mu \lambda d^4x,$$

$$= S + \int J^\mu \partial_\mu \lambda d^4x = S + \int \partial_\mu (\lambda J^\mu) d^4x - \int \lambda \partial_\mu J^\mu d^4x,$$

$$= S + \int_\Sigma \lambda J^\mu d\Sigma_\mu.$$

(4.45)

As usual, Σ here is the 3-cylinder of infinite radius in the spatial directions, with endcaps at \(t = t_1\) and \(t = t_2\). The current \(J^\mu\) will vanish on the sides of the cylinder, since they are at spatial infinity and we take \(J^\mu\) to vanish there. If we restrict attention to gauge transformations that vanish at \(t = t_1\) and \(t = t_2\) then the surface integral will therefore give zero, and so \(S\) is unchanged. Even if \(\lambda\) is non-zero at \(t = t_1\) and \(t = t_2\) then the surface integral will just give a constant, independent of \(A_\mu\), and so the original and the gauge transformed actions will give the same equations of motion.

### 4.4 Energy density and energy flux

Here, we review the calculation of energy density and energy flux in the 3-dimensional language. After that, we shall give the more elegant 4-dimensional description.

Consider the two Maxwell equations

$$\nabla \times \vec{B} - \frac{\partial \vec{E}}{\partial t} = 4\pi \vec{J}, \quad \nabla \times \vec{E} + \frac{\partial \vec{B}}{\partial t} = 0.$$  

(4.46)
From these, we can deduce
\[
\vec{E} \cdot \frac{\partial \vec{E}}{\partial t} + \vec{B} \cdot \frac{\partial \vec{B}}{\partial t} = \vec{E} \cdot (\nabla \times \vec{B} - 4\pi \vec{J}) - \vec{B} \cdot (\nabla \times \vec{E}),
\]
\[
= \epsilon_{ijk}(E_i \partial_j B_k - B_i \partial_j E_k) - 4\pi \vec{J} \cdot \vec{E},
\]
\[
= -\epsilon_{ijk}(B_i \partial_j E_k + E_k \partial_j B_i) - 4\pi \vec{J} \cdot \vec{E},
\]
\[
= -\partial_j (\epsilon_{jki} E_k B_i) - 4\pi \vec{J} \cdot \vec{E},
\]
\[
= -\nabla \cdot (\vec{E} \times \vec{B}) - 4\pi \vec{J} \cdot \vec{E}.
\] (4.47)

We then define the Poynting vector
\[
\vec{S} \equiv \frac{1}{4\pi} \vec{E} \times \vec{B},
\] (4.48)
and so
\[
\frac{1}{2} \frac{\partial}{\partial t}(\vec{E}^2 + \vec{B}^2) = -4\pi \nabla \cdot \vec{S} - 4\pi \vec{J} \cdot \vec{E},
\] (4.49)
since \(\vec{E} \cdot \partial \vec{E}/\partial t = \frac{1}{2} \partial /\partial t(\vec{E}^2)\), etc.

We now assume that the \(\vec{E}\) and \(\vec{B}\) fields are confined to some finite region of space. Integrating (4.49) over all space, we obtain
\[
\int \vec{J} \cdot \vec{E} d^3 x + \frac{1}{8\pi} \frac{d}{dt} \int (\vec{E}^2 + \vec{B}^2) d^3 x = -\int \nabla \cdot \vec{S} d^3 x,
\]
\[
= -\int_{\Sigma} \vec{S} \cdot d\Sigma,
\]
\[
= 0.
\] (4.50)

We get zero on the right-hand side because, having used the divergence theorem to convert it to an integral over \(\Sigma\), the “sphere at infinity,” the integral vanishes since \(\vec{E}\) and \(\vec{B}\), and hence \(\vec{S}\), are assumed to vanish there.

If the current \(\vec{J}\) is assumed to be due to the motion of a set of charges \(q_a\) with 3-velocities \(\vec{v}_a\) and rest masses \(m_a\), we shall have from (4.32) and (3.12) that
\[
\int \vec{J} \cdot \vec{E} d^3 x = \sum_a q_a \vec{v}_a \cdot \vec{E}(\vec{r}_a) = \frac{d\mathcal{E}_{\text{mech}}}{dt},
\] (4.51)
where
\[
\mathcal{E}_{\text{mech}} = \sum_a m_a \gamma_a
\] (4.52)
is the total mechanical energy for the set of particles, as defined in (3.7). Note that here
\[
\gamma_a \equiv (1 - v_a^2)^{-1/2}.
\] (4.53)
Thus we conclude that
\[ \frac{d}{dt}(E_{\text{mech}} + \frac{1}{8\pi} \int (\vec{E}^2 + \vec{B}^2) d^3x) = 0. \]  
(4.54)

This is the equation of total energy conservation. It says that the sum of the total mechanical energy plus the energy contained in the electromagnetic fields is a constant. Thus we interpret
\[ W \equiv \frac{1}{8\pi} (\vec{E}^2 + \vec{B}^2) \]  
(4.55)
as the energy density of the electromagnetic field.

Returning now to equation (4.49), we can consider integrating it over just a finite volume \( V \), bounded by a closed 2-surface \( \Sigma \). We will have
\[ \frac{d}{dt}(E_{\text{mech}} + \int_V W d^3x) = -\int_\Sigma \vec{S} \cdot d\Sigma . \]  
(4.56)

We now know that the left-hand side should be interpreted as the rate of change of total energy in the volume \( V \) and so clearly, since the total energy must be conserved, we should interpret the right-hand side as the flux of energy passing through the boundary surface \( \Sigma \). Thus we see that the Poynting vector
\[ \vec{S} = \frac{1}{4\pi} \vec{E} \times \vec{B} \]  
(4.57)
is to be interpreted as the energy flux across the boundary; i.e. the energy per unit area per unit time.

### 4.5 Energy-momentum tensor

The discussion above was presented within the 3-dimensional framework. In this section we shall give a 4-dimensional spacetime description, which involves the introduction of the energy-momentum tensor. We shall begin with a rather general introduction. In order to simplify this discussion, we shall first describe the construction of the energy-momentum tensor for a scalar field \( \phi(x^\mu) \). When we then apply these ideas to electromagnetism, we shall need to make the rather simple generalisation to the case of a Lagrangian for the vector field \( A_\mu(x^\nu) \).

Recall that if we write the Maxwell tensor \( F_{\mu\nu} \) in terms of the 4-vector potential \( A_\mu \), namely \( F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu \), then the Bianchi identity \( \partial_\mu F_{\rho\nu} + \partial_\nu F_{\rho\mu} + \partial_\rho F_{\mu\nu} = 0 \) is automatically solved, and so the remaining content of the source-free Maxwell equations is just the field equation \( \partial^\mu F_{\mu\nu} = 0 \), which implies
\[ \Box A_\mu - \partial_\mu(\partial_\nu A^\nu) = 0 , \]  
(4.58)
where  is the d’Alembertian. If we choose to work in the Lorenz gauge, \( \partial_{\nu}A^\nu = 0 \), the field equation reduces to

\[
\Box A_\mu = 0. \tag{4.59}
\]

In the analogous, but simpler, example of a scalar field theory, we could consider the field equation

\[
\Box \phi = 0. \tag{4.60}
\]

A slightly more general possibility would be to add a “mass term” for the scalar field, and consider the equation of motion

\[
\Box \phi - m^2 \phi = 0, \tag{4.61}
\]

where \( m \) is a constant, describing the mass of the field. (As we shall discuss in detail later in the course, electromagnetism is described by a massless field. At the level of a particle description, this corresponds to the fact that the photon is a massless particle.)

The equation of motion (4.61) for the scalar field can be derived from an action. Consider the Lagrangian density

\[
\mathcal{L} = -\frac{1}{2}(\partial^\mu \phi)(\partial_\mu \phi) - \frac{1}{2}m^2 \phi^2. \tag{4.62}
\]

Varying the action \( S = \int \mathcal{L} d^4x \) with respect to \( \phi \), we obtain

\[
\delta S = \int \left(- (\partial^\mu \phi) \partial_\nu \delta \phi - m^2 \phi \delta \phi \right) d^4x,
\]

\[
= \int \left( \partial_\mu \partial^\mu \phi - m^2 \phi \right) \delta \phi d^4x, \tag{4.63}
\]

where we have, as usual, dropped the boundary term when performing the integration by parts to obtain the second line. Requiring \( \delta S = 0 \) for all possible \( \delta \phi \) consistent with the boundary conditions, we conclude that the quantity in the parentheses on the second line must vanish, and hence we arrive at the equation of motion (4.61).

We can now extend the discussion by considering an abstract Lagrangian density \( \mathcal{L} \) describing a scalar field \( \phi \). We shall assume that \( \mathcal{L} \) depends on \( \phi \), and on its first derivatives \( \partial_\nu \phi \), but that it has no explicit dependence\(^{16} \) on the spacetime coordinates \( x^\mu \):

\[
\mathcal{L} = \mathcal{L}(\phi, \partial_\nu \phi). \tag{4.64}
\]

The action is then given by

\[
S = \int \mathcal{L}(\phi, \partial_\nu \phi) d^4x. \tag{4.65}
\]

\(^{16}\)This is the analogue of a Lagrangian in classical mechanics that depends on the coordinates \( q_i \) and velocities \( \dot{q}_i \), but which does not have explicit time dependence. Energy is conserved in a system described by such a Lagrangian.
The Euler-Lagrange equations for the scalar field then follow from requiring that the action be stationary. Thus we have

$$\delta S = \int \left[ \frac{\partial L}{\partial \phi} \delta \phi + \frac{\partial L}{\partial (\partial_\nu \phi)} \partial_\nu \delta \phi \right] d^4 x,$$

$$= \int \left[ \frac{\partial L}{\partial \phi} \delta \phi - \partial_\nu \left( \frac{\partial L}{\partial (\partial_\nu \phi)} \right) \delta \phi \right] d^4 x + \int \Sigma \frac{\partial L}{\partial (\partial_\nu \phi)} \delta \phi d\Sigma_\nu,$$

$$= \int \left[ \frac{\partial L}{\partial \phi} \delta \phi - \partial_\nu \left( \frac{\partial L}{\partial (\partial_\nu \phi)} \right) \delta \phi \right] d^4 x,$$

(4.67)

where, in getting to the last line, we have as usual dropped the surface term integrated over the boundary cylinder $\Sigma$, since we shall insist that $\delta \phi$ vanishes on the endcaps of $\Sigma$ at $t = t_1$ and $t = t_2$, and that $\phi$ goes to zero sufficiently fast at spatial infinity. Thus the requirement that $\delta S = 0$ for all such $\delta \phi$ implies the Euler-Lagrange equations

$$\frac{\partial L}{\partial \phi} - \partial_\nu \left( \frac{\partial L}{\partial (\partial_\nu \phi)} \right) = 0.$$

(4.68)

Now consider the expression $\partial_\rho L = \partial L/\partial x^\rho$. Since we are assuming $L$ has no explicit dependence on the spacetime coordinates, it follows that $\partial_\rho L$ is given by the chain rule,

$$\partial_\rho L = \frac{\partial L}{\partial \phi} \partial_\rho \phi + \frac{\partial L}{\partial (\partial_\nu \phi)} \partial_\rho \partial_\nu \phi.$$

(4.69)

Now, using the Euler-Lagrange equations (4.68), we can write this as

$$\partial_\rho L = \partial_\nu \left( \frac{\partial L}{\partial (\partial_\nu \phi)} \right) \partial_\rho \phi + \frac{\partial L}{\partial (\partial_\nu \phi)} \partial_\rho \partial_\nu \phi,$$

(4.70)

and thus we have

$$\partial_\nu \left[ \frac{\partial L}{\partial (\partial_\nu \phi)} \partial_\rho \phi - \delta_\nu^\rho L \right] = 0.$$

(4.71)

We are therefore led to define the 2-index tensor

$$T_\rho^\nu \equiv - \frac{\partial L}{\partial (\partial_\nu \phi)} \partial_\rho \phi + \delta_\nu^\rho L,$$

(4.72)

17Note that $\partial L/\partial (\partial_\nu \phi)$ means taking the partial derivative of $L$ viewed as a function of $\phi$ and $\partial_\rho \phi$, with respect to $\partial_\nu \phi$. For example, if $L = -\frac{1}{2} (\partial_\nu \phi) (\partial_\nu \phi) + \frac{1}{2} m^2 \phi^2$, then

$$\partial L/\partial (\partial_\nu \phi) = -(\partial_\nu \phi) \frac{\partial (\partial_\nu \phi)}{\partial (\partial_\nu \phi)} = -(\partial_\nu \phi) \delta_\nu^\nu = -\partial_\nu \phi.$$

(4.66)

Of course, in this example $\partial L/\partial \phi$ is just equal to $-m^2 \phi$.

18These are the analogue of the Euler-Lagrange equations $\partial L/\partial q_i - d/dt (\partial L/\partial \dot{q}_i) = 0$ in particle mechanics for a system of particles with coordinates $q_i$ and velocities $\dot{q}_i$, derived from the Lagrangian $L = L(q_i, \dot{q}_i)$ by requiring stationarity of the action $S = \int L dt$. 68
which then satisfies
\[ \partial_\nu T^\nu_\rho = 0 \, . \] \hspace{1cm} (4.73)

\( T^\nu_\rho \) is called an energy-momentum tensor.

In the specific example of the Lagrangian density (4.62) for a free massive scalar field, we see that the energy-momentum tensor will be given by
\[ T^\nu_\rho = \partial^\nu \phi \partial_\rho \phi - \frac{1}{2} \eta^\mu_\rho (\partial^\nu \phi)(\partial_\sigma \phi) - \frac{1}{4} m^2 \phi^2 \delta^\nu_\rho \, . \] \hspace{1cm} (4.74)

Raising the lower index and relabelling indices, we therefore have
\[ T^\mu_\nu = \partial^\mu \phi \partial^\nu \phi - \frac{1}{2} \eta^\mu_\nu (\partial^\sigma \phi)(\partial_\sigma \phi) - \frac{1}{2} m^2 \phi^2 \eta^\mu_\nu \, . \] \hspace{1cm} (4.75)

It so happens in this example that \( T^\mu_\nu \) has turned out to be symmetric in the indices \( \mu \) and \( \nu \), but for a more general Lagrangian density this may not necessarily happen. We shall discuss this further below.

We saw previously that the equation \( \partial_\mu J^\mu = 0 \) for the 4-vector current density \( J^\mu \) implies that there is a conserved charge
\[ Q = \int_{t=const} \rho d^3 x = \int_{t=const} J^0 d^3 x \, . \] \hspace{1cm} (4.76)

The conservation of this charge, i.e. \( \frac{dQ}{dt} = 0 \) follows from
\[
\frac{dQ}{dt} = \int_{t=const} \partial_0 J^0 d^3 x = - \int_{t=const} \partial_i J^i d^3 x = - \int_S J^i dS_i ,
\]

\[ = 0 \, . \] \hspace{1cm} (4.77)

(The first step follows from \( 0 = \partial_\mu J^\mu = \partial_0 J^0 + \partial_i J^i \); the second step from the 3-dimensional divergence theorem; and the final step from the assumption that \( J^i \) vanishes on \( S \), the sphere at spatial infinity.)

By an identical argument, it follows that the equation \( \partial_\nu T^\nu_\rho = 0 \) implies that there is a conserved 4-vector:
\[ P^\mu \equiv \int_{t=const} T^\mu_0 d\Sigma_0 = \int_{t=const} T^\mu_\nu d\Sigma_\nu \, . \] \hspace{1cm} (4.78)

(Of course \( T^\mu_\nu = \eta^\mu_\rho T^\rho_\nu \) ) Thus we may check
\[
\frac{dP^\mu}{dt} = \partial_0 \int_{t=const} T^\mu_0 d^3 x = \int_{t=const} \partial_0 T^\mu_0 d^3 x = - \int_{t=const} \partial_i T^\mu_i d^3 x ,
\]

\[ = - \int_S T^\mu_i dS_i = 0 \, , \] \hspace{1cm} (4.79)
where in the last line we have used the divergence theorem to turn the integral into a 2-
dimensional integral over the boundary sphere \( S \) at infinity. This vanishes since we shall
assume the fields are zero at infinity.

Notice that \( T^{00} = -T_0^0 \) and from (4.72) we therefore have

\[
T^{00} = \frac{\partial L}{\partial \partial_0 \phi} \partial_0 \phi - \mathcal{L}. \tag{4.80}
\]

Now for a Lagrangian \( L = L(q^i, \dot{q}^i) \) we have the canonical momentum \( \pi_i = \partial L/\partial \dot{q}^i \), and the Hamiltonian

\[
H = \pi_i \dot{q}^i - L. \tag{4.81}
\]

Since there is no explicit time dependence, \( H \) is conserved, and is equal to the total energy
of the system. Comparing with (4.80), we can recognise that \( T^{00} \) is the energy density.

From (4.78) we therefore have that

\[
P^0 = \int T^{00} d^3x \tag{4.82}
\]

is the total energy. Since it is manifest from its construction that \( P^\mu \) is a 4-vector, and
since its 0 component is the energy, it follows that \( P^\mu \) is the 4-momentum.

The essential point in the discussion above is that \( P^\mu \) given in (4.78) should be conserved,
which requires \( \partial_\nu T^\rho_\nu = 0 \). The quantity \( T^\rho_\nu \) we constructed is not the unique tensor with
this property. We can define a new one, according to

\[
T\rho^\nu \longrightarrow T\rho^\nu + \partial_\sigma \psi\rho^\nu_\sigma, \tag{4.83}
\]

where \( \psi\rho^\nu_\sigma \) is an arbitrary tensor that is antisymmetric in its last two indices,

\[
\psi\rho^\nu_\sigma = -\psi\rho_\sigma^\nu. \tag{4.84}
\]

We shall take \( \psi\rho^\nu_\sigma \) to vanish at spatial infinity.

The antisymmetry implies, since partial derivatives commute, that

\[
\partial_\nu \partial_\sigma \psi\rho^\nu_\sigma = 0, \tag{4.85}
\]

and hence that the modified energy-momentum tensor defined by (4.83) is conserved too.

Furthermore, the modification to \( T^\rho_\nu \) does not alter \( P^\mu \), since, from (4.78), the extra term
will be

\[
P^{\mu}_{\text{extra}} = \int_{t=\text{const}} \partial_\sigma \psi^\mu_\nu_\sigma d\Sigma_\nu = \int_{t=\text{const}} \partial_\sigma \psi^{\mu 0}_\sigma d \Sigma_0, \]

\[
= \int_{t=\text{const}} \partial_0 \psi^{\mu 0}_\sigma d^3x + \int_{t=\text{const}} \partial_1 \psi^{\mu 0}_i d^3x = \int_{t=\text{const}} \partial_i \psi^{\mu 0}_i d^3x, \]

\[
= \int_S \psi^{\mu 0}_i dS_i = 0, \tag{4.86}
\]
where \( S \) is the sphere at spatial infinity. (The fact that the integral is over a \( t = \text{constant} \) surface means that only the \( d\Sigma_0 \) term contributes on the first line. The antisymmetry of \( \psi^{\mu\nu\sigma} \) in \( \nu \) and \( \sigma \) implies that \( \psi^{\mu 00} \) on the second line is zero.) The modification to \( P^\mu \) therefore vanishes since we are requiring that \( \psi_\rho^{\mu\nu\sigma} \) vanishes at spatial infinity.

The energy-momentum tensor can be pinned down uniquely by requiring that the four-dimensional angular momentum \( M^{\mu\nu} \), defined by

\[
M^{\mu\nu} = \int (x^\mu dP^\nu - x^\nu dP^\mu)
\]

be conserved, where \( dP^\mu = T^{\mu\rho} d\Sigma_\rho \), i.e. it is the integrand of the 4-momentum integral (4.78). First, let us make a remark about angular momentum in four dimensions. In three dimensions, we define the angular momentum 3-vector as \( \vec{L} = \vec{r} \times \vec{p} \). In other words,

\[
L_i = \epsilon_{ijk} x^j p^k = \frac{1}{2} \epsilon_{ijk} (x^j p^k - x^k p^j) = \frac{1}{2} \epsilon_{ijk} M^{jk},
\]

(4.88)

where \( M^{jk} \equiv x^j p^k - x^k p^j \). Thus taking \( M^{\mu\nu} = x^\mu p^\nu - x^\nu p^\mu \) in four dimensions is a plausible-looking generalisation. It should be noted that in a general dimension, angular momentum is described by a 2-index antisymmetric tensor; in other words, angular momentum is associated with a rotation in a 2-dimensional plane. It is a very special feature of three dimensions that we can use the \( \epsilon_{ijk} \) tensor to map the 2-index antisymmetric tensor \( M^{jk} \) into the vector \( L_i = \frac{1}{2} \epsilon_{ijk} M^{jk} \). Put another way, a very special feature of three dimensions is that a rotation in the \((x,y)\) plane can equivalently be described as a rotation about the orthogonal (i.e. \( z \)) axis. In higher dimensions, rotations do not occur around axes, but rather, in 2-planes. It is amusing, therefore, to try to imagine what the analogue of an axle is for a higher-dimensional car!

Getting back to our discussion of angular momentum and the energy-momentum tensor in four dimensions, we are defining

\[
M^{\mu\nu} = \int (x^\mu dP^\nu - x^\nu dP^\mu) = \int (x^\mu T^{\nu\rho} - x^\nu T^{\mu\rho}) d\Sigma_\rho .
\]

(4.89)

By analogous arguments to those we used earlier, this will be conserved (i.e. \( dM^{\mu\nu}/dt = 0 \)) if

\[
\partial_\rho (x^\mu T^{\nu\rho} - x^\nu T^{\mu\rho}) = 0 .
\]

(4.90)

Distributing the derivative, we therefore have the requirement that

\[
\delta_\rho T^{\nu\rho} + x^\mu \partial_\rho T^{\nu\rho} - \delta_\rho T^{\mu\rho} - x^\nu \partial_\rho T^{\mu\rho} = 0 ,
\]

(4.91)

and hence, since \( \partial_\rho T^{\mu\rho} = 0 \), that \( T^{\mu\nu} \) is symmetric,

\[
T^{\mu\nu} = T^{\nu\mu} .
\]

(4.92)
Using the freedom to add $\partial_\sigma \psi^{\mu\nu\sigma}$ to $T^{\mu\nu}$, as we discussed earlier, it is always possible to arrange for $T^{\mu\nu}$ to be symmetric. From now on, we shall assume that this is done.

We already saw that $P^\mu = \int T^{\mu0} d^3x$ is the 4-momentum, so $T^{00}$ is the energy density, and $T^{i0}$ is the 3-momentum density. Let us now look at the conservation equation $\partial_\nu T^{\mu\nu} = 0$ in more detail. Taking $\mu = 0$, we have

$$\frac{\partial}{\partial t} T^{00} + \partial_j T^{0j} = 0 .$$

Integrating over a spatial 3-volume $V$ with boundary $S$, we therefore find

$$\frac{d}{dt} \int_V T^{00} d^3x = - \int_V \partial_j T^{0j} d^3x = - \int_S T^{0j} dS_j .$$

The left-hand side is the rate of change of field energy in the volume $V$, and so we can deduce, from energy conservation, that $T^{0j}$ is the energy flux 3-vector. But since we are now working with a symmetric energy-momentum tensor, we have that $T^{0j} = T^{j0}$, and we already identified $T^{j0}$ as the 3-momentum density. Thus we have that

energy flux = momentum density.

From the $\mu = i$ components of $\partial_\nu T^{\mu\nu} = 0$, we have

$$\frac{\partial}{\partial t} T^{i0} + \partial_j T^{ij} = 0 ,$$

and so, integrating over the 3-volume $V$, we get

$$\frac{d}{dt} \int_V T^{i0} d^3x = - \int_V \partial_j T^{ij} d^3x = - \int_S T^{ij} dS_j .$$

The left-hand side is the rate of change of 3-momentum, and so we deduce that $T^{ij}$ is the 3-tensor of momentum flux density. It gives the $i$ component of 3-momentum that flows, per unit time, through the 2-surface perpendicular to the $x^j$ axis. $T^{ij}$ is sometimes called the 3-dimensional stress tensor.

### 4.6 Energy-momentum tensor for the electromagnetic field

Recall that for a scalar field $\phi$, the original construction of the energy-momentum tensor $T_{\rho\nu}$ (which we later modified by adding $\partial_\sigma \psi^{\rho\nu\sigma}$ where $\psi^{\rho\nu\sigma} = -\psi^{\rho\sigma\nu}$) was given by

$$T_{\rho\nu} = - \frac{\partial \mathcal{L}}{\partial (\partial_\nu \phi)} \partial_\rho \phi + \delta^\nu_\rho \mathcal{L} .$$

If we have a set of $N$ scalar fields $\phi_a$, then it is easy to see that the analogous conserved tensor is

$$T_{\rho\nu} = - \sum_{a=1}^N \frac{\partial \mathcal{L}}{\partial (\partial_\nu \phi_a)} \partial_\rho \phi_a + \delta^\nu_\rho \mathcal{L} .$$
A similar calculation shows that if we consider instead a vector field $A_\sigma$, with Lagrangian density $\mathcal{L}(A_\sigma, \partial_\nu A_\sigma)$, the construction will give a conserved energy-momentum tensor

$$T_\rho^\nu = -\frac{\partial \mathcal{L}}{\partial (\partial_\nu A_\sigma)} \partial_\rho A_\sigma + \delta_\rho^\nu \mathcal{L}.$$  

(4.100)

Let us apply this to the Lagrangian density for pure electrodynamics (without sources),

$$\mathcal{L} = -\frac{1}{16\pi} F_{\mu\nu} F^{\mu\nu}.$$  

(4.101)

We have

$$\delta \mathcal{L} = -\frac{1}{8\pi} F^{\mu\nu} \delta F_{\mu\nu} = -\frac{1}{4\pi} F^{\mu\nu} \partial_\mu \delta A_\nu,$$  

(4.102)

and so

$$\frac{\partial \mathcal{L}}{\partial (\partial_\mu A_\nu)} = -\frac{1}{4\pi} F^{\mu\nu}.$$  

(4.103)

Thus from (4.100) we find

$$T_\rho^\nu = \frac{1}{4\pi} F^{\nu\sigma} \partial_\rho A_\sigma - \frac{1}{16\pi} \delta_\rho^\nu F^\sigma_\lambda F_\sigma^\lambda,$$  

(4.104)

and so

$$T^{\mu\nu} = \frac{1}{4\pi} F^{\nu\sigma} \partial^\mu A_\sigma - \frac{1}{16\pi} \eta^{\mu\nu} F^\sigma_\lambda F_\sigma^\lambda.$$  

(4.105)

This expression is not symmetric in $\mu$ and $\nu$. However, following our previous discussion, we can add a term $\partial_\sigma \psi^{\mu\nu}$ to it, where $\psi^{\mu\nu} = -\psi^{\nu\mu}$, without upsetting the conservation condition $\partial_\nu T^{\mu\nu} = 0$. Specifically, we shall choose $\psi^{\mu\nu} = -1/(4\pi) A^\mu F^{\nu\sigma}$, and so

$$\partial_\sigma \psi^{\mu\nu} = \frac{1}{4\pi} \partial_\sigma (A^\mu F^{\nu\sigma}),$$

$$= -\frac{1}{4\pi} (\partial_\sigma A^\mu) F^{\nu\sigma} - \frac{1}{4\pi} A^\mu \partial_\sigma F^{\nu\sigma} = -\frac{1}{4\pi} (\partial_\sigma A^\mu) F^{\nu\sigma}.$$  

(4.106)

(the $\partial_\sigma F^{\nu\sigma}$ term drops as a consequence of the source-free field equation.) This leads to the new energy-momentum tensor

$$T^{\mu\nu} = \frac{1}{4\pi} F^{\nu\sigma} (\partial_\mu A_\sigma - \partial_\sigma A^\mu) - \frac{1}{16\pi} \eta^{\mu\nu} F^\sigma_\lambda F_\sigma^\lambda,$$  

(4.107)

or, in other words,

$$T^{\mu\nu} = \frac{1}{4\pi} \Big( F^\mu_\sigma F^{\nu\sigma} - \frac{1}{4} \eta^{\mu\nu} F^\sigma_\lambda F_\sigma^\lambda \Big).$$  

(4.108)

This is indeed manifestly symmetric in $\mu$ and $\nu$. From now on, it will be understood when we speak of the energy-momentum tensor for electrodynamics that this is the one we mean.

It is a straightforward exercise to verify directly, using the source-free Maxwell field equation and the Bianchi identity, that indeed $T^{\mu\nu}$ given by (4.108) is conserved, $\partial_\nu T^{\mu\nu} = 0$. Note that it has another simple property, namely that it is trace-free, in the sense that

$$\eta_{\mu\nu} T^{\mu\nu} = 0.$$  

(4.109)
This is easily seen from (4.108), as a consequence of the fact that \( \eta^{\mu\nu}\eta_{\mu\nu} = 4 \) in four dimensions. The trace-free property is related to a special feature of the Maxwell equations in four dimensions, known as conformal invariance.

Having obtained the energy-momentum tensor (4.108) for the electromagnetic field, it is instructive to look at its components from the three-dimensional point of view. First, recall that we showed earlier that

\[
F_{\sigma\lambda}F^{\sigma\lambda} = 2(\vec{B}^2 - \vec{E}^2) .
\]  

Then, we find

\[
T^{00} = \frac{1}{4\pi} (F^{0}_{\sigma}F^{0\sigma} - \frac{1}{3}\eta^{00}F_{\sigma\lambda}F^{\sigma\lambda}) ,
\]

\[
= \frac{1}{4\pi} (F^{0i}F^{0i} + \frac{1}{2}\vec{B}^2 - \frac{1}{2}\vec{E}^2) ,
\]

\[
= \frac{1}{4\pi} (\vec{E}^2 + \frac{1}{2}\vec{B}^2 - \frac{1}{2}\vec{E}^2) ,
\]

\[
= \frac{1}{8\pi} (\vec{E}^2 + \vec{B}^2) .
\]  

(4.111)

Thus \( T^{00} \) is equal to the energy density \( W \) that we introduced in (4.55).

Now consider \( T^{0i} \). Since \( \eta^{0i} = 0 \), we have

\[
T^{0i} = \frac{1}{4\pi} F^{0}_{\sigma}F^{i\sigma} = \frac{1}{4\pi} F^{0j}F^{ij} ,
\]

\[
= \frac{1}{4\pi} E_j\epsilon_{ijk}B_k = S_i ,
\]  

(4.112)

where \( \vec{S} = 1/(4\pi)\vec{E} \times \vec{B} \) is the Poynting vector introduced in (4.48). Thus \( T^{0i} \) is the energy flux. As we remarked earlier, since we now have \( T^{0i} = T^{i0} \), it can be equivalently interpreted as the 3-momentum density vector.

Finally, we consider the components \( T^{ij} \). We have

\[
T^{ij} = \frac{1}{4\pi} (F^{i\sigma}F^{j\sigma} - \frac{1}{4}\eta^{ij}2(\vec{B}^2 - \vec{E}^2)) ,
\]

\[
= \frac{1}{4\pi} (F^{i0}F^{j0} + F^{i}_{k}F^{jk} - \frac{1}{2}\delta_{ij}(\vec{B}^2 - \vec{E}^2)) ,
\]

\[
= \frac{1}{4\pi} (-E_iE_j + \epsilon_{ik\ell}\epsilon_{jkm}B_{\ell}B_{m} - \frac{1}{2}\delta_{ij}(\vec{B}^2 - \vec{E}^2)) ,
\]

\[
= \frac{1}{4\pi} (-E_iE_j + \delta_{ij}\vec{B}^2 - B_iB_j - \frac{1}{2}\delta_{ij}(\vec{B}^2 - \vec{E}^2)) ,
\]

\[
= \frac{1}{4\pi} (-E_iE_j - B_iB_j + \frac{1}{2}\delta_{ij}(\vec{E}^2 + \vec{B}^2)) .
\]  

(4.113)

To summarise, we have

\[
T^{\mu\nu} = \begin{pmatrix} T^{00} & T^{0j} \\ T^{i0} & \sigma_{ij} \end{pmatrix} = \begin{pmatrix} W & S_j \\ S_i & \sigma_{ij} \end{pmatrix} ,
\]  

(4.114)
where $W$ and $\vec{S}$ are the energy density and Poynting flux,

$$W = \frac{1}{8\pi} (\vec{E}^2 + \vec{B}^2), \quad \vec{S} = \frac{1}{4\pi} \vec{E} \times \vec{B},$$  \hfill (4.115)

and

$$\sigma_{ij} = -\frac{1}{4\pi} (E_i E_j + B_i B_j) + W \delta_{ij}. \hfill (4.116)$$

**Canonical Forms for $T_{\mu\nu}$:**

- Unless $\vec{E}$ and $\vec{B}$ are perpendicular and equal in magnitude, we can always choose a Lorentz frame where $\vec{E}$ and $\vec{B}$ are parallel at a point. (In the case that $\vec{E}$ and $\vec{B}$ are perpendicular (but unequal in magnitude), one or other of $\vec{E}$ or $\vec{B}$ will be zero, at the point, in the new Lorentz frame.)

Let the direction of $\vec{E}$ and $\vec{B}$ then be along $z$:

$$\vec{E} = (0,0,E), \quad \vec{B} = (0,0,B). \hfill (4.117)$$

Then we have $\vec{S} = 1/(4\pi) \vec{E} \times \vec{B} = 0$ and

$$\sigma_{11} = \sigma_{22} = W, \quad \sigma_{33} = -W, \quad \sigma_{ij} = 0 \text{ otherwise}, \hfill (4.118)$$

and so $T^{\mu\nu}$ is diagonal, given by

$$T^{\mu\nu} = \begin{pmatrix}
W & 0 & 0 & 0 \\
0 & W & 0 & 0 \\
0 & 0 & W & 0 \\
0 & 0 & 0 & -W
\end{pmatrix}, \hfill (4.119)$$

with $W = 1/(8\pi)(E^2 + B^2)$.

- If $\vec{E}$ and $\vec{B}$ are perpendicular and $|\vec{E}| = |\vec{B}|$ at a point, then at that point we can choose axes so that

$$\vec{E} = (E,0,0), \quad \vec{B} = (0,B,0) = (0,E,0). \hfill (4.120)$$

(We shall see later on that this $\vec{E} \cdot \vec{B} = 0$ and $|\vec{E}| = |\vec{B}|$ case arises for electromagnetic plane waves.) Then we have

$$W = \frac{1}{4\pi} E^2, \quad \vec{S} = (0,0,W),$$

$$\sigma_{11} = \sigma_{22} = 0, \quad \sigma_{33} = W, \quad \sigma_{ij} = 0 \text{ otherwise}, \hfill (4.121)$$

75
and therefore $T^{\mu\nu}$ is given by

$$T^{\mu\nu} = \begin{pmatrix} W & 0 & 0 & W \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ W & 0 & 0 & W \end{pmatrix}.$$  \hspace{2cm} (4.122)

### 4.7 Inclusion of massive charged particles

We now consider the energy-momentum tensor for a particle with rest mass $m$. We proceed by analogy with the construction of the 4-current density $J^\mu$ for charged non-interacting particles. Thus we define first a mass density, $\varepsilon$, for a point mass $m$ located at $\vec{r} = \vec{r}_0(t)$. This will simply be given by a 3-dimensional delta function, with strength $m$, located at the instantaneous position of the mass point:

$$\varepsilon = m\delta^3(\vec{r} - \vec{r}_0(t)).$$  \hspace{2cm} (4.123)

The energy density for the particle will then be its mass density times the corresponding $\gamma$ factor, i.e. $T^{00} = \varepsilon \gamma$, where $\gamma = (1 - v^2)^{-1/2}$ and $\vec{v} = \frac{d\vec{r}_0(t)}{dt}$ is the velocity of the particle. Since the coordinate time $t$ and the proper time $\tau$ in the frame of the particle are related, as usual, by $dt = \gamma d\tau$, we then have

$$T^{00} = \varepsilon \frac{dt}{d\tau}.$$  \hspace{2cm} (4.124)

The 3-momentum density will be $v^i$ times the energy density, and so

$$T^{0i} = \varepsilon \gamma \frac{dx^i}{dt} = \varepsilon \frac{dt}{d\tau} \frac{dx^i}{dt}.$$  \hspace{2cm} (4.125)

We can therefore write

$$T^{0\nu} = \varepsilon \frac{dt}{d\tau} \frac{dx^\nu}{dt} = \frac{dx^0}{d\tau} \frac{dx^\nu}{dt}.$$  \hspace{2cm} (4.126)

On general grounds of Lorentz covariance, it must therefore be that

$$T^{\mu\nu} = \frac{\varepsilon}{\gamma} \frac{dx^\mu}{d\tau} \frac{dx^\nu}{d\tau},$$  

By writing it as we have done in the second line here, it becomes manifest that $T^{\mu\nu}$ for the particle is symmetric in $\mu$ and $\nu$. 

76
Consider now a system consisting of a particle with mass $m$ and charge $q$, moving in an electromagnetic field. Clearly, since the particle interacts with the field, we should not expect either the energy-momentum tensor (4.108) for the electromagnetic field or the energy-momentum tensor (4.127) for the particle to be conserved separately. This is because energy, and momentum, is being exchanged between the particle and the field. We can expect, however, that the total energy-momentum tensor for the system, i.e. the sum of (4.108) and (4.127), to be conserved.

In order to distinguish clearly between the various energy-momentum tensors, let us define

$$T_{\text{tot.}}^{\mu\nu} = T_{\text{e.m.}}^{\mu\nu} + T_{\text{part.}}^{\mu\nu},$$

(4.128)

where $T_{\text{e.m.}}^{\mu\nu}$ and $T_{\text{part.}}^{\mu\nu}$ are the energy-momentum tensors for the electromagnetic field and the particle respectively:

$$T_{\text{e.m.}}^{\mu\nu} = \frac{1}{4\pi} \left( F_\mu^\sigma F_\nu^\sigma - \frac{1}{4} \eta_\mu^\nu F_\sigma^\lambda F_\sigma^\lambda \right),$$

$$T_{\text{part.}}^{\mu\nu} = \varepsilon \frac{dx^\mu}{d\tau} \frac{dx^\nu}{dt},$$

(4.129)

where $\varepsilon = m \delta^3(\vec{r} - \vec{r}_0(t))$.

Consider $T_{\text{e.m.}}^{\mu\nu}$ first. Taking the divergence, we find

$$\partial_\nu T_{\text{e.m.}}^{\mu\nu} = \frac{1}{4\pi} \left( \partial_\nu F_\mu^\sigma F_\nu^\sigma + F_\mu^\sigma \partial_\nu F_\nu^\sigma - \frac{1}{2} F_\sigma^\lambda \partial_\nu F_\sigma^\lambda \right),$$

$$= \frac{1}{4\pi} \left( \partial_\nu F_\mu^\sigma F_\nu^\sigma + F_\mu^\sigma \partial_\nu F_\nu^\sigma + \frac{1}{2} F_\nu^\sigma \partial_\nu F_\nu^\sigma - \frac{1}{2} F_\sigma^\lambda \partial_\nu F_\sigma^\lambda \right),$$

$$= \frac{1}{4\pi} \left( \partial_\nu F_\mu^\sigma F_\nu^\sigma - \frac{1}{2} F_\sigma^\lambda \partial_\nu F_\sigma^\lambda - \frac{1}{2} F_\lambda^\sigma \partial_\nu F_\lambda^\sigma + F_\mu^\sigma \partial_\nu F_\nu^\sigma \right),$$

$$= \frac{1}{4\pi} F_\sigma^\lambda \partial_\nu F_\sigma^\lambda,$n

(4.130)

In getting to the second line we used the Bianchi identity on the last term in the top line. The third line is obtained by swapping indices on a field strength in the terms with the $\frac{1}{2}$ factors, and this reveals that all except one term cancel, leading to the result on the fourth line. Finally, the result on the fifth lines follows after using the Maxwell equation of motion $\partial_\nu F_\nu^\sigma = -4\pi J^\sigma$. As expected, the energy-momentum tensor for the electromagnetic field by itself is not conserved when there are sources.

Now we want to show that this non-conservation is balanced by an equal and opposite non-conservation for the energy-momentum tensor of the particle, which is given in (4.129). We have

$$\partial_\nu T_{\text{part.}}^{\mu\nu} = \partial_\nu \left( \varepsilon \frac{dx^\nu}{dt} \right) \frac{dx^\mu}{dt} + \varepsilon \frac{dx^\nu}{dt} \partial_\nu \left( \frac{dx^\mu}{d\tau} \right).$$

(4.131)
The first term is zero. This can be seen from the fact that the calculation is identical to
the one which we used a while back in section 4.3 to show that the 4-current \( J^\mu = \rho \frac{dx^\mu}{dt} \) for
a charged particle is conserved. Thus we have

\[
\partial_\nu T_{\mu\nu}^{\text{part.}} = \varepsilon \frac{dx^\nu}{dt} \partial_\nu \left( \frac{dx^\mu}{d\tau} \right) = \varepsilon \frac{dx^\nu}{dt} \partial_\nu U^\mu,
\]

\[
= \varepsilon \frac{dU^\mu}{dt}.
\]

(4.132)

By the Lorentz force equation \( m \frac{dU^\mu}{d\tau} = q F_{\mu\nu} U^\nu \), we have

\[
\varepsilon \frac{dU^\mu}{d\tau} = \rho F_{\mu\nu} U^\nu = \rho F_{\mu\nu} \frac{dx^\nu}{d\tau},
\]

(4.133)

and so

\[
\varepsilon \frac{dU^\mu}{dt} = \rho F_{\mu\nu} \frac{dx^\nu}{dt} = F_{\mu\nu} J^\nu,
\]

(4.134)

since \( J^\mu = \rho \frac{dx^\mu}{dt} \). Thus we conclude that

\[
\partial_\nu T_{\mu\nu}^{\text{part.}} = F_{\mu\nu} J^\nu,
\]

(4.135)

and so, combining this with (4.130), we conclude that the total energy-momentum tensor
for the particle plus electromagnetic field, defined in (4.128) is conserved,

\[
\partial_\nu T_{\mu\nu}^{\text{tot.}} = 0.
\]

(4.136)

5 Coulomb’s Law

5.1 Potential of a point charge

Consider first a static point charge, for which the Maxwell equations therefore reduce to

\[
\vec{\nabla} \times \vec{E} = 0, \quad \vec{\nabla} \cdot \vec{E} = 4\pi \rho.
\]

(5.1)

The first equation implies, of course, that we can write

\[
\vec{E} = -\vec{\nabla} \phi,
\]

(5.2)

and then the second equation implies that \( \phi \) satisfies the Poisson equation

\[
\nabla^2 \phi = -4\pi \rho.
\]

(5.3)

If the point charge is located at the origin, and the charge is \( e \), then the charge density \( \rho \) is given by

\[
\rho = e \delta^3(\vec{r}).
\]

(5.4)
Away from the origin, (5.3) implies that $\phi$ should satisfy the Laplace equation,

$$\nabla^2 \phi = 0, \quad |\vec{r}| > 0. \quad (5.5)$$

Since the charge density (5.4) is spherically symmetric, we can assume that $\phi$ will be spherically symmetric too, $\phi(\vec{r}) = \phi(r)$, where $r = |\vec{r}|$. From $r^2 = x^j x^j$ we deduce, by acting with $\partial_i$, that

$$\partial_i r = \frac{x^i}{r}. \quad (5.6)$$

From this it follows by the chain rule that

$$\partial_i \phi = \phi' \partial_i r = \phi' \frac{x^i}{r}, \quad (5.7)$$

where $\phi' \equiv d\phi/dr$, and hence

$$\nabla^2 \phi = \partial_i \partial_i \phi = \partial_i \left( \phi' \frac{x^i}{r} \right) = \phi'' \frac{x^i x^i}{r^2} + \phi' \frac{\partial_i x^i}{r} + \phi' x^i \partial_i \frac{1}{r},$$

$$= \phi'' + \frac{2}{r} \phi'. \quad (5.8)$$

Thus the Laplace equation (5.5) can be written as

$$(r^2 \phi')' = 0, \quad r > 0, \quad (5.9)$$

which integrates to give

$$\phi = \frac{q}{r}, \quad (5.10)$$

where $q$ is a constant, and we have dropped an additive constant of integration by using the gauge freedom to choose $\phi(\infty) = 0$.

To determine the constant $q$, we integrate the Poisson equation (5.3) over the interior $V_R$ of a sphere of radius $R$ centred on the origin, and use the divergence theorem:

$$\int_{V_R} \nabla^2 \phi \, d^3 x = -4\pi e \int_{V_R} \delta^3(\vec{r}) \, d^3 x = -4\pi e,$$

$$= \int_{S_R} \vec{\nabla} \phi \cdot d\vec{S} = \int_{S_R} \partial_i \left( \frac{q}{r} \right) dS_i,$$

$$= -q \int_{S_R} \frac{x^i dS_i}{r^3} = -q \int_{S_R} \frac{n^i dS_i}{R^2}, \quad (5.11)$$

where $S_R$ is the surface of the sphere of radius $R$ that bounds the volume $V_R$, and $n^i \equiv x^i / r$ is the outward-pointing unit vector. Clearly we have

$$n^i dS_i = R^2 d\Omega, \quad (5.12)$$
where $d\Omega$ is the area element on the unit-radius sphere, and so
\[
-q \int_{S_R} \frac{n \cdot dS_i}{r^2} = -q \int d\Omega = -4\pi q, \tag{5.13}
\]
and so we conclude that $q$ is equal to $e$, the charge on the point charge at $r = 0$.

Note that if the point charge $e$ were located at $\vec{r}'$, rather than at the origin, then by trivially translating the coordinate system we will have the potential
\[
\phi(\vec{r}) = \frac{e}{|\vec{r} - \vec{r}'|}, \tag{5.14}
\]
and this will satisfy
\[
\nabla^2 \phi = -4\pi e \delta^3(\vec{r} - \vec{r}'). \tag{5.15}
\]

5.2 Electrostatic energy

In general, the energy density of an electromagnetic field is given by $W = 1/(8\pi)(\vec{E}^2 + \vec{B}^2)$. A purely electrostatic system therefore has a field energy $U$ given by
\[
U = \int W d^3x = \frac{1}{8\pi} \int \vec{E}^2 d^3x,
\]

\[
= -\frac{1}{8\pi} \int \vec{E} \cdot \vec{\nabla} \phi d^3x,
\]

\[
= -\frac{1}{8\pi} \int \vec{\nabla} \cdot (\vec{E} \phi) d^3x + \frac{1}{8\pi} \int (\vec{\nabla} \cdot \vec{E}) \phi d^3x,
\]

\[
= -\frac{1}{8\pi} \int S \phi \vec{E} \cdot d\vec{S} + \frac{1}{2} \int \rho \phi d^3x,
\]

\[
= \frac{1}{2} \int \rho \phi d^3x. \tag{5.16}
\]

Note that the surface integral over the sphere at infinity gives zero because the electric field is assumed to die away to zero there. Thus we conclude that the electrostatic field energy is given by
\[
U = \frac{1}{2} \int \rho \phi d^3x. \tag{5.17}
\]

We can apply this formula to a system of $N$ charges $q_a$, located at points $\vec{r}_a$, for which we shall have
\[
\rho = \sum_{a=1}^{N} q_a \delta^3(\vec{r} - \vec{r}_a). \tag{5.18}
\]

However, a naive application of (5.17) would give nonsense, since we find
\[
U = \frac{1}{2} \sum_{a=1}^{N} q_a \int \delta^3(\vec{r} - \vec{r}_a) \phi(\vec{r}) d^3x = \frac{1}{2} \sum_{a=1}^{N} q_a \phi(\vec{r}_a), \tag{5.19}
\]
where \( \phi(\vec{r}) \) is given by (5.14),

\[
\phi(\vec{r}) = \sum_{b=1}^{N} \frac{q_b}{|\vec{r} - \vec{r}_b|}.
\]

(5.20)

This means that (5.19) will give infinity since \( \phi(\vec{r}) \), not unreasonably, diverges at the location of each point charge.

This is the classic “self-energy” problem, which one encounters even for a single point charge. There is no totally satisfactory way around this in classical electromagnetism, and so one has to adopt a “fudge.” The fudge consists of observing that the true self-energy of a charge, whatever that might mean, is a constant. Naively, it appears to be an infinite constant, but that is clearly the result of making the idealised assumption that the charge is literally located at a single point. In any case, one can argue that the constant self-energy will not be observable, as far as energy-conservation considerations are concerned, and so one might as well just drop it for now. Thus the way to make sense of the ostensibly divergent energy (5.19) for the system of point charges is to replace \( \phi(\vec{r}_a) \), which means the potential at \( \vec{r} = \vec{r}_a \) due to all the charges, by \( \phi_a \), which is defined to be the potential at \( \vec{r} = \vec{r}_a \) due to all the charges except the charge \( q_a \) that is itself located at \( \vec{r} = \vec{r}_a \). Thus we have

\[
\phi_a \equiv \sum_{b \neq a} \frac{q_b}{|\vec{r}_a - \vec{r}_b|},
\]

(5.21)

and so (5.19) is now interpreted to mean that the total energy of the system of charges is

\[
U = \frac{1}{2} \sum_a \sum_{b \neq a} \frac{q_a q_b}{|\vec{r}_a - \vec{r}_b|}.
\]

(5.22)

5.3 Field of a uniformly moving charge

Suppose a charge \( e \) is moving with uniform velocity \( \vec{v} \) in the Lorentz frame \( S \). We may transform to a frame \( S' \), moving with velocity \( \vec{v} \) relative to \( S \), in which the charge is at rest. For convenience, we shall choose the origin of axes so that the charge is located at the origin of the frame \( S' \).

It follows that in the frame \( S' \), the field due to the charge can be described purely by the electric scalar potential \( \phi' \):

\[
\text{In the frame } S': \quad \phi' = \frac{e}{\vec{r}'}, \quad \vec{A}' = 0.
\]

(5.23)

(Note that the primes here all signify that the quantities are those of the primed frame \( S' \).)
We know that $A^\mu = (\phi, \vec{A})$ is a 4-vector, and so the components $A^\mu$ transform under Lorentz boosts in exactly the same way as the components of $x^\mu$. Thus we shall have

$$
\phi' = \gamma (\phi - \vec{v} \cdot \vec{A}) , \quad \vec{A}' = \vec{A} + \frac{\gamma - 1}{v^2} (\vec{v} \cdot \vec{A}) \vec{v} - \gamma \vec{v} \phi ,
$$

(5.24)

where $\gamma = (1 - v^2)^{-1/2}$. Clearly the inverse Lorentz transformation is obtained by switching the roles of the primed and unprimed fields and sending $\vec{v} \rightarrow -\vec{v}$, and so we shall have

$$
\phi = \gamma (\phi' + \vec{v} \cdot \vec{A}' ) , \quad \vec{A} = \vec{A}' + \frac{\gamma - 1}{v^2} (\vec{v} \cdot \vec{A}') \vec{v} + \gamma \vec{v} \phi' .
$$

(5.25)

From (5.23), we therefore find that the potentials in the frame $S$, in which the particle is moving with velocity $\vec{v}$, are given by

$$
\phi = \gamma \phi' = \frac{e \gamma}{r'}, \quad \vec{A} = \gamma \vec{v} \phi' = \frac{e \gamma \vec{v}}{r'}. 
$$

(5.26)

Note that we still have $r'$ appearing in the denominator, which we would now like to express in terms of the unprimed coordinates.

Suppose, for example, that we orient the axes so that $\vec{v}$ lies along the $x$ direction. Then we shall have

$$
x' = \gamma (x - vt), \quad y' = y, \quad z' = z ,
$$

(5.27)

and so

$$
r'^2 = x'^2 + y'^2 + z'^2 = \gamma^2 (x - vt)^2 + y^2 + z^2 ,
$$

(5.28)

$$
= \gamma^2 \left[ (x - vt)^2 + (1 - v^2) (y^2 + z^2) \right]. 
$$

(5.29)

It follows therefore from (5.26) that the scalar and 3-vector potentials in the frame $S$ are given by

$$
\phi = \frac{e}{R_s}, \quad \vec{A} = \frac{e \vec{v}}{R_s} ,
$$

(5.30)

where we have defined

$$
R_s^2 \equiv (x - vt)^2 + (1 - v^2)(y^2 + z^2) .
$$

(5.31)

The electric and magnetic fields can now be calculated in the standard way from $\phi$ and $\vec{A}$, as in (2.8). Alternatively, and equivalently, we can first calculate $\vec{E}'$ and $\vec{B}'$ in the primed frame, and then Lorentz transform these back to the unprimed frame. In the frame $S'$, we shall of course have

$$
\vec{E}' = \frac{e \vec{v}'}{r'^3}, \quad \vec{B}' = 0 .
$$

(5.32)
The transformation to the unprimed frame is then given by inverting the standard results (2.56) and (2.57) that express \( \vec{E}' \) and \( \vec{B}' \) in terms of \( \vec{E} \) and \( \vec{B} \). Again, this is simply achieved by interchanging the primed and unprimed fields, and sending \( \vec{v} \) to \( -\vec{v} \). This gives

\[
\begin{align*}
\vec{E} &= \gamma (\vec{E}' - \vec{v} \times \vec{B}') - \frac{\gamma - 1}{v^2} (\vec{v} \cdot \vec{E}') \vec{v}, \\
\vec{B} &= \gamma (\vec{B}' + \vec{v} \times \vec{E}') - \frac{\gamma - 1}{v^2} (\vec{v} \cdot \vec{B}') \vec{v},
\end{align*}
\]

and so from (5.32), we find that \( \vec{E} \) and \( \vec{B} \) in the frame \( S \) are given by

\[
\begin{align*}
\vec{E} &= e\gamma \vec{r}' / r'^3 - \frac{e - 1}{v^2} \vec{v} \times \vec{r}' / r'^3, \\
\vec{B} &= \gamma \vec{v} \times \vec{E}' = e\gamma \vec{v} \times \vec{r}' / r'^3.
\end{align*}
\]

Let us again assume that we orient the axes so that \( \vec{v} \) lies along the \( x \) direction. Then from the above we find that

\[
\begin{align*}
E_x &= \frac{e x'}{r'^3}, \quad E_y = \frac{e y'}{r'^3}, \quad E_z = \frac{e z'}{r'^3},
\end{align*}
\]

and so

\[
\begin{align*}
E_x &= \frac{e\gamma(x - vt)}{r'^3}, \quad E_y = \frac{e\gamma y}{r'^3}, \quad E_z = \frac{e\gamma z}{r'^3}.
\end{align*}
\]

Since the charge is located at the (uniformly moving) point \((vt, 0, 0)\) in the frame \( S \), it follows that the vector from the charge to the point \( \vec{r} = (x, y, z) \) is

\[
\vec{R} = (x - vt, y, z).
\]

From (5.36), we then find that the electric field is given by

\[
\vec{E} = \frac{e\gamma \vec{R}}{r'^3} = \frac{e(1 - v^2) \vec{R}}{R^3},
\]

where \( R^* \) was defined in (5.31).

If we now define \( \theta \) to be the angle between the vector \( \vec{R} \) and the \( x \) axis, then the coordinates \((x, y, z)\) of the observation point \( P \) will be such that

\[
y^2 + z^2 = R^2 \sin^2 \theta, \quad \text{where} \quad R^2 = |\vec{R}|^2 = (x - vt)^2 + y^2 + z^2.
\]

This implies, from (5.31), that

\[
R^* = R^2 - v^2(y^2 + z^2) = R^2(1 - v^2 \sin^2 \theta),
\]

and so the electric field due to the moving charge is

\[
\vec{E} = \frac{e\vec{R}}{R^3} \frac{1 - v^2}{(1 - v^2 \sin^2 \theta)^{3/2}}.
\]
For an observation point \( P \) located on the \( x \) axis, the electric field will be \( E_\parallel \) (parallel to the \( x \) axis), and given by setting \( \theta = 0 \) in (5.41). On the other hand, we can define the electric field \( E_\perp \) in the \((y, z)\) plane (corresponding to \( \theta = \pi/2 \)). From (5.41) we therefore have
\[
E_\parallel = \frac{e(1 - v^2)}{R^2}, \quad E_\perp = \frac{e(1 - v^2)^{-1/2}}{R^2}.
\]
(5.42)
Note that \( E_\parallel \) has the smallest magnitude, and \( E_\perp \) has the largest magnitude, that \( \vec{E} \) attains as a function of \( \theta \).

When the velocity is very small, the magnitude of the electric field is (as one would expect) more or less independent of \( \theta \). However, as \( v \) approaches 1 (the speed of light), we find that \( E_\parallel \) decreases to zero, while \( E_\perp \) diverges. Thus for \( v \) near to the speed of light the electric field is very sharply peaked around \( \theta = \pi/2 \). If we set
\[
\theta = \frac{\pi}{2} - \psi,
\]
then
\[
|\vec{E}| = \frac{e(1 - v^2)}{R^2(1 - v^2 \cos^2 \psi)^{3/2}} \approx \frac{e(1 - v^2)}{(1 - v^2 + \frac{1}{2} \psi^2)^{3/2}}
\]
(5.44)
if \( v \approx 1 \). Thus the angular width of the peak is of the order of
\[
\psi \sim \sqrt{1 - v^2}.
\]
(5.45)

We saw previously that the magnetic field in the frame \( S \) is given by \( \vec{B} = \gamma \vec{v} \times \vec{E}' \). From (5.34) we have \( \vec{v} \times \vec{E} = \gamma \vec{v} \times \vec{E}' \), and so therefore
\[
\vec{B} = \vec{v} \times \vec{E} = \frac{e(1 - v^2)\vec{v} \times \vec{R}}{R^3}.
\]
(5.46)
Note that if \( |\vec{v}| \ll 1 \) we get the usual non-relativistic expressions
\[
\vec{E} \approx \frac{e\vec{R}}{R^3}, \quad \vec{B} \approx \frac{e\vec{v} \times \vec{R}}{R^3}.
\]
(5.47)

5.4 Motion of a charge in a Coulomb potential

We shall consider a particle of mass \( m \) and charge \( e \) moving in the field of a static charge \( Q \). The classic "Newtonian" result is very familiar, with the orbit of the particle being a conic section; an ellipse, a parabola or a hyperbola, depending on the charges and the orbital parameters. In this section we shall consider the fully relativistic problem, when the velocity of the particle is not necessarily small compared with the speed of light. Note that we shall be assuming here that the charge \( Q \) is fixed, at the origin. In practice, if \( Q \)
represented a nucleus and an electron, the two objects would move around their common centre of gravity. Allowing for this would be a straightforward, if rather tedious, minor distraction from the main points to be studied here, and so we shall make the idealisation that the centre of attraction or repulsion, namely the charge \( Q \), is fixed.

The Lagrangian for the system is given by (2.84), with \( \phi = \frac{Q}{r} \) and \( \vec{A} = 0 \):

\[
L = -m(1 - \dot{x}^i \dot{x}^i)^{1/2} - \frac{eQ}{r},
\]

(5.48)

where \( \dot{x}^i = \frac{dx^i}{dt} \), and \( r^2 = x^i x^i \). The charges occur in the combination \( eQ \) throughout the calculation, and so for convenience we shall define

\[
\kappa \equiv eQ.
\]

(5.49)

Note that if \( \kappa > 0 \) the force between the charges will be repulsive, while if \( \kappa < 0 \) it will be attractive.

It is convenient to introduce spherical polar coordinates in the standard way,

\[
x = r \sin \theta \cos \varphi, \quad y = r \sin \theta \sin \varphi, \quad z = r \cos \theta,
\]

(5.50)

and then the Lagrangian becomes

\[
L = -m(1 - \dot{r}^2 - r^2 \dot{\theta}^2 - r^2 \sin^2 \theta \dot{\varphi}^2)^{1/2} - \frac{\kappa}{r}.
\]

(5.51)

The Lagrangian is of the form \( L = L(q_i, \dot{q}_i) \) for coordinates \( q_i \) and velocities \( \dot{q}_i \). The Euler-Lagrange equations are

\[
\frac{\partial L}{\partial q_i} - \frac{d}{dt} \left( \frac{\partial L}{\partial \dot{q}_i} \right) = 0.
\]

(5.52)

Note that if the Lagrangian is independent of a particular coordinate, say \( q_j \) for some particular value \( j \), then there is an associated conserved quantity \( \partial L / \partial \dot{q}_j \):

\[
\frac{d}{dt} \frac{\partial L}{\partial \dot{q}_j} = 0.
\]

(5.53)

(A coordinate that enters the Lagrangian only through its velocity is sometimes known as an ignorable coordinate.)

The Euler-Lagrange equation for \( \theta \) gives

\[
r^2 \sin \theta \cos \varphi \dot{\varphi}^2 (1 - \dot{r}^2 - r^2 \dot{\theta}^2 - r^2 \sin^2 \theta \dot{\varphi}^2)^{-1/2} - \frac{d}{dt} \left( r^2 \dot{\theta}(1 - \dot{r}^2 - r^2 \dot{\theta}^2 - r^2 \sin^2 \theta \dot{\varphi}^2)^{-1/2} \right) = 0.
\]

(5.54)

It can be seen that a solution to this equation is to take \( \theta = \frac{\pi}{2} \), and \( \dot{\varphi} = 0 \). In other words, if the particle starts out moving in the \( \theta = \frac{\pi}{2} \) plane (i.e. the \( (x, y) \) plane at \( z = 0 \)), it will
remain in this plane. This is just the familiar result that the motion of a particle moving in a central force lies in a plane. We may therefore assume now, without loss of generality, that \( \theta = \frac{\pi}{2} \) for all time. We are left with just \( r \) and \( \varphi \) as polar coordinates in the \((x,y)\) plane. The Lagrangian for the reduced system, where we consistently can set \( \theta = \frac{\pi}{2} \), is then simply

\[
L = -m(1 - \dot{r}^2 - r^2 \dot{\varphi}^2)^{1/2} - \frac{\kappa}{r}.
\]  

(5.55)

We note that \( \frac{\partial L}{\partial \dot{\varphi}} = 0 \), and so there is a conserved quantity

\[
\frac{\partial L}{\partial \dot{\varphi}} = m r^2 \dot{\varphi} (1 - \dot{r}^2 - r^2 \dot{\varphi}^2)^{-1/2} = \ell,
\]

(5.56)

where \( \ell \) is a constant. Since \((1 - \dot{r}^2 - r^2 \dot{\varphi}^2)^{-1/2} = (1 - \dot{x}^2 \dot{y}^2)^{-1/2} = \gamma \), we simply have

\[
m \gamma r^2 \dot{\varphi} = \ell.
\]

(5.57)

In fact \( \ell \) is the angular momentum of the particle, measured with respect to the origin. Note that we can also write (5.56 as

\[
m r^2 \frac{d\varphi}{d\tau} = \ell,
\]

(5.58)

since coordinate time \( t \) and proper time \( \tau \) are related by \( d\tau = dt/\gamma \).

Since the Lagrangian does not depend explicitly on \( t \), the total energy \( \mathcal{E} \) is also conserved. Thus we have that

\[
\mathcal{E} = H = \sqrt{p^2 + m^2 + \frac{\kappa}{r}}
\]

(5.59)

is a constant. Here,

\[
p^2 = m^2 \gamma^2 \dot{r}^2 = m^2 \gamma^2 \dot{r}^2 + m^2 \gamma^2 r^2 \dot{\varphi}^2,
\]

\[
= m^2 \left( \frac{dr}{d\tau} \right)^2 + m^2 r^2 \left( \frac{d\varphi}{d\tau} \right)^2,
\]

(5.60)

since, as usual, coordinate time and proper time are related by \( d\tau = dt/\gamma \).

We therefore have

\[
\left( \mathcal{E} - \frac{\kappa}{r} \right)^2 = p^2 + m^2 = m^2 \left( \frac{dr}{d\tau} \right)^2 + m^2 r^2 \left( \frac{d\varphi}{d\tau} \right)^2 + m^2.
\]

(5.61)

We now perform the standard change of variables in orbit calculations, and let

\[
r = \frac{1}{u}.
\]

(5.62)

This implies

\[
\frac{dr}{d\tau} = -\frac{1}{u^2} \frac{du}{d\tau} = -\frac{1}{u^2} \frac{du}{d\varphi} \frac{d\varphi}{d\tau} = -\frac{\ell}{m} u',
\]

(5.63)
where we have used (5.58) and also we have defined

\[ u' \equiv \frac{du}{d\phi}. \] (5.64)

It now follows that (5.61) becomes

\[ (E - \kappa u)^2 = \ell^2 u'^2 + \ell^2 u^2 + m^2. \] (5.65)

This ordinary differential equation can be solved in order to find \( u \) as a function of \( \phi \), and hence \( r \) as a function of \( \phi \). The solution determines the shape of the orbit of the particle around the fixed charge \( Q \).

Rewriting (5.65) as

\[ \ell^2 u'^2 = \left( u \sqrt{\kappa^2 - \ell^2} - \frac{\kappa E}{\sqrt{\kappa^2 - \ell^2}} \right)^2 - m^2 - \frac{E^2 \ell^2}{\kappa^2 - \ell^2}, \] (5.66)

we see that it is convenient to make a change of variable from \( u \) to \( w \), defined by

\[ u \sqrt{\kappa^2 - \ell^2} - \frac{\kappa E}{\sqrt{\kappa^2 - \ell^2}} = \pm \sqrt{m^2 + \frac{E^2 \ell^2}{\kappa^2 - \ell^2}} \cosh w, \] (5.67)

where the + sign is chosen if \( \kappa < 0 \) (attractive potential, since \( \kappa = eQ/Q \)), and the − sign if \( \kappa > 0 \) (repulsive potential). We can then integrate (5.66), to obtain

\[ \ell \sqrt{\kappa^2 - \ell^2} w = \phi. \] (5.68)

Here, we have made a convenient choice, without loss of generality, for the additive constant of integration, by absorbing it into a choice of origin for the azimuthal angle \( \phi \). Thus we have

\[ \sqrt{\kappa^2 - \ell^2} u = \pm \sqrt{m^2 + \frac{E^2 \ell^2}{\kappa^2 - \ell^2}} \cosh \left[ (\frac{\kappa^2}{\ell^2} - 1)^{1/2} \phi \right] + \frac{\kappa E}{\sqrt{\kappa^2 - \ell^2}}. \] (5.69)

In other words, the orbit is given, in terms of \( r = r(\phi) = 1/u(\phi) \), by

\[ \frac{\kappa^2 - \ell^2}{r} = \pm \sqrt{E^2 \ell^2 + m^2(\kappa^2 - \ell^2)} \cosh \left[ (\frac{\kappa^2}{\ell^2} - 1)^{1/2} \phi \right] + \kappa E. \] (5.70)

The solution (5.70) is presented for the case where \( |\ell| < |\kappa| \). If instead \( |\ell| > |\kappa| \) we should instead complete the square as given in footnote 19, leading us to make a cosine substitution rather than a hyperbolic cosine, and the solution becomes

\[ \frac{\ell^2 - \kappa^2}{r} = \sqrt{E^2 \ell^2 + m^2(\ell^2 - \kappa^2)} \cos \left[ (\frac{\kappa^2}{\ell^2} - 1)^{1/2} \phi \right] - \kappa E. \] (5.71)

19This “completing of the square” is appropriate for the case where \( |\ell| < |\kappa| \). If instead \( |\ell| > |\kappa| \), we would write

\[ \ell^2 u'^2 = -\left( u \sqrt{\ell^2 - \kappa^2} + \frac{\kappa E}{\sqrt{\ell^2 - \kappa^2}} \right)^2 - m^2 + \frac{E^2 \ell^2}{\ell^2 - \kappa^2}. \]
Finally, if \(|\ell| = |\kappa|\), we see that equation (5.65) leads to the solution
\[
\frac{2\kappa \mathcal{E}}{r} = \mathcal{E}^2 - m^2 - \mathcal{E}^2 \varphi^2.
\] (5.72)

The situation described above for relativistic orbits should be contrasted with what happens in the non-relativistic case. In this limit, the Lagrangian (after restricting to motion in the \((x, y)\) plane again) is simply given by
\[
L = \frac{1}{2} m (\dot{r}^2 + r^2 \dot{\varphi}^2) - \frac{\kappa}{r}.
\] (5.73)

Note that this can be obtained from the relativistic Lagrangian (5.55) we studied above, by taking \(\dot{r}\) and \(r \dot{\varphi}\) to be small compared to 1 (the speed of light), and then expanding the square root to quadratic order in velocities. As discussed previously, one can ignore the leading-order term \(-m\) in the expansion, since this is just a constant (the rest-mass energy of the particle) and so it does not enter in the Euler-Lagrange equations. The analysis of the Euler-Lagrange equations for the non-relativistic Lagrangian (5.73) is a standard one. There energy \(E\) and angular momentum \(\ell\) are conserved, and given by
\[
E = \frac{1}{2} m (\dot{r}^2 + r^2 \dot{\varphi}^2) + \frac{\kappa}{r}, \quad \ell = mr^2 \dot{\varphi}.
\] (5.74)

Substituting the latter into the former give the standard radial equation, whose solution implies orbits given by
\[
\frac{1}{r} = -\frac{mk}{\ell^2} \left( \sqrt{1 + \frac{2E\ell^2}{mk^2} \cos \varphi + 1} \right).
\] (5.75)

If \(\kappa \equiv \epsilon Q < 0\) (attractive potential), solutions exist for \(E \geq -\frac{mk^2}{2\ell^2}\). Closed orbits occur when \(-\frac{mk^2}{2\ell^2} \leq E < 0\); these are elliptical generically, and circular if the lower bound is saturated. The orbits are parabolic if \(E = 0\) and hyperbolic if \(E > 0\). If \(\kappa > 0\) (repulsive potential), solutions exist if \(E > 0\) and they are all hyperbolic.

The key difference in the relativistic case is that the orbits do not have a \(2\pi\) periodicity in \(\varphi\), even when \(|\ell| > |\kappa|\), as in (5.71), for which the radius \(r\) is a trigonometric function of \(\varphi\). The reason for this is that the argument of the trigonometric function is
\[
\left(1 - \frac{\kappa^2}{\ell^2}\right)^{1/2} \varphi,
\] (5.76)
and so \(\varphi\) has to increase through an angle \(\Delta \varphi\) given by
\[
\Delta \varphi = 2\pi \left(1 - \frac{\kappa^2}{\ell^2}\right)^{-1/2}
\] (5.77)
before the cosine completes one cycle. If we assume that \(|\kappa/\ell|\) is small compared with 1, then the shape of the orbit is still approximately like an ellipse, except that the “perihelion”
of the ellipse advances by an angle

$$\delta \varphi = 2\pi \left[ \left( 1 - \frac{\kappa^2}{\ell^2} \right)^{-1/2} - 1 \right] \approx \frac{\pi \kappa^2}{\ell^2}$$  \hspace{1cm} (5.78)$$

per orbit. Generically, the orbits are not closed, although they will be in the special case that $$\left( 1 - \frac{\kappa^2}{\ell^2} \right)^{-1/2}$$ is rational.

The fact that the major axis of the ellipse remains fixed in the non-relativistic case is a reflection of the fact that there is an additional “hidden” symmetry in the non-relativistic system, which is broken by the relativistic corrections. Specifically, there is a conserved quantity called the Runge-Lenz vector in the non-relativistic theory of a particle of mass $$m$$ moving in a central $$1/r^2$$ force $$\vec{F} = \kappa \vec{r}/r^3$$. It is given by

$$\vec{W} = \vec{p} \times \vec{L} + \frac{m \kappa \vec{r}}{r}$$,  \hspace{1cm} (5.79)$$

where $$\vec{L} = \vec{r} \times \vec{p}$$ is the angular momentum of the particle about the force centre. It is straightforward to verify that the equations of motion following from the non-relativistic Lagrangian $$L = \frac{1}{2} m \dot{x}_i^2 + \kappa/r$$ imply that $$d\vec{W}/dt = 0$$. In the case of an elliptical orbit (i.e. $$\kappa < 0$$ and $$-\frac{m \kappa^2}{2 \ell^2} < E < 0$$), the Runge-Lenz vector points along the major axis of the ellipse.

Going back to the full relativistic discussion, if $$|\ell| \leq |\kappa|$$ and if $$\kappa < 0$$ (which means $$eQ < 0$$ and hence an attractive force between the charges), we see from the solution (5.70) that the particle spirals inwards and eventually reaches $$r = 0$$ within a finite time. This can never happen in the non-relativistic case; the orbit of the particle can never reach the origin at $$r = 0$$, unless the angular momentum $$\ell$$ is exactly zero. The reason for this is that in the non-relativistic case the centrifugal potential term $$\ell^2/r^2$$ always tends to throw the particle away from the origin if $$r$$ tries to get too small. By contrast, in the relativistic case the effect of the centrifugal term is reduced at small $$r$$, and it cannot prevent the collapse of the orbit to $$r = 0$$. This can be seen by looking at the conserved quantity $$\mathcal{E}$$ in the fully relativistic analysis, which, from our discussion above, can be written as

$$\mathcal{E} = \left( m^2 + m^2 \left( \frac{dr}{d\tau} \right)^2 + \frac{\ell^2}{r^2} \right)^{1/2} + \frac{\kappa}{r}.$$  \hspace{1cm} (5.80)$$

First, consider the non-relativistic limit, for which the rest-mass term dominates inside the square root:

$$\mathcal{E} \approx m + \frac{1}{2} m \left( \frac{dr}{d\tau} \right)^2 + \frac{\ell^2}{2mr^2} + \frac{\kappa}{r}.$$  \hspace{1cm} (5.81)$$

In other words, written in terms of the non-relativistic energy $$E$$ of eqn (??) we have

$$E = \frac{1}{2} m r^2 + \frac{\ell^2}{2mr^2} + \frac{\kappa}{r}.$$  \hspace{1cm} (5.82)$$
Here, we see that even if \( \kappa < 0 \) (an attractive force), the repulsive centrifugal term always wins over the attractive charge term \( \kappa/r \) at small enough \( r \).

On the other hand, if we keep the full relativistic expression (5.80), then at small enough \( r \) the competition between the centrifugal term and the charge term becomes evenly matched, in the sense that each has the same \( 1/r \) power-law behaviour

\[
\mathcal{E} \approx \frac{|\ell|}{r} + \frac{\kappa}{r},
\]

and clearly if \( \kappa < -|\ell| \) the attraction between the charges wins the contest.

6 Electromagnetic Waves

6.1 Wave equation

As discussed at the beginning of the course (see section 1.1), Maxwell’s equations admit wave-like solutions. These solutions can exist in free space, in a region where there are no source currents, for which the equations take the form

\[
\nabla \cdot \vec{E} = 0, \quad \nabla \times \vec{B} - \frac{\partial \vec{E}}{\partial t} = 0, \\
\nabla \cdot \vec{B} = 0, \quad \nabla \times \vec{E} + \frac{\partial \vec{B}}{\partial t} = 0.
\]

(6.1)

As discussed in section 1.1, taking the curl of the \( \nabla \times \vec{E} \) equation, and using the \( \nabla \times \vec{B} \) equation, one finds

\[
\nabla^2 \vec{E} - \frac{\partial^2 \vec{E}}{\partial t^2} = 0, 
\]

(6.2)

and similarly,

\[
\nabla^2 \vec{B} - \frac{\partial^2 \vec{B}}{\partial t^2} = 0.
\]

(6.3)

Thus each component of \( \vec{E} \) and each component of \( \vec{B} \) satisfies d’Alembert’s equation

\[
\nabla^2 f - \frac{\partial^2 f}{\partial t^2} = 0.
\]

(6.4)

This can, of course, be written as

\[
\Box f \equiv \partial^\mu \partial_\mu f = 0,
\]

(6.5)

which shows that d’Alembert’s operator is Lorentz invariant.

The wave equation (6.4) admits plane-wave solutions, where \( f \) depends on \( t \) and on a single linear combination of the \( x, y \) and \( z \) coordinates. By choosing the orientation of the
axes appropriately, we can make this linear combination become simply $x$. Thus we may seek solutions of (6.4) of the form $f = f(t, x)$. The function $f$ will then satisfy
\[
\frac{\partial^2 f}{\partial x^2} - \frac{\partial^2 f}{\partial t^2} = 0,
\]
(6.6)
which can be written in the factorised form
\[
\left(\frac{\partial}{\partial x} - \frac{\partial}{\partial t}\right)\left(\frac{\partial}{\partial x} + \frac{\partial}{\partial t}\right)f(t, x) = 0.
\]
(6.7)
Now introduce “light-cone coordinates”
\[
u = x - t, \quad \nu = x + t.
\]
(6.8)
We see that
\[
\frac{\partial}{\partial x} = \frac{\partial}{\partial u} + \frac{\partial}{\partial v}, \quad \frac{\partial}{\partial t} = -\frac{\partial}{\partial u} + \frac{\partial}{\partial v},
\]
(6.9)
and so (6.7) becomes
\[
\frac{\partial^2 f}{\partial u \partial v} = 0.
\]
(6.10)
The general solution to this is
\[
f = f_+(u) + f_-(v) = f_+(x - t) + f_-(x + t),
\]
(6.11)
where $f_+$ and $f_-$ are arbitrary functions.

The functions $f_\pm$ determine the profile of a wave-like disturbance that propagates at the speed of light (i.e. at speed 1). In the case of a wave described by $f_+(x - t)$, the disturbance propagates at the speed of light in the positive $x$ direction. This can be seen from the fact that if we sit at a given point on the profile (i.e. at a fixed value of the argument of the function $f_+$), then as $t$ increases the $x$ value must increase too. This means that the disturbance moves, with speed 1, along the positive $x$ direction. Likewise, a wave described by $f_-(x + t)$ moves in the negative $x$ direction as time increases.

More generally, we can consider a plane-wave disturbance moving along the direction of a unit 3-vector $\vec{n}$:
\[
f(t, \vec{r}) = f_+ (\vec{n} \cdot \vec{r} - t) + f_- (\vec{n} \cdot \vec{r} + t).
\]
(6.12)
The $f_+$ wave moves in the direction of $\vec{n}$ as $t$ increases, while the $f_-$ wave moves in the direction of $-\vec{n}$. The previous case of propagation along the $x$ axis, corresponds to taking $\vec{n} = (1, 0, 0)$.

Let us now return to the discussion of electromagnetic waves. Following the discussion above, there will exist plane-wave solutions of (6.2), propagating along the $\vec{n}$ direction, of the form
\[
\vec{E}(\vec{r}, t) = \vec{E}(\vec{n} \cdot \vec{r} - t).
\]
(6.13)
That is, $\vec{E}$ is a function of the single argument $(\vec{n} \cdot \vec{r} - t)$.

From the Maxwell equation $\partial \vec{B} / \partial t = -\vec{\nabla} \times \vec{E}$, we shall therefore have

$$\frac{\partial B_i}{\partial t} = -\epsilon_{ijk} \partial_j E_k(n_\ell x_\ell - t),$$

$$= -\epsilon_{ijk} n_j E'_k(n_\ell x_\ell - t), \quad (6.14)$$

where $E'_k$ denotes the derivative of $E_k$ with respect to its argument. We also have that $\partial E_k(n_\ell x_\ell - t) / \partial t = -E'_k(n_\ell x_\ell - t)$, and so we can write (6.14) as

$$\frac{\partial B_i}{\partial t} = \epsilon_{ijk} n_j \frac{\partial}{\partial t} E_k(n_\ell x_\ell - t). \quad (6.15)$$

This can be integrated with respect to $t$, dropping the constant of integration since an additional static $\vec{B}$ field term is of no interest to us when discussing electromagnetic waves. Thus we have

$$B_i = \epsilon_{ijk} n_j E_k, \quad \text{i.e.} \quad \vec{B} = \vec{n} \times \vec{E}. \quad (6.16)$$

The source-free Maxwell equation $\vec{\nabla} \cdot \vec{E} = 0$ implies

$$\partial_i E_i(n_j x_j - t) = n_i E'_i(n_j x_j - t) = -\frac{\partial}{\partial t} \vec{n} \cdot \vec{E} = 0. \quad (6.17)$$

Again, we can drop the constant of integration since we are not interested in including static electric fields in this discussion about electromagnetic waves, and conclude that for the plane wave

$$\vec{n} \cdot \vec{E} = 0. \quad (6.18)$$

Since $\vec{B} = \vec{n} \times \vec{E}$, it immediately follows that $\vec{n} \cdot \vec{B} = 0$ and $\vec{E} \cdot \vec{B} = 0$ also. Thus we see that for a plane electromagnetic wave propagating along the $\vec{n}$ direction, the $\vec{E}$ and $\vec{B}$ vectors are orthogonal to $\vec{n}$ and also orthogonal to each other:

$$\vec{n} \cdot \vec{E} = 0, \quad \vec{n} \cdot \vec{B} = 0, \quad \vec{E} \cdot \vec{B} = 0. \quad (6.19)$$

It also follows from $\vec{B} = \vec{n} \times \vec{E}$ that

$$|\vec{E}| = |\vec{B}|, \quad \text{i.e.} \quad E = B. \quad (6.20)$$

Thus we find that the energy density $W$ is given by

$$W = \frac{1}{8\pi}(E^2 + B^2) = \frac{1}{4\pi}E^2. \quad (6.21)$$

The Poynting flux $\vec{S} = (\vec{E} \times \vec{B})/(4\pi)$ is given by

$$S_i = \frac{1}{4\pi}\epsilon_{ijk}E_j\epsilon_{k\ell m}n_\ell E_m = \frac{1}{4\pi}n_i E_j E_j - \frac{1}{4\pi}E_i n_j E_j,$$

$$= \frac{1}{4\pi}n_i E_j E_j, \quad (6.22)$$
and so we have

\[ W = \frac{1}{4\pi} E^2, \quad \vec{S} = \frac{1}{4\pi} \vec{n} E^2 = \vec{n} W. \] (6.23)

Note that the argument \( \vec{n} \cdot \vec{r} - t \) can be written as

\[ \vec{n} \cdot \vec{r} - t = n_\mu x^\mu, \] (6.24)

where \( n_\mu = (-1, \vec{n}) \) and hence

\[ n^\mu = (1, \vec{n}). \] (6.25)

Since \( \vec{n} \) is a unit vector, \( \vec{n} \cdot \vec{n} = 1 \), we have

\[ n_\mu n^\mu = \eta_{\mu\nu} n^\mu n^\nu = 0. \] (6.26)

\( n^\mu \) is called a Null Vector. This is a non-vanishing vector whose norm \( n_\mu n^\mu \) vanishes. Such vectors can arise because of the minus sign in the \( \eta_{00} \) component of the 4-metric. By contrast, in a metric of positive-definite signature, such as the 3-dimensional Euclidean metric \( \delta_{ij} \), a vector whose norm vanishes is itself necessarily zero.

We can now evaluate the various components of the energy-momentum tensor, which are given by (4.114) and the equations that follow it. Thus we have

\[ T^{00} = W = \frac{1}{4\pi} E^2 = \frac{1}{4\pi} B^2, \]
\[ T^{0i} = T^{i0} = S_i = n_i W, \]
\[ T^{ij} = \frac{1}{4\pi} \left(-E_i E_j - B_i B_j + \frac{1}{2}(E^2 + B^2)\delta_{ij}\right), \]
\[ = \frac{1}{4\pi} \left(-E_i E_j - \epsilon_{ikl}\epsilon_{jmn}n_k n_m E_l E_n + E^2 \delta_{ij}\right), \]
\[ = \frac{1}{4\pi} \left(-E_i E_j - \delta_{ij} E^2 - n_i n_k E_k E_j - n_j n_k E_k E_i + \delta_{ij} n_k n_m E_k E_\ell \right. \]
\[ + n_i n_j E_\ell E_\ell + n_k n_k E_i E_j + E^2 \delta_{ij}\right), \]
\[ = \frac{1}{4\pi} \left( n_i n_j E^2 = n_i n_j W. \right) \] (6.27)

Note that in deriving this last result, we have used the identity

\[ \epsilon_{ikl}\epsilon_{jmn} = \delta_{ij}\delta_{km}\delta_{ln} + \delta_{im}\delta_{kn}\delta_{lj} + \delta_{in}\delta_{kj}\delta_{lm} - \delta_{in}\delta_{kj}\delta_{lm} - \delta_{ij}\delta_{kn}\delta_{lm} - \delta_{in}\delta_{km}\delta_{lj}. \] (6.28)

The expressions for \( T^{00}, T^{0i} \) and \( T^{ij} \) can be combined into the single Lorentz-covariant expression

\[ T^{\mu\nu} = n^\mu n^\nu W. \] (6.29)

From this, we can compute the conserved 4-momentum

\[ P^\mu = \int_{t=\text{const.}} T^{\mu\nu} d\Sigma_\nu = \int T^{\mu 0} d^3 x, \]
\[ = \int n^\mu W d^3 x = n^\mu \int W d^3 x, \] (6.30)
and hence we have

\[ P^\mu = \eta^\mu \mathcal{E} , \]  

(6.31)

where

\[ \mathcal{E} = \int W d^3x , \]  

(6.32)

the total energy of the electromagnetic field. Note that \( P^\mu \) is also a null vector,

\[ P^\mu P_\mu = \mathcal{E}^2 \eta^{\mu \nu} n_\mu n_\nu = 0 . \]  

(6.33)

### 6.2 Monochromatic plane waves

In the discussion above, we considered plane electromagnetic waves with an arbitrary profile. A special case is to consider the situation when the plane wave has a definite frequency \( \omega \), so that its time dependence is of the form \( \cos \omega t \). Thus we can write

\[ \vec{E} = \vec{E}_0 e^{i(\vec{k} \cdot \vec{r} - \omega t)} , \quad \vec{B} = \vec{B}_0 e^{i(\vec{k} \cdot \vec{r} - \omega t)} , \]  

(6.34)

where \( \vec{E}_0 \) and \( \vec{B}_0 \) are (possibly complex) constants. The physical \( \vec{E} \) and \( \vec{B} \) fields are obtained by taking the real parts of \( \vec{E} \) and \( \vec{B} \). (Since the Maxwell equations are linear, we can always choose to work in such a complex notation, with the understanding that we take the real parts to get the physical quantities.) It is customary to use the same symbols \( \vec{E} \) and \( \vec{B} \) both for the complex fields and for the physical fields obtained by taking their real parts. There is usually no risk of confusion, since it should be clear from the context which fields are intended.

As we shall discuss in some detail later, more general electromagnetic wave solutions, including the plane wave with arbitrary profile discussed previously, can be built up as linear combinations of the monochromatic plane-wave solutions. The most general wave solutions correspond to superpositions of monochromatic plane waves summed over all wavevectors and frequencies.

Of course, for the fields in (6.34) to solve the Maxwell equations, there must be relations among the constants \( \vec{k} \), \( \omega \), \( \vec{E}_0 \) and \( \vec{B}_0 \). Specifically, since \( \vec{E} \) and \( \vec{B} \) must satisfy the wave equations (6.2) and (6.3), we must have

\[ \vec{k}^2 = \omega^2 , \]  

(6.35)

and since \( \vec{\nabla} \cdot \vec{E} = 0 \) and \( \vec{\nabla} \cdot \vec{B} = 0 \), we must have

\[ \vec{k} \cdot \vec{E}_0 = 0 , \quad \vec{k} \cdot \vec{B}_0 = 0 . \]  

(6.36)
Finally, following the discussion in the more general case above, it follows from \( \nabla \times \vec{E} + \partial \vec{B} / \partial t = 0 \) and \( \nabla \times \vec{B} - \partial \vec{E} / \partial t = 0 \) that

\[
\vec{B} = \frac{\vec{k} \times \vec{E}}{\omega}.
\] (6.37)

It is natural, therefore, to introduce the 4-vector

\[
k^\mu = (\omega, \vec{k}) = \omega n^\mu,
\] (6.38)

where \( n^\mu = (1, \vec{n}) \) and \( \vec{n} = \vec{k} / |\vec{k}| = \vec{k} / \omega \). Equation (6.35) then becomes simply the statement that \( k^\mu \) is a null vector,

\[
k^\mu k_\mu = 0.
\] (6.39)

Note that the argument of the exponentials in (6.34) can now be written as

\[
\vec{k} \cdot \vec{r} - \omega t = k_\mu x^\mu,
\] (6.40)

which we shall commonly write as \( k \cdot x \). Thus we may rewrite (6.34) more briefly as

\[
\vec{E} = \vec{E}_0 e^{i k \cdot x}, \quad \vec{B} = \vec{B}_0 e^{i k \cdot x}.
\] (6.41)

As usual, we have a plane transverse wave, propagating in the direction of the unit 3-vector \( \vec{n} = \vec{k} / \omega \). The term “transverse” here signifies that \( \vec{E} \) and \( \vec{B} \) are perpendicular to the direction in which the wave is propagating. In fact, we have

\[
\vec{n} \cdot \vec{E} = \vec{n} \cdot \vec{B} = 0, \quad \vec{B} = \vec{n} \times \vec{E},
\] (6.42)

and so we have also that \( \vec{E} \) and \( \vec{B} \) are perpendicular to each other, and that \( |\vec{E}| = |\vec{B}| \).

Consider the case where \( \vec{E}_0 \) is taken to be real, which means that \( \vec{B}_0 \) is real too. Then the physical fields (obtained by taking the real parts of the fields given in (6.34)), are given by

\[
\vec{E} = \vec{E}_0 \cos(\vec{k} \cdot \vec{r} - \omega t), \quad \vec{B} = \vec{B}_0 \cos(\vec{k} \cdot \vec{r} - \omega t).
\] (6.43)

The energy density is then given by

\[
W = \frac{1}{8\pi} (E^2 + B^2) = \frac{1}{4\pi} E_0^2 \cos^2(\vec{k} \cdot \vec{r} - \omega t).
\] (6.44)

If we define the time average of \( W \) by

\[
\langle W \rangle \equiv \frac{1}{T} \int_0^T W dt,
\] (6.45)
where $T = 2\pi/\omega$ is the period of the oscillation, then we shall have

$$\langle W \rangle = \frac{1}{8\pi} E_0^2 = \frac{1}{8\pi} B_0^2. \quad (6.46)$$

Note that in terms of the complex expressions (6.34), we can write this as

$$\langle W \rangle = \frac{1}{8\pi} \vec{E} \cdot \vec{E}^* = \frac{1}{8\pi} \vec{B} \cdot \vec{B}^*, \quad (6.47)$$

where the * denotes complex conjugation, since the time and position dependence of $\vec{E}$ or $\vec{B}$ is cancelled when multiplied by the complex conjugate field.\textsuperscript{20}

In general, when $\vec{E}_0$ and $\vec{B}_0$ are not real, we shall also have the same expressions (6.47) for the time-averaged energy density.

In a similar manner, we can evaluate the time average of the Poynting flux vector $\vec{S} = (\vec{E} \times \vec{B})/(4\pi)$. If we first consider the case where $\vec{E}_0$ is real, we shall have

$$\vec{S} = \frac{1}{4\pi} \vec{E} \times \vec{B} = \frac{1}{4\pi} \vec{E}_0 \times \vec{B}_0 \cos^2(\vec{n} \cdot \vec{r} - \omega t) = \frac{1}{4\pi} \vec{n} E_0^2 \cos^2(\vec{n} \cdot \vec{r} - \omega t), \quad (6.48)$$

and so

$$\langle \vec{S} \rangle = \frac{1}{8\pi} \vec{E}_0 \times \vec{B}_0 = \frac{1}{8\pi} \vec{n} E_0^2. \quad (6.49)$$

In general, even if $\vec{E}_0$ and $\vec{B}_0$ are not real, we can write $\langle \vec{S} \rangle$ in terms of the complex $\vec{E}$ and $\vec{B}$ fields as

$$\langle \vec{S} \rangle = \frac{1}{8\pi} \vec{E} \times \vec{B}^* = \frac{1}{8\pi} \vec{n} \vec{E} \cdot \vec{E}^*. \quad (6.50)$$

Note that we have

$$\langle \vec{S} \rangle = \vec{n} \langle W \rangle. \quad (6.51)$$

### 6.3 Motion of a point charge in a linearly-polarised E.M. wave

Consider a plane wave propagating in the $z$ direction, with

$$\vec{E} = (E_0 \cos \omega(z - t), 0, 0), \quad \vec{B} = (0, E_0 \cos \omega(z - t), 0). \quad (6.52)$$

(Recall that $|\vec{E}| = \omega$.) Suppose now that there is a particle of mass $m$ and charge $e$ in this field. By the Lorentz force equation we shall have

$$\frac{d\vec{p}}{dt} = e\vec{E} + e\vec{v} \times \vec{B}. \quad (6.53)$$

\textsuperscript{20}This “trick,” of expressing the time-averaged energy density in terms of the dot product of the complex field with its complex conjugate, is rather specific to this situation, where the quantity being time-averaged is quadratic in the electric and magnetic fields.
For simplicity, we shall make the assumption that the motion of the particle can be treated non-relativistically, and so
\[
\vec{p} = m\vec{v} = m\frac{d\vec{r}}{dt}.
\] (6.54)

Let us suppose that the particle is initially located at the point \( z = 0 \), and that it moves only by a small amount in comparison to the wavelength \( 2\pi/\omega \) of the electromagnetic wave. Therefore, to a good approximation, we can assume that the particle is sitting in the uniform, although time-dependent, electromagnetic field obtained by setting \( z = 0 \) in (6.52). Thus
\[
\vec{E} = (E_0 \cos \omega t, 0, 0), \quad \vec{B} = (0, E_0 \cos \omega t, 0),
\] (6.55)
and so the Lorentz force equation gives
\[
\begin{align*}
m\ddot{x} &= eE_0 \cos \omega t - e\dot{z}E_0 \cos \omega t \approx eE_0 \cos \omega t, \\
m\ddot{y} &= 0, \\
m\ddot{z} &= e\dot{x}E_0 \cos \omega t.
\end{align*}
\] (6.56)

Note that the approximation in the first line follows from our assumption that the motion of the particle is non-relativistic, so \( |\dot{z}| << 1 \).

With convenient and inessential choices for the constants of integration, we first obtain
\[
\begin{align*}
\dot{x} &= \frac{eE_0}{m\omega} \sin \omega t, \\
x &= -\frac{eE_0}{m\omega^2} \cos \omega t,
\end{align*}
\] (6.57)
Substituting into the \( z \) equation then gives
\[
\ddot{z} = \frac{e^2 E_0^2}{m^2\omega^2} \sin \omega t \cos \omega t = \frac{e^2 E_0^2}{2m^2\omega} \sin 2\omega t,
\] (6.58)
which integrates to give (dropping inessential constants of integration)
\[
z = -\frac{e^2 E_0^2}{8m^2\omega^3} \sin 2\omega t.
\] (6.59)

The motion in the \( y \) direction is purely linear, and since we are not interested in the case where the particle drifts uniformly through space, we can just focus on the solution where \( y \) is constant, say \( y = 0 \).

Thus the interesting motion of the particle in the electromagnetic field is of the form
\[
\begin{align*}
x &= \alpha \cos \omega t, \\
z &= \beta \sin 2\omega t = 2\beta \sin \omega t \cos \omega t,
\end{align*}
\] (6.60)
where
\[
\alpha = -\frac{eE_0}{m\omega^2}, \quad \beta = -\frac{e^2 E_0^2}{8m^2\omega^3}.
\] (6.61)
Thus we find
\[ z = \frac{2\beta}{\alpha} x \sqrt{1 - \frac{x^2}{\alpha^2}}. \]  
(6.62)

This describes a “figure of eight” lying on its side in the \((x, z)\) plane. The assumptions we made in deriving this, namely non-relativistic motion and a small \(z\) displacement relative to the wavelength of the electromagnetic wave, can be seen to be satisfied provided the amplitude \(E_0\) of the wave is sufficiently small. Note that the displacement in the \(z\) direction is small compared with the displacement in the \(x\) direction, since the dimensionless ratio \(2\beta/\alpha\) is given by
\[ \frac{2\beta}{\alpha} = \frac{eE_0}{4m\omega}, \]  
(6.63)

and we see from the expression for \(\dot{x}\) in (6.57) that this ratio is small since \(\dot{x}\) is assumed small compared to the speed of light (i.e. 1).

The response of the charge particle to electromagnetic wave provides a model for how the electrons in a receiving antenna behave in the presence of an electromagnetic wave. This shows how the wave is converted into oscillatory currents in the antenna, which are then amplified and processed into the final output signal in a radio receiver.

### 6.4 Circular and elliptical polarisation

The electromagnetic wave described in section 6.2 is linearly polarised. For example, we could consider the solution with
\[ \vec{E}_0 = (0, E_0, 0), \quad \vec{B}_0 = (0, 0, B_0), \quad \vec{n} = (1, 0, 0). \]  
(6.64)

This corresponds to a linearly polarised electromagnetic wave propagating along the \(x\) direction.

By taking a linear superposition of waves propagating along a given direction \(\vec{n}\), we can obtain circularly polarised, or more generally, elliptically polarised, waves. Let \(\vec{e}\) and \(\vec{f}\) be two orthogonal unit vectors, that are also both orthogonal to \(\vec{n}\):
\[ \vec{e} \cdot \vec{e} = 1, \quad \vec{f} \cdot \vec{f} = 1, \quad \vec{n} \cdot \vec{n} = 1, \]
\[ \vec{e} \cdot \vec{f} = 0, \quad \vec{n} \cdot \vec{e} = 0, \quad \vec{n} \cdot \vec{f} = 0. \]  
(6.65)

Suppose now we consider a plane wave given by
\[ \vec{E} = (E_0 \vec{e} + \vec{E}_0 \vec{f}) e^{i(\vec{k} \cdot \vec{r} - \omega t)}, \quad \vec{B} = \vec{n} \times \vec{E}, \]  
(6.66)
where \( E_0 \) and \( \tilde{E}_0 \) are complex constants. If \( E_0 \) and \( \tilde{E}_0 \) both have the same phase (i.e. \( \tilde{E}_0/E_0 \) is real), then we again have a linearly-polarised electromagnetic wave. If instead the phases of \( E_0 \) and \( \tilde{E}_0 \) are different, then the wave is in general elliptically polarised.

Consider as an example the case where

\[
\tilde{E}_0 = \pm i E_0 ,
\]

(with \( E_0 \) taken to be real, without loss of generality), for which the electric field will be given by

\[
\tilde{E} = E_0(\vec{e} \pm i \vec{f}) e^{i(\vec{k} \cdot \vec{r} - \omega t)} .
\]

Taking the real part, to get the physical electric field, we obtain

\[
\tilde{E} = E_0 \vec{e} \cos(\vec{k} \cdot \vec{r} - \omega t) \mp E_0 \vec{f} \sin(\vec{k} \cdot \vec{r} - \omega t) .
\]

For example, if we choose

\[
\vec{n} = (0,0,1) , \quad \vec{e} = (1,0,0) , \quad \vec{f} = (0,1,0) ,
\]

then the electric field is given by

\[
E_x = E_0 \cos(\omega z - \omega t) , \quad E_y = \mp E_0 \sin(\omega z - \omega t) .
\]

It is clear from this that the magnitude of the electric field is constant,

\[
|\tilde{E}| = E_0 .
\]

If we fix a value of \( z \), then the \( \tilde{E} \) vector can be seen to be rotating around the \( z \) axis (the direction of motion of the wave). This rotation is \textit{anticlockwise} in the \((x,y)\) plane if we choose the plus sign in (6.67), and \textit{clockwise} if we choose the minus sign instead. These two choices correspond to having a circularly polarised wave of positive or negative helicity respectively. (Positive helicity means the rotation is parallel to the direction of propagation, while negative helicity means the rotation is anti-parallel to the direction of propagation.)

In more general cases, where the magnitudes of \( E_0 \) and \( \tilde{E}_0 \) are unequal, or where the phase angle between them is not equal to 0 (linear polarisation) or 90 degrees, the electromagnetic wave will be elliptically polarised. Consider, for example, the case where the electric field is given by

\[
\tilde{E} = (a_1 e^{i\delta_1}, a_2 e^{i\delta_2}, 0) e^{i\omega(z-t)} ,
\]

with the propagation direction being \( \vec{n} = (0,0,1) \). Then we shall have

\[
\vec{B} = \vec{n} \times \tilde{E} = (-a_2 e^{i\delta_2}, a_1 e^{i\delta_1}, 0) e^{i\omega(z-t)} .
\]
The real constants $a_1$, $a_2$, $\delta_1$ and $\delta_2$ determine the nature of this plane wave propagating along the $z$ direction. Of course the overall phase is unimportant, so really it is only the difference $\delta_2 - \delta_1$ between the phase angles that is important.

The magnitude and phase information is sometimes expressed in terms of the Stokes Parameters $(s_0, s_1, s_2, s_3)$, which are defined by

\begin{align*}
    s_0 &= E_x E_x + E_y E_y^* = a_1^2 + a_2^2, \\
    s_1 &= E_x E_x^* - E_y E_y^* = a_1^2 - a_2^2, \\
    s_2 &= 2 \Re(E_x^* E_y) = 2a_1 a_2 \cos(\delta_2 - \delta_1), \\
    s_3 &= 2 \Im(E_x^* E_y) = 2a_1 a_2 \sin(\delta_2 - \delta_1).
\end{align*}

(The last two involve the real and imaginary parts of $(E_x^* E_y)$ respectively.) The four Stokes parameters are not independent:

\begin{equation}
    s_0^2 = s_1^2 + s_2^2 + s_3^2. \tag{6.76}
\end{equation}

The parameter $s_0$ characterises the intensity of the electromagnetic wave, while $s_1$ characterises the amount of $x$ polarisation versus $y$ polarisation, with

\begin{equation}
    -s_0 \leq s_1 \leq s_0. \tag{6.77}
\end{equation}

The third independent parameter, which could be taken to be $s_2$, characterises the phase difference between the $x$ and the $y$ polarised waves. Circular polarisation with $\pm$ helicity corresponds to

\begin{equation}
    s_1 = 0, \quad s_2 = 0, \quad s_3 = \pm s_0. \tag{6.78}
\end{equation}

### 6.5 General superposition of plane waves

So far in the discussion of electromagnetic waves, we have considered the case where there is a single direction of propagation (i.e. a plane wave), and a single frequency (monochromatic). The most general wave-like solutions of the Maxwell equations can be expressed as linear combinations of these basic monochromatic plane-wave solutions.

In order to discuss the general wave solutions, it is helpful to work with the gauge potential $A^\mu = (\phi, \vec{A})$. Recall that we have the freedom to make gauge transformations $A_\mu \rightarrow A_\mu + \partial_\mu \lambda$, where $\lambda$ is an arbitrary function. For the present purposes, of describing wave solutions, a convenient choice of gauge is to set $\phi = 0$. Such a gauge choice would not be convenient when discussing solutions in electrostatics, but in the present case, where we know that the wave solutions are necessarily time-dependent, it is quite helpful. It is know as the Radiation Gauge.
Thus, we shall first write a single monochromatic plane wave in terms of the 3-vector potential, as
\[ \vec{A} = a\vec{e} e^{i(\vec{k} \cdot \vec{r} - \omega t)} , \] (6.79)
where \( \vec{e} \) is a unit polarisation vector, and \( a \) is a constant. As usual, we must have \( |\vec{k}|^2 = \omega^2 \).

The electric and magnetic fields will be given by
\[ \vec{E} = -\vec{\nabla} \phi - \frac{\partial \vec{A}}{\partial t} = i a\omega \vec{e} e^{i(\vec{k} \cdot \vec{r} - \omega t)} , \]
\[ \vec{B} = \vec{\nabla} \times \vec{A} = i a \vec{k} \times \vec{e} e^{i(\vec{k} \cdot \vec{r} - \omega t)} = \frac{\vec{k} \times \vec{E}}{\omega} . \] (6.80)

We can immediately see that \( \vec{E} \) and \( \vec{B} \) satisfy the wave equation, and that we must impose \( \vec{e} \cdot \vec{k} = 0 \) in order to satisfy \( \vec{\nabla} \cdot \vec{E} = 0 \).

We have established, therefore, that (6.79) describes a monochromatic plane wave propagating along the \( \vec{k} \) direction, with electric field along \( \vec{e} \), provided that \( \vec{e} \cdot \vec{k} = 0 \) and \( |\vec{k}| = \omega \).

More precisely, the gauge potential that gives the physical (i.e. real) electric and magnetic fields is given by taking the real part of \( \vec{A} \) in (6.79). Thus, when we want to describe the actual physical quantities, we shall write
\[ \vec{A} = a\vec{e} e^{i(\vec{k} \cdot \vec{r} - \omega t)} + a^* \vec{e} e^{-i(\vec{k} \cdot \vec{r} - \omega t)} . \] (6.81)
(We have absorbed a factor of \( \frac{1}{2} \) here into a rescaling of \( a \), in order to avoid carrying \( \frac{1}{2} \) factors around in all the subsequent equations.) For brevity, we shall usually write the “physical” \( \vec{A} \) as
\[ \vec{A} = a\vec{e} e^{i(\vec{k} \cdot \vec{r} - \omega t)} + \text{c.c.} , \] (6.82)
where \( \text{c.c} \) stands for “complex conjugate.”

Now consider a general linear superposition of monochromatic plane waves, with different wave-vectors \( \vec{k} \), different polarisation vectors \( \vec{e} \), and different amplitudes \( a \). We shall therefore label the polarisation vectors and amplitudes as follows:
\[ \vec{e} \rightarrow \vec{e}_\lambda(\vec{k}) , \quad a \rightarrow a_\lambda(\vec{k}) . \] (6.83)
Here \( \lambda \) is an index which ranges over the values 1 and 2, which labels 2 real orthonormal vectors \( \vec{e}_1(\vec{k}) \) and \( \vec{e}_2(\vec{k}) \) that span the 2-plane perpendicular to \( \vec{k} \). The general wave solution can then be written as the sum over all such monochromatic plane waves of the form (6.82). Since a continuous range of wave-vectors is allowed, the summation over these will be a 3-dimensional integral. Thus we can write
\[ \vec{A} = \sum_{\lambda=1}^{2} \int \frac{d^3\vec{k}}{(2\pi)^3} \left[ \vec{e}_\lambda(\vec{k}) a_\lambda(\vec{k}) e^{i(\vec{k} \cdot \vec{r} - \omega t)} + \text{c.c.} \right] , \] (6.84)
where $\omega = |\vec{k}|$, and

$$\vec{k} \cdot \vec{e}_\lambda(\vec{k}) = 0, \quad \vec{e}_\lambda(\vec{k}) \cdot \vec{e}_\lambda(\vec{k}) = \delta_{\lambda\lambda'}.$$ (6.85)

For many purposes, it will be convenient to expand $\vec{A}$ in a basis of circularly-polarised monochromatic plane waves, rather than linearly-polarised waves. In this case, we should choose the 2-dimensional basis of polarisation vectors $\vec{e}_\pm$, related to the previous basis by

$$\vec{e}_\pm = \frac{1}{\sqrt{2}} (\vec{e}_1 \pm i \vec{e}_2).$$ (6.86)

Since we have $\vec{e}_i \cdot \vec{e}_j = \delta_{ij}$, it follows that

$$\vec{e}_+ \cdot \vec{e}_+ = 0, \quad \vec{e}_- \cdot \vec{e}_- = 0, \quad \vec{e}_+ \cdot \vec{e}_- = 1.$$ (6.87)

Note that $\vec{e}_\pm^* = \vec{e}_\mp$. We can label the $\vec{e}_\pm$ basis vectors by $\vec{e}_\lambda$, where $\lambda$ is now understood to take the two “values” + and −. We then write the general wave solution as

$$\vec{A} = \sum_{\lambda=\pm} \int \frac{d^3 \vec{k}}{(2\pi)^3} \left[ \vec{e}_\lambda(\vec{k}) a_{\lambda}(\vec{k}) e^{i(\vec{k} \cdot \vec{r} - \omega t)} + \text{c.c.} \right],$$ (6.88)

Of course, we also have $\vec{k} \cdot \vec{e}_\lambda = 0$, and $\omega = |\vec{k}|$.

### 6.5.1 Helicity and energy of circularly-polarised waves

The angular-momentum tensor $M^{\mu\nu}$ for the electromagnetic field is defined by

$$M^{\mu\nu} = \int_{t=\text{const}} (x'^\mu T^{\nu\rho} - x'^\nu T^{\mu\rho}) d\Sigma_\rho,$$ (6.89)

and so the three-dimensional components $M^{ij}$ are

$$M^{ij} = \int_{t=\text{const}} (x'^i T^{j0} - x'^j T^{i0}) d\Sigma_0 = \int (x'^i S^j - x'^j S^i) d^3 x,$$ (6.90)

Thus, since $\vec{S} = (\vec{E} \times \vec{B})/(4\pi)$, the three-dimensional angular momentum $L_i = \frac{1}{2} \epsilon_{ijk} M^{jk}$ is given by

$$L_i = \int \epsilon_{ijk} x'^j S^k d^3 x,$$ (6.91)

i.e.

$$\vec{L} = \frac{1}{4\pi} \int \vec{r} \times (\vec{E} \times \vec{B}) d^3 x.$$ (6.92)
Now, since \( \vec{B} = \vec{\nabla} \times \vec{A} \), we have

\[
[\vec{r} \times (\vec{E} \times \vec{B})]_i = \epsilon_{ijk} \epsilon_{klm} x_j E_l B_m ,
= \epsilon_{ijk} \epsilon_{klm} \epsilon_{mpq} x_j E_l \partial_p A_q ,
= \epsilon_{ijk} (\delta_{kp} \delta_{lq} - \delta_{kq} \delta_{lp}) x_j E_l \partial_p A_q ,
= \epsilon_{ijk} x_j E_l \partial_k A_l - \epsilon_{ijk} x_j E_l \partial_l A_k ,
\]

(6.93)

and so

\[
L_i = \frac{1}{4 \pi} \int (\epsilon_{ijk} x_j E_l \partial_k A_l - \epsilon_{ijk} x_j E_l \partial_l A_k) d^3 x ,
= \frac{1}{4 \pi} \int \left( - \epsilon_{ijk} \partial_k (x_j E_l) A_l + \partial_l (x_j E_l) A_k \right) d^3 x ,
= \frac{1}{4 \pi} \int \left( - \epsilon_{ijk} x_j (\partial_k E_l) A_l + \epsilon_{ijk} E_l A_k \right) d^3 x .
\]

(6.94)

Note that in performing the integrations by parts here, we have, as usual, assumed that the fields fall off fast enough at infinity that the surface term can be dropped. We have also used the source-free Maxwell equation \( \partial_t E_l = 0 \) in getting to the final line. Thus, we conclude that the angular momentum 3-vector can be expressed as

\[
\vec{L} = \frac{1}{4 \pi} \int (\vec{E} \times \vec{A} - A_i (\vec{r} \times \vec{\nabla}) E_i) d^3 x .
\]

(6.95)

The two terms in (6.95) can be interpreted as follows. The second term can be viewed as an “orbital angular momentum,” since it clearly depends on the choice of origin. It is rather analogous to an \( \vec{r} \times \vec{p} \) contribution to the angular momentum of a system of particles. On the other hand, the first term in (6.95) can be viewed as an “intrinsic spin” term, since it is constructed purely from the electromagnetic fields themselves, and is independent of the choice of origin. We shall calculate this spin contribution,

\[
\vec{L}_{\text{spin}} = \frac{1}{4 \pi} \int \vec{E} \times \vec{A} d^3 x
\]

(6.96)

to the angular momentum in the case of the sum over circularly-polarised waves that we introduced in the previous section. Recall that for this sum, the 3-vector potential is given by

\[
\vec{A} = \sum_{\lambda=\pm} \int \frac{d^3 \vec{k}'}{(2\pi)^3} \left[ \vec{e}_\lambda(\vec{k}') a_\lambda(\vec{k}') e^{i(\vec{k}' \cdot \vec{r} - \omega' t)} + \text{c.c.} \right] ,
\]

(6.97)

The electric field is then given by

\[
\vec{E} = -\frac{\partial \vec{A}}{\partial t} = \sum_{\lambda=\pm} \int \frac{d^3 \vec{k}}{(2\pi)^3} \left[ i \omega \vec{e}_\lambda(\vec{k}) a_\lambda(\vec{k}) e^{i(\vec{k} \cdot \vec{r} - \omega t)} + \text{c.c.} \right] ,
\]

(6.98)
Note that we have put primes on the summation and integration variables \( \lambda \) and \( \vec{k} \) in the expression for \( \vec{A} \). This is so that we can take the product \( \vec{E} \times \vec{A} \) and not have a clash of “dummy” summation variables, in what will follow below. We have also written the frequency as \( \omega' \equiv |\vec{k}'| \) in the expression for \( \vec{A} \).

Our interest will be to calculate the time average

\[
\langle \vec{L}_{\text{spin}} \rangle \equiv \frac{1}{T} \int_0^T \vec{L}_{\text{spin}} dt.
\]  

(6.99)

Since we are considering a wave solution with an entire “chorus” of frequencies now, we define the time average by taking \( T \) to infinity. (It is easily seen that this coincides with the previous definition of the time average for a monochromatic wave of frequency \( \omega \), where \( T \) was taken to be \( 2\pi/\omega \).) Note that the time average will be zero for any quantity whose time dependence is of the oscillatory form \( e^{i\nu t} \), because we would have

\[
\frac{1}{T} \int_0^T e^{i\nu t} dt = \frac{1}{i\nu T} (e^{i\nu T} - 1),
\]  

(6.100)

which clearly goes to zero as \( T \) goes to infinity. Since the time dependence of all the quantities we shall consider is precisely of the form \( e^{i\nu t} \), it follows that in order to survive the time averaging, it must be that \( \nu = 0 \). Thus we have \( \langle e^{i\nu t} \rangle = 0 \) if \( \nu \neq 0 \) and \( \langle e^{i\nu t} \rangle = 1 \) if \( \nu = 0 \).

We are interested in calculating the time average of \( \vec{E} \times \vec{A} \), where \( \vec{A} \) and \( \vec{E} \) are given by (6.97) and (6.98). The quantities \( \omega \) appearing there are, by definition, positive, since we have defined \( \omega \equiv |\vec{k}| \). The only way that we shall get terms in \( \vec{E} \times \vec{A} \) that have zero frequency (i.e. \( \nu = 0 \)) is from the product of one of the terms that is explicitly written times one of the “c.c.” terms, since these, of course, have the opposite sign for their frequency dependence.

The upshot of this discussion is that when we evaluate the time average of \( \vec{E} \times \vec{A} \), with \( \vec{A} \) and \( \vec{E} \) given by (6.97) and (6.98), the only terms that survive will be coming from the product of the explicitly-written term for \( \vec{E} \) times the “c.c.” term for \( \vec{A} \), plus the “c.c.” term for \( \vec{E} \) times the explicitly-written term for \( \vec{A} \). Furthermore, in order for the products to have zero frequency, and therefore survive the time averaging, it must be that \( \omega' = \omega \).

We therefore find

\[
\langle \vec{E} \times \vec{A} \rangle = \sum_{\lambda \lambda'} \int \frac{d^3\vec{k}}{(2\pi)^3} \frac{d^3\vec{k}'}{(2\pi)^3} i \omega \left[ \epsilon_{\lambda}(\vec{k}) \times \epsilon^{\ast}_{\lambda'}(\vec{k}') a_{\lambda}\langle\vec{k}\rangle a^{\ast}_{\lambda'}(\vec{k}') e^{i(\vec{k}-\vec{k}')} \cdot \vec{r} - \epsilon^{\ast}_{\lambda}(\vec{k}) \times \epsilon_{\lambda'}(\vec{k}') a_{\lambda'}\langle\vec{k}\rangle a_{\lambda'}(\vec{k}') e^{-i(\vec{k}-\vec{k}')} \cdot \vec{r} \right]
\]  

(6.101)
We now need to integrate $\langle \vec{E} \times \vec{A} \rangle$ over all 3-space, which we shall write as

$$\int \langle \vec{E} \times \vec{A} \rangle \, d^3\vec{r}. \quad (6.102)$$

We now make use of the result from the theory of delta functions that

$$\int e^{i(\vec{k} - \vec{k}') \cdot \vec{r}} \, d^3\vec{r} = (2\pi)^3 \delta^3(\vec{k} - \vec{k}'). \quad (6.103)$$

Therefore, from (6.101) we find

$$\int \langle \vec{E} \times \vec{A} \rangle \, d^3\vec{r} = \sum_{\lambda' \lambda} \int \frac{d^3\vec{k}}{(2\pi)^3} \, i\omega \left[ \vec{e}_\lambda(\vec{k}) \times \vec{e}^*_{\lambda'}(\vec{k}) a_\lambda(\vec{k}) a^*_{\lambda'}(\vec{k}) - \vec{e}^*_\lambda(\vec{k}) \times \vec{e}_{\lambda'}(\vec{k}) a^*_{\lambda}(\vec{k}) a_{\lambda'}(\vec{k}) \right]. \quad (6.104)$$

Finally, we recall that the polarization vectors $\vec{e}_\pm(\vec{k})$ span the 2-dimensional space orthogonal to the wave-vector $\vec{k}$. In terms of the original real basis unit vectors $\vec{e}_1(\vec{k})$ and $\vec{e}_2(\vec{k})$ we have

$$\vec{e}_1(\vec{k}) \times \vec{e}_2(\vec{k}) = \frac{\vec{k}}{\omega}, \quad (6.105)$$

and so it follows from (6.86) that

$$\vec{e}_+(\vec{k}) \times \vec{e}^*_+(\vec{k}) = -\frac{i\vec{k}}{\omega}, \quad \vec{e}_-(\vec{k}) \times \vec{e}^*_-(\vec{k}) = \frac{i\vec{k}}{\omega}. \quad (6.106)$$

From this, it follows that (6.104) becomes

$$\int \langle \vec{E} \times \vec{A} \rangle \, d^3\vec{r} = 2 \int \frac{d^3\vec{k}}{(2\pi)^3} \vec{k} \left[ a_+(\vec{k}) a^*_+(\vec{k}) - a_-(\vec{k}) a^*_-(\vec{k}) \right], \quad (6.107)$$

and so we have

$$\langle \vec{L}_{\text{spin}} \rangle = \frac{1}{2\pi} \int \frac{d^3\vec{k}}{(2\pi)^3} \vec{k} \left( |a_+(\vec{k})|^2 - |a_-(\vec{k})|^2 \right). \quad (6.108)$$

It can be seen from this result that the modes associated with the coefficients $a_+(\vec{k})$ correspond to circularly-polarised waves of positive helicity; i.e. their spin is parallel to the wave-vector $\vec{k}$. Conversely, the modes with coefficients $a_-(\vec{k})$ correspond to circularly-polarised waves of negative helicity; i.e. with spin that is anti-parallel to the wave-vector $\vec{k}$.

In a similar fashion, we may evaluate the energy of the general wave solution as a sum over the individual modes. The total energy $\mathcal{E}$ is given by\textsuperscript{21}

$$\mathcal{E} = \frac{1}{8\pi} \int (E^2 + B^2) \, d^3x \to \frac{1}{4\pi} \int E^2 \, d^3x. \quad (6.109)$$

\textsuperscript{21}We are being a little bit sloppy here, in invoking the result, shown earlier for a single monochromatic plane wave, that the electric and magnetic fields give equal contributions to the energy. It is certainly not true any longer that $\vec{E}^2 = \vec{B}^2$ for a general superposition of plane waves. However, after integrating over
Since $\vec{E} = -\partial \vec{A} / \partial t$ here, we have
\begin{equation}
\langle E^2 \rangle = \sum_{\lambda, \lambda'} \int \frac{d^3 \vec{k}}{(2\pi)^3} \frac{d^3 \vec{k}'}{(2\pi)^3} \omega^2 \left[ \vec{\epsilon}_\lambda(\vec{k}) \cdot \vec{\epsilon}^*_\lambda(\vec{k}') a_\lambda(\vec{k}) a^*_\lambda(\vec{k}') e^{i(\vec{k} - \vec{k}') \cdot \vec{r}} + \vec{\epsilon}^*_\lambda(\vec{k}) \cdot \vec{\epsilon}_\lambda(\vec{k}') a^*_\lambda(\vec{k}) a_\lambda(\vec{k}') e^{-i(\vec{k} - \vec{k}') \cdot \vec{r}} \right],
\end{equation}

where again, the time-averaging has picked out only the terms whose total frequency adds to zero. The integration over all space then again gives a three-dimensional delta function $\delta^3(\vec{k} - \vec{k}')$, and so we find
\begin{equation}
\int \langle E^2 \rangle d^3 \vec{r} = \sum_{\lambda, \lambda'} \int \frac{d^3 \vec{k}}{(2\pi)^3} \omega^2 \left[ \vec{\epsilon}_\lambda(\vec{k}) \cdot \vec{\epsilon}^*_\lambda(\vec{k}) a_\lambda(\vec{k}) a^*_\lambda(\vec{k}) + \vec{\epsilon}^*_\lambda(\vec{k}) \cdot \vec{\epsilon}_\lambda(\vec{k}) a^*_\lambda(\vec{k}) a_\lambda(\vec{k}) \right],
\end{equation}

Finally, using the orthogonality relations (6.87), and the conjugation identity $\vec{\epsilon}_\pm = \vec{\epsilon}^*_\mp$, we obtain
\begin{equation}
\langle E \rangle = \frac{1}{2\pi} \int \frac{d^3 \vec{k}}{(2\pi)^3} \omega^2 \left( |a_+(\vec{k})|^2 + |a_-(\vec{k})|^2 \right).
\end{equation}

From the two results (6.108) and (6.112), we see that for a given mode characterised by helicity $\lambda$ and wave-vector $\vec{k}$, we have
\begin{equation}
\langle L_{\text{spin}} \rangle_{\vec{k}, \lambda} = \frac{1}{2\pi} \vec{k} |a_\lambda(\vec{k})|^2 \left( \text{sign } \lambda \right),
\end{equation}
\begin{equation}
\langle E \rangle_{\vec{k}, \lambda} = \frac{1}{2\pi} \omega^2 |a_\lambda(\vec{k})|^2,
\end{equation}

where $\text{sign } \lambda$ is $+1$ for $\lambda = +$ and $-1$ for $\lambda = -$. The helicity $\sigma$, which is the component of spin along the direction of the wave-vector $\vec{k}$, is therefore given by
\begin{equation}
\sigma = \frac{1}{2\pi} |\vec{k}| |a_\lambda(\vec{k})|^2 \left( \text{sign } \lambda \right),
\end{equation}
\begin{equation}
= \frac{1}{2\pi} \omega |a_\lambda(\vec{k})|^2 \left( \text{sign } \lambda \right),
\end{equation}
\begin{equation}
= \frac{1}{\omega} \langle E \rangle_{\vec{k}, \lambda} \left( \text{sign } \lambda \right).
\end{equation}

In other words, we have that
\begin{equation}
\text{energy} = \pm (\text{helicity}) \omega,
\end{equation}

all space and performing a time averaging, as we shall do below, the contribution of the electric field to the final result will just be a sum over the contributions of all the individual modes. Likewise, the contribution of the magnetic field will be a sum over all the individual modes. It is now true that the electric and magnetic contributions of each mode will be equal, and so one does indeed get the correct answer by simply doubling the result for the electric field alone. Any reader who has doubts about this is invited to perform the somewhat more complicated direct calculation of the contribution from the magnetic field, to confirm that it is true.
and so we can write
\[ \mathcal{E} = |\sigma| \omega. \] (6.116)
This can be compared with the result in quantum mechanics, that
\[ E = \hbar \omega. \] (6.117)
Planck’s constant \( \hbar \) has the units of angular momentum, and in fact the basic “unit” of angular momentum for the photon is one unit of \( \hbar \). In the transition from classical to quantum physics, the helicity of the electromagnetic field becomes the spin of the photon.

### 6.6 Gauge invariance and electromagnetic fields

In the previous discussion, we described electromagnetic waves in terms of the gauge potential \( A_\mu = (-\phi, \vec{A}) \), working in the gauge where \( \phi = 0 \), i.e. \( A_0 = 0 \). Since the gauge symmetry of Maxwell’s equations is
\[ A_\mu \rightarrow A_\mu + \partial_\mu \lambda, \] (6.118)
one might think that all the gauge freedom had been used up when we imposed the condition \( \phi = 0 \), on the grounds that one arbitrary function (the gauge parameter \( \lambda \)) has been used in order to set one function (the scalar potential \( \phi \)) to zero. This is, in fact, not the case. To see this, recall that for the electromagnetic wave we wrote \( \vec{A} \) as a superposition of terms of the form
\[ \vec{A} = \vec{c} e^{i(\vec{k} \cdot \vec{r} - \omega t)}, \] (6.119)
which implied that
\[ \vec{k} \cdot \vec{c} = 0, \] (6.120)
From this we have
\[ \vec{\nabla} \cdot \vec{E} = -\omega \vec{k} \cdot \vec{c} e^{i(\vec{k} \cdot \vec{r} - \omega t)}, \] (6.121)
and so the Maxwell equation \( \vec{\nabla} \cdot \vec{E} = 0 \) implies that \( \vec{k} \cdot \vec{c} = 0 \), and hence
\[ \vec{k} \cdot \vec{A} = 0. \] (6.122)
This means that as well as having \( A_0 = -\phi = 0 \), we also have a component of \( \vec{A} \) vanishing, namely the projection along \( \vec{k} \).

To see how this can happen, it is helpful to go back to a Lorentz-covariant gauge choice instead. First, consider the Maxwell field equation, in the absence of source currents:
\[ \partial^\mu F_{\mu\nu} = 0. \] (6.123)
Since $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$, this implies
\[
\partial^\mu \partial_\mu A_\nu - \partial^\nu \partial_\nu A_\mu = 0. \tag{6.124}
\]
We now choose the Lorenz gauge condition,
\[
\partial^\mu A_\mu = 0. \tag{6.125}
\]
The field equation (6.124) then reduces to
\[
\partial^\mu \partial_\mu A_\nu = 0, \quad \text{i.e.} \quad \Box A_\mu = 0. \tag{6.126}
\]
One might again think that all the gauge symmetry had been “used up” in imposing the Lorenz gauge condition (6.125), on the grounds that the arbitrary function $\lambda$ in the gauge transformation
\[
A_\mu \rightarrow A_\mu + \partial_\mu \lambda \tag{6.127}
\]
that allowed one to impose (6.125) would no longer allow any freedom to impose further conditions on $A_\mu$. This is not quite true, however.

To see this, let us suppose we are already in Lorenz gauge, and then try performing a further gauge transformation, as in (6.127), insisting that we must remain in the Lorenz gauge. This means that $\lambda$ should satisfy
\[
\partial^\mu \partial_\mu \lambda = 0, \quad \text{i.e.} \quad \Box \lambda = 0. \tag{6.128}
\]
Non-trivial such functions $\lambda$ can of course exist; any solution of the wave equation will work.

To see what this implies, let us begin with a general solution of the wave equation (6.126), working in the Lorenz gauge (6.125). We can decompose this solution as a sum over plane waves, where a typical mode in the sum is
\[
A_\mu = a_\mu e^{i(k^\nu x_\nu - \omega t)} = a_\mu e^{ik_\nu x_\nu} = a_\mu e^{ik\cdot x}, \tag{6.129}
\]
where $a_\mu$ and $k_\nu$ are constant. Substituting into the wave equation (6.126) we find
\[
0 = \Box A_\mu = \partial^\sigma \partial_\sigma (a_\mu e^{ik\cdot x}) = -k^\sigma k_\sigma a_\mu e^{ik\cdot x}, \tag{6.130}
\]
whilst the Lorenz gauge condition (6.125) implies
\[
0 = \partial^\mu A_\mu = \partial^\mu (a_\mu e^{ik\cdot x}) = i k^\mu a_\mu e^{ik\cdot x}. \tag{6.131}
\]
In other words, $k_\mu$ and $a_\mu$ must satisfy
\[
k^\mu k_\mu = 0, \quad k^\mu a_\mu = 0. \tag{6.132}
\]
The first of these equations implies that \( k^\mu \) is a null vector, as we had seen earlier. The second equation implies that 1 of the 4 independent components that a 4-vector \( a_\mu \) generically has is restricted in this case, so that \( a_\mu \) has only 3 independent components.

Now we perform the further gauge transformation \( A_\mu \rightarrow A_\mu + \partial_\mu \lambda \), where, as discussed above, \( \Box \lambda = 0 \) so that we keep the gauge-transformed \( A_\mu \) in Lorenz gauge. Specifically, we shall choose

\[
\lambda = i h e^{i k \cdot x},
\]

(6.133)

where \( h \) is a constant. Thus we shall have

\[
A_\mu \rightarrow A_\mu - h k^\mu e^{i k \cdot x}.
\]

(6.134)

With \( A_\mu \) given by (6.129) this means we shall have

\[
a_\mu e^{i k \cdot x} \rightarrow a_\mu e^{i k \cdot x} - h k^\mu e^{i k \cdot x},
\]

(6.135)

which implies

\[
a_\mu \rightarrow a_\mu - h k^\mu.
\]

(6.136)

As a check, we can see that the redefined \( a_\mu \) indeed still satisfies \( k^\mu a_\mu = 0 \), as it should, since \( k^\mu \) is a null vector.

The upshot of this discussion is that the freedom to take the constant \( h \) to be anything we like allows us to place a second restriction on the components of \( a_\mu \). Thus not merely are its ostensible 4 components reduced to 3 by virtue of \( k^\mu a_\mu = 0 \), but a further component can be eliminated by means of the residual gauge freedom, leaving just 2 independent components in the polarisation vector \( a_\mu \). Since the physical degrees of freedom are, by definition, the independent quantities that cannot be changed by making gauge transformations, we see that there are 2 degrees of freedom in the electromagnetic wave, and not 3 as one might naively have supposed.

These 2 physical degrees of freedom can be organised as the + and - helicity states, just as we did in our earlier discussion. These are the circularly-polarised waves rotating anti-clockwise and clockwise, respectively. In other words, these are the states whose spin is either parallel, or anti-parallel, to the direction of propagation. One way of understanding why we have only 2, and not 3, allowed states is that the wave is travelling at the speed of light, and so it is not possible for it to have a helicity that projects other than fully parallel or anti-parallel to its direction of propagation.

We can make contact with the \( \phi = 0 \) gauge choice that we made in our previous discussion of electromagnetic waves. Starting in Lorenz gauge, we make use of the residual
gauge transformation (6.136) by choosing $h$ so that

$$a_0 - h k_0 = 0, \quad \text{i.e.} \quad h = \frac{-a_0}{\omega}. \quad (6.137)$$

this means that after performing the residual gauge transformation we shall have

$$a_0 = 0, \quad (6.138)$$

and so, from (6.129), we shall have

$$A_0 = 0, \quad \text{i.e.} \quad \phi = 0. \quad (6.139)$$

The original Lorenz gauge condition (6.125) then reduces to

$$\partial_i A_i = 0, \quad \text{i.e.} \quad \vec{\nabla} \cdot \vec{A} = 0. \quad (6.140)$$

This implies $\vec{k} \cdot \vec{A} = 0$, and so we have reproduced precisely the $\phi = 0, \vec{k} \cdot \vec{A} = 0$ gauge conditions that we used previously in our analysis of the general electromagnetic wave solutions. The choice $\phi = 0$ and $\vec{\nabla} \cdot \vec{A} = 0$ is known as Radiation Gauge.

In $D$ spacetime dimensions, the analogous result can easily be seen to be that the electromagnetic wave has $(D - 2)$ degrees of freedom.

### 6.7 Fourier decomposition of electrostatic fields

We saw earlier in 6.5 that an electromagnetic wave, expressed in the radiation gauge in terms of the 3-vector potential $\vec{A}$, could be decomposed into Fourier modes as in (6.88). For each mode $\vec{A}(\vec{k}, \lambda)$ in the sum, we have $\vec{e}_\lambda(\vec{k}) \cdot \vec{k} = 0$, and so each mode of the electric field $\vec{E}(\vec{k}, \lambda) = -\partial A(\vec{k}, \lambda)/\partial t$ satisfies the transversality condition

$$\vec{k} \cdot \vec{E}(\vec{k}, \lambda) = 0. \quad (6.141)$$

By contrast, an electrostatic field $\vec{E}$ is longitudinal. Consider, for example, a point charge at the origin, whose potential therefore satisfies

$$\nabla^2 \phi = -4\pi e \delta^3(\vec{r}). \quad (6.142)$$

We can express $\phi(\vec{r})$ in terms of its Fourier transform $\Phi(\vec{k})$ as

$$\phi(\vec{r}) = \int \frac{d^3 \vec{k}}{(2\pi)^3} \Phi(\vec{k}) e^{i\vec{k} \cdot \vec{r}}. \quad (6.143)$$

This is clearly a sum over zero-frequency waves, as one would expect since the fields are static.