Chapter 6

Electrodynamics and Relativity

6.1 Four-Vectors

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Using the Einstein convention, the transformation law (6.1.3) becomes

$$x'^{\mu} = \Lambda^{\mu}{}_{\nu} x^{\nu}, \tag{6.1.7}$$

and the expression for the space-time interval is

$$\ell^2 = x^{\mu} g_{\mu\nu} x^{\nu}. \tag{6.1.8}$$

Tensors (i.e. linear relations between a number of four-vectors) can be represented conveniently in this notation. For example, if four-vector x^{μ} is related to four-vectors y^{μ}, z^{μ} and w^{μ} through a linear relation (i.e., a rank 4 tensor), this can be expressed as

$$x^{\mu} = M^{\mu}{}_{\nu\rho\sigma} y^{\nu} z^{\rho} w^{\sigma}. \tag{6.1.9}$$

We can see that in this notation, a rank N tensor appears simply as an object with N Lorentz indices. In contrast, the matrix notation we used to represent the inertia tensor is only suitable for rank 2 tensor.

The component notation is also more flexible than the matrix notation. A product of two matrices (or rank 2 tensors) $\overline{\overline{\mathbf{A}}}$ and $\overline{\overline{\mathbf{B}}}$, with components $A^{\mu}{}_{\nu}$ and $B^{\mu}{}_{\nu}$, can be written as

$$(\overline{\overline{\mathbf{A}}} \cdot \overline{\overline{\mathbf{B}}})^{\mu}{}_{\nu} = A^{\mu}{}_{\rho} B^{\rho}{}_{\nu}. \tag{6.1.10}$$

In matrix multiplication, the order of the factors matters, $\overline{\overline{\mathbf{A}}} \cdot \overline{\overline{\mathbf{B}}} \neq \overline{\overline{\mathbf{B}}} \cdot \overline{\overline{\mathbf{A}}}$, but in the component notation the symbols represent matrix elements, which are real or complex numbers. Therefore we can change the order of the factors freely, i.e., $A^{\mu}{}_{\rho} B^{\rho}{}_{\nu} = B^{\rho}{}_{\nu} A^{\mu}{}_{\rho}$. The labelling of the indices keeps track of how the tensors are multiplied. To actually compute numerical values, it it often convenient to switch to the matrix notation, and it is then important to write the matrices in the right order. Note also that you can choose freely which Greek letter you use for each summation index, but the same letter can only be used once in one expression (i.e. once as a superscript and once as a subscript).

Under this transformation, the space-time interval transforms as

$$\ell^2 = x^{\mu} g_{\mu\nu} x^{\nu} \to x'^{\mu} g_{\mu\nu} x'^{\nu} = \Lambda^{\mu}{}_{\rho} x^{\rho} g_{\mu\nu} \Lambda^{\nu}{}_{\sigma} x^{\sigma} = x^{\rho} \Lambda^{\mu}{}_{\rho} g_{\mu\nu} \Lambda^{\nu}{}_{\sigma} x^{\sigma} = x^{\mu} \Lambda^{\rho}{}_{\mu} g_{\rho\sigma} \Lambda^{\sigma}{}_{\nu} x^{\nu}, \tag{6.1.11}$$

where, in the last step, we used the freedom to change the labelling of the summation indices and swapped $\mu \leftrightarrow \rho$ and $\nu \leftrightarrow \sigma$. In order for the transformation to leave the space-time interval ℓ^2 invariant, Eq. (6.1.11) has to be equal to $x^{\mu} g_{\mu\nu} x^{\nu}$, and this requires

$$\Lambda^{\rho}{}_{\mu} g_{\rho\sigma} \Lambda^{\sigma}{}_{\nu} = g_{\mu\nu}. \tag{6.1.12}$$

We can therefore use Eq. (6.1.12) as the definition of a Lorentz transformation.

Using the metric tensor, we can also define a scalar product of two four-vectors x and y as

$$x \cdot y = x^{\mu} g_{\mu\nu} y^{\nu} = x^0 y^0 - x^1 y^1 - x^2 y^2 - x^3 y^3. \tag{6.1.13}$$

This is also invariant under Lorentz transformations,

$$\ell^2 = x^{\mu} g_{\mu\nu} y^{\nu} \to x'^{\mu} g_{\mu\nu} y'^{\nu} = \Lambda^{\mu}{}_{\rho} x^{\rho} g_{\mu\nu} \Lambda^{\nu}{}_{\sigma} y^{\sigma} = x^{\mu} \Lambda^{\rho}{}_{\mu} g_{\rho\sigma} \Lambda^{\sigma}{}_{\nu} y^{\nu} = x^{\mu} g_{\mu\nu} y^{\nu}. \tag{6.1.14}$$

To simplify the notation further, we define a *covariant* vector x_{μ} by

$$x_{\mu} = g_{\mu\nu} x^{\nu} = (x^0, -x^1, -x^2, -x^3), \tag{6.1.15}$$

and indicate it by using a subscript index. We say that we use the metric to *lower the index*. The original position four-vector x^{μ} with a superscript index is called a *contravariant* vector. For example, the scalar product (6.1.13) is then simply

$$x \cdot y = x^{\mu} y_{\mu}. \tag{6.1.16}$$

To *raise the index*, i.e., turn a covariant vector back to a contravariant one, we need the inverse $\overline{\overline{\mathbf{g^{-1}}}}$ of the metric tensor, so that

$$x^{\mu} = (g^{-1})^{\mu\nu} x_{\nu}. \tag{6.1.17}$$

(Note that we use superscript indices to be consistent with the Einstein convention.) Eq. (6.1.17) is equivalent to saying that it is the inverse matrix of $g_{\mu\nu}$, defined in the usual way by

$$(g^{-1})^{\mu\nu} g_{\nu\rho} = \delta^{\mu}_{\rho}, \tag{6.1.18}$$

where

$$\delta^\mu_\rho = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{6.1.19}$$

is the 4×4 unit matrix.

For the Minkowski metric (6.1.5) it is easy to find the inverse, and it turns out to be the same matrix as the metric itself,

$$(g^{-1})^{\mu\nu} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} = g_{\mu\nu}, \tag{6.1.20}$$

but this is not the case in general relativity. In any case, using the definition of the inverse metric (6.1.18), we can see that it satisfies

$$g_{\mu\rho}g_{\nu\sigma}(g^{-1})^{\rho\sigma} = g_{\mu\nu}. \tag{6.1.21}$$

This means that when we lower the indices of the inverse metric $(g^{-1})^{\mu\nu}$ in the same way as in Eq. (6.1.15) we obtain the original metric $g_{\mu\nu}$. We can therefore think of the inverse metric $(g^{-1})^{\mu\nu}$ as simply the contravariant counterpart of the covariant metric $g_{\mu\nu}$. In particular, this means that there is no need to indicate the inverse metric by "-1" and we can simply write

$$(g^{-1})^{\mu\nu} = g^{\mu\nu} \tag{6.1.22}$$

without any risk of confusion. Then Eq. (6.1.18) becomes

$$g^{\mu\nu}g_{\nu\rho} = \delta^\mu_\rho, \tag{6.1.23}$$

and the expression for raising the index (6.1.17) simplifies to

$$x^\mu = g^{\mu\nu}x_\nu. \tag{6.1.24}$$

We can treat all Lorentz indices in this way, using $g_{\mu\nu}$ to lower a contravariant superscript index to a covariant subscript, and $g^{\mu\nu}$ to raise a covariant subscript index to a contravariant superscript. In particular, if we multiply both sides of Eq. (6.1.12) by $g^{\lambda\mu}$, we find

$$g^{\lambda\mu}\Lambda^\rho{}_\mu g_{\rho\sigma}\Lambda^\sigma{}_\nu = g^{\lambda\mu}g_{\mu\nu} = \delta^\lambda_\nu. \tag{6.1.25}$$

Comparing this with the definition of the *inverse Lorentz transformation* $\overline{\overline{\mathbf{\Lambda^{-1}}}}$, which takes the system back from the boosted to the original frame,

$$(\Lambda^{-1})^\lambda{}_\sigma\Lambda^\sigma{}_\nu = \delta^\lambda_\nu, \tag{6.1.26}$$

we find that

$$(\Lambda^{-1})^\lambda{}_\sigma = g^{\lambda\mu}\Lambda^\rho{}_\mu g_{\rho\sigma} \equiv \Lambda_\sigma{}^\lambda. \tag{6.1.27}$$

We can also derive the transformation law for covariant vectors,

$$x'_\mu = g_{\mu\nu}x'^\nu = g_{\mu\nu}\Lambda^\nu{}_\rho x^\rho = g_{\mu\nu}\Lambda^\nu{}_\rho g^{\rho\lambda}x_\lambda = \Lambda_\mu{}^\lambda x_\lambda. \tag{6.1.28}$$

Comparing with Eq. (6.1.27) we see that the transformation matrix for the covariant vectors is the inverse of the contravariant transformation matrix.

Besides the position four-vector x^μ, there are other quantities that transform in the same way under Lorentz transformations and can therefore be naturally written as four-vectors. These include

- The four-velocity u^μ, defined as
$$u^\mu = \frac{dx^\mu}{d\tau}, \tag{6.1.29}$$
where τ is the *proper time* is the time measured by the observer moving along the trajector defined by x^μ. It is defined as $c^2d\tau^2 = c^2dt^2 - dx^2 - dy^2 - dz^2$, or $dt = \gamma d\tau$. It follows that
$$u^\mu = (\gamma c, \gamma v_x, \gamma v_y, \gamma v_z). \tag{6.1.30}$$
- Four-momentum
$$p^\mu = mu^\mu = (E/c, p_x, p_y, p_z). \tag{6.1.31}$$
- Four-current density
$$j^\mu = nqu^\mu = (\gamma nqc, \gamma nqv_x, \gamma nqv_y, \gamma nqv_z) = (\rho c, j_x, j_y, j_z). \tag{6.1.32}$$

These all transform as contravariant vectors, i.e., $u'^\mu = \Lambda^\mu{}_\nu u^\nu$ etc., although of course we can always lower the index with the metric to turn them into the covariant form, $u_\mu = g_{\mu\nu}u^\nu$, when it is more convenient.

For an example of a four-vector that is more natural to think of as a covariant vector, consider a scalar function $f(x)$ of spacetime, and its derivative with respect to the contravariant position vector x^μ. Using the chain rule of derivatives, and the inverse Lorentz transformation $x^\nu = (\Lambda^{-1})^\nu{}_\mu x'^\mu = \Lambda_\mu{}^\nu x'^\mu$, we find

$$\frac{\partial f(x)}{\partial x'^\mu} = \sum_\nu \frac{\partial x^\nu}{\partial x'^\mu}\frac{\partial f(x)}{\partial x^\nu} = \Lambda_\mu{}^\nu \frac{\partial f(x)}{\partial x^\nu}. \tag{6.1.33}$$

Comparing with Eq. (6.1.28), we see that a derivative with respect to a contravariant vector transforms as a covariant vector. Therefore we use the notation

$$\partial_\mu \equiv \frac{\partial}{\partial x^\mu}, \tag{6.1.34}$$

to make this explicit. In this notation, Eq. (6.1.33) becomes

$$\partial'_\mu f = \Lambda_\mu{}^\nu \partial_\nu f. \tag{6.1.35}$$

Similarly, a derivative with respect to a covariant vector transforms as a contravariant vector, and therefore we write

$$\partial^\mu \equiv \frac{\partial}{\partial x_\mu}. \tag{6.1.36}$$

Remember that a tensor is a linear relationship between two or more vectors. In special relativity we expect that the same linear relationship remains valid in all reference frames, in which case it is called a *Lorentz tensor*. Consider, for example, the rank 4 tensor $M^\mu_{\nu\rho\sigma}$ in Eq. (6.1.9). Lorentz boosting the right-hand-side, we find

$$x'^\mu = \Lambda^\mu{}_\lambda x^\lambda = \Lambda^\mu{}_\lambda M^\lambda{}_{\nu\rho\sigma} y^\nu z^\rho w^\sigma = \Lambda^\mu{}_\lambda M^\lambda{}_{\nu\rho\sigma} \Lambda_\alpha{}^\nu y'^\alpha \Lambda_\beta{}^\rho z'^\beta \Lambda_\gamma{}^\sigma w'^\gamma, \tag{6.1.37}$$

where in the last step we used the inverse Lorentz transformation. We want to be able to write this as

$$x'^\mu = M'^\mu{}_{\alpha\beta\gamma} y'^\alpha z'^\beta w'^\gamma, \tag{6.1.38}$$

which means that the boosted tensor has to be

$$M'^\mu{}_{\alpha\beta\gamma} = \Lambda^\mu{}_\lambda \Lambda_\alpha{}^\nu \Lambda_\beta{}^\rho \Lambda_\gamma{}^\sigma M^\lambda{}_{\nu\rho\sigma} \tag{6.1.39}$$

We can see that each superscript index transforms with the contravariant transformation matrix, and each subscript index with the covariant transformation matrix, just like in four-vectors.

Whenever an index is summed over (*contracted*) according to the Einstein convention, the sum is Lorentz invariant, so summed indices can be ignored when doing Lorentz transformations. For example,

$$M'^\mu{}_{\mu\beta\gamma} = \Lambda^\mu{}_\lambda \Lambda_\mu{}^\nu \Lambda_\beta{}^\rho \Lambda_\gamma{}^\sigma M^\lambda{}_{\nu\rho\sigma} = \delta^\nu_\lambda \Lambda_\beta{}^\rho \Lambda_\gamma{}^\sigma M^\lambda{}_{\nu\rho\sigma} = \Lambda_\beta{}^\rho \Lambda_\gamma{}^\sigma M^\mu{}_{\mu\rho\sigma} \tag{6.1.40}$$

where we used the property (6.1.25). This shows why the Einstein convention is so useful in special relativity: Because the laws of nature are supposed to be the same in all inertial frames, pairs of indices should only appear in this Lorentz invariant form.

6.2 Vector Potential

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

$$\begin{aligned}
\nabla\cdot\mathbf{E} &= \frac{\rho}{\epsilon_0}, \quad \text{(Gauss's law)} \\
\nabla\times\mathbf{E} &= -\frac{\partial\mathbf{B}}{\partial t}, \quad \text{(Faraday's law)} \\
\nabla\cdot\mathbf{B} &= 0, \quad \text{(magnetic Gauss's law)} \\
\nabla\times\mathbf{B} &= \mu_0\mathbf{J} + \mu_0\epsilon_0\frac{\partial\mathbf{E}}{\partial t}. \quad \text{(Ampère's law)}
\end{aligned} \tag{6.2.1}$$

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

$$\mathbf{E} = -\nabla\phi\,. \tag{6.2.2}$$

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

$$\nabla^2\phi = -\frac{\rho}{\epsilon_0}. \tag{6.2.3}$$

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

On the other hand, Eq. (6.2.2) is clearly not sufficient in time-dependent problems, which can be seen by considering Faraday's law,

$$\nabla\times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}\,. \tag{6.2.4}$$

Using Eq. (6.2.2), we can write the curl of the electric field as

$$\nabla\times \mathbf{E} = -\nabla \times \nabla\phi\,, \tag{6.2.5}$$

but this is vanishes because the curl of a gradient is identically zero. Therefore Eqs. (6.2.2) and (6.2.5) are incompatible.

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

$$\begin{aligned} \mathbf{B} &= \nabla\times \mathbf{A}, \\ \mathbf{E} &= -\frac{\partial \mathbf{A}}{\partial t} - \nabla\phi. \end{aligned} \tag{6.2.6}$$

Let us see how Maxwell's equations (6.2.1) appear in terms of ϕ and $\mathbf{A}$. First, magnetic Gauss's law is trivally satisfied

$$\nabla \cdot \mathbf{B} = \nabla \cdot (\nabla\times \mathbf{A}) = 0, \tag{6.2.7}$$

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

Faraday's law is also automatically satisfied,

$$\nabla\times \mathbf{E} = \nabla \times \left(-\frac{\partial \mathbf{A}}{\partial t} - \nabla\phi\right) = -\frac{\partial\, \nabla\times \mathbf{A}}{\partial t} - \nabla\times \nabla\boldsymbol{\phi} = -\frac{\partial \mathbf{B}}{\partial t}. \tag{6.2.8}$$

Gauss's law becomes

$$\nabla \cdot \mathbf{E} = -\frac{\partial(\nabla \cdot \mathbf{A})}{\partial t} - \nabla^2\phi = \frac{\rho}{\epsilon_0}. \tag{6.2.9}$$

This is a non-trivial equation that the potentials $\mathbf{A}$ and ϕ have to satisfy. It is essentially Poisson's equation (6.2.3) with an additional term.

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

Finally, Ampère's law becomes

$$\boldsymbol{\nabla}\times \mathbf{B} = \boldsymbol{\nabla} \times (\boldsymbol{\nabla}\times \mathbf{A}) = \mu_0\mathbf{J} + \mu_0\epsilon_0\frac{\partial \mathbf{E}}{\partial t} = \mu_0\mathbf{J} - \mu_0\epsilon_0\boldsymbol{\nabla}\frac{\partial\phi}{\partial t} - \mu_0\epsilon_0\frac{\partial^2\mathbf{A}}{\partial t^2}. \tag{6.2.10}$$

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

$$\frac{1}{c^2}\frac{\partial^2\mathbf{A}}{\partial t^2} = -\boldsymbol{\nabla} \times (\boldsymbol{\nabla}\times \mathbf{A}) - \frac{1}{c^2}\boldsymbol{\nabla}\frac{\partial\phi}{\partial t} + \mu_0\mathbf{J} = \nabla^2\mathbf{A} - \boldsymbol{\nabla}\left(\boldsymbol{\nabla}\cdot\mathbf{A} + \frac{1}{c^2}\frac{\partial\phi}{\partial t}\right) + \mu_0\mathbf{J}. \tag{6.2.11}$$

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

Eq. (6.2.11) has the form of a wave equation with some additional terms, so we can try to look for plane wave solutions. As an Ansatz, let us consider a plane wave polarised in the x direction and travelling in the z direction at the speed of light

$$\begin{aligned} \mathbf{A} &= A_0 e^{ik(z-ct)}\hat{\mathbf{x}}, \\ \phi &= 0. \end{aligned} \tag{6.2.12}$$

Substituting this to Eq. (6.2.11), we find

$$\frac{1}{c^2}\frac{\partial^2\mathbf{A}}{\partial t^2} - \nabla^2\mathbf{A} - \boldsymbol{\nabla}\left(\boldsymbol{\nabla}\cdot\mathbf{A} - \frac{1}{c^2}\frac{\partial\phi}{\partial t}\right) - \mu_0\mathbf{J} = \frac{1}{c^2}\frac{\partial^2\mathbf{A}}{\partial t^2} - \frac{\partial^2\mathbf{A}}{\partial z^2} = -k^2\mathbf{A} + k^2\mathbf{A} = 0, \tag{6.2.13}$$

so this satisfies Ampère's law. Using Eq. (6.2.6), we find the magnetic and electric fields

$$\begin{aligned} \mathbf{B} &= \boldsymbol{\nabla}\times\mathbf{A} = \begin{vmatrix} \hat{\mathbf{x}} & \hat{\mathbf{y}} & \hat{\mathbf{z}} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ A_x & 0 & 0 \end{vmatrix} = \frac{\partial A_x}{\partial z}\hat{\mathbf{y}} = ikA_0e^{ik(z-ct)}\hat{\mathbf{y}}, \\ \mathbf{E} &= -\frac{\partial\mathbf{A}}{\partial t} - \boldsymbol{\nabla}\phi = ikcA_0e^{ik(z-ct)}\hat{\mathbf{x}}. \end{aligned} \tag{6.2.14}$$

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

6.3 Gauge Invariance

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

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

$$\begin{aligned} \mathbf{A} &\to \mathbf{A} + \boldsymbol{\alpha}(\mathbf{x}, t), \\ \phi &\to \phi + f(\mathbf{x}, t). \end{aligned} \tag{6.3.1}$$

This would change the electric and magnetic fields by

$$\begin{aligned}\mathbf{B} &\rightarrow \boldsymbol{\nabla} \times (\mathbf{A} + \boldsymbol{\alpha}) = \boldsymbol{\nabla}\times \mathbf{A} + \boldsymbol{\nabla}\times \boldsymbol{\alpha} = \mathbf{B} + \boldsymbol{\nabla}\times \boldsymbol{\alpha}, \\ \mathbf{E} &\rightarrow -\frac{\partial(\mathbf{A}+\boldsymbol{\alpha})}{\partial t} - \boldsymbol{\nabla}(\phi + f) = \mathbf{E} - \frac{\partial \boldsymbol{\alpha}}{\partial t} - \boldsymbol{\nabla} f. \end{aligned} \tag{6.3.2}$$

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

$$\boldsymbol{\nabla}\times \boldsymbol{\alpha} = 0, \quad \text{and} \quad \frac{\partial \boldsymbol{\alpha}}{\partial t} + \boldsymbol{\nabla} f = 0. \tag{6.3.3}$$

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

$$\boldsymbol{\alpha} = \boldsymbol{\nabla}\lambda. \tag{6.3.4}$$

The second condition in Eq. (6.3.3) then becomes

$$\boldsymbol{\nabla} f = -\frac{\partial \boldsymbol{\alpha}}{\partial t} = -\boldsymbol{\nabla}\frac{\partial \lambda}{\partial t}. \tag{6.3.5}$$

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

$$\begin{aligned}\mathbf{A} &\rightarrow \mathbf{A} + \boldsymbol{\nabla}\lambda, \\ \phi &\rightarrow \phi - \frac{\partial \lambda}{\partial t}, \end{aligned} \tag{6.3.6}$$

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

Note that while $\mathbf{E}$ and $\mathbf{B}$ are invariant under gauge transformations, Eqs. (6.2.9) and (6.2.11) are not. This means that we can use a gauge transformation to make those equations simpler and easier to solve. This is known as fixing the gauge. For example, the divergence $\boldsymbol{\nabla} \cdot \mathbf{A}$ is not gauge invariant but transforms as

$$\boldsymbol{\nabla} \cdot \mathbf{A} \rightarrow \boldsymbol{\nabla} \cdot (\mathbf{A} + \boldsymbol{\nabla}\lambda) = \boldsymbol{\nabla} \cdot \mathbf{A} + \nabla^2\lambda. \tag{6.3.7}$$

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

One popular way to fix the gauge is the *Coulomb gauge*, in which $\boldsymbol{\nabla} \cdot \mathbf{A} = 0$. In this gauge the non-trivial Maxwell equations (6.2.9) and (6.2.11) become

$$\begin{aligned}\nabla^2\phi &= -\frac{\rho}{\epsilon_0}, \\ \frac{1}{c^2}\frac{\partial^2 \mathbf{A}}{\partial t^2} &= \nabla^2\mathbf{A} - \frac{1}{c^2}\boldsymbol{\nabla}\frac{\partial \phi}{\partial t} + \mu_0\mathbf{J}. \end{aligned} \tag{6.3.8}$$

The main benefit of this gauge is that the equations are simpler: The first equation is simply the familiar Poisson equation, and the second equation is a wave equation with a source term.

However, the drawback is that the Poisson equation appears to violate causality because a change in the charge distribution affects the scalar potential immediately at all distances. This is not a serious problem because ϕ is not observable, and the observable fields $\mathbf{E}$ and $\mathbf{B}$ still behave causally.

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

$$\boldsymbol{\nabla}\cdot\mathbf{A}+\frac{1}{c^2}\frac{\partial\phi}{\partial t}=0. \tag{6.3.9}$$

In this gauge, the equations of motion are

$$\begin{aligned}\frac{1}{c^2}\frac{\partial^2\phi}{\partial t^2}-\nabla^2\phi &= \frac{\rho}{\epsilon_0},\\ \frac{1}{c^2}\frac{\partial^2\mathbf{A}}{\partial t^2}-\nabla^2\mathbf{A} &= \mu_0\mathbf{J}.\end{aligned} \tag{6.3.10}$$

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

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

$$\begin{aligned}\frac{\partial}{\partial t}\left(\boldsymbol{\nabla}\cdot\mathbf{A}\right) &= -\frac{\rho}{\epsilon_0},\\ \frac{1}{c^2}\frac{\partial^2\mathbf{A}}{\partial t^2} &= \nabla^2\mathbf{A}-\boldsymbol{\nabla}\left(\boldsymbol{\nabla}\cdot\mathbf{A}\right)+\mu_0\mathbf{J}.\end{aligned} \tag{6.3.11}$$

6.4 Relativistic Electrodynamics

Historically, electrodynamics played a key role in the development of the theory of relativity, and electrodynamics appears much more elegent in a fully relativistic formulation. However, it is not entirely trivial to write electric and magnetic fields in a four-vector form. For example, when moving from one frame of reference to another, the electric and magnetic field transform into one another and are therefore not independent entities. They cannot be two separate four-vectors.

To derive the relativistic formulation, we start from the expression for the Lorentz force,

$$\frac{d\mathbf{p}}{dt}=\mathbf{F}=q(\mathbf{E}+\mathbf{v}\times\mathbf{B}). \tag{6.4.1}$$

It follows directly that the derivative with respect to the proper time is

$$\frac{d\mathbf{p}}{d\tau}=\gamma\frac{d\mathbf{p}}{dt}=q\left(\frac{u^0}{c}\mathbf{E}+\mathbf{u}\times\mathbf{B}\right), \tag{6.4.2}$$

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

where $\mathbf{u} = \gamma\mathbf{v}$ is the spatial part of the four-velocity u^μ.

The time derivative of the energy of the particle is

$$\frac{dE}{dt} = \mathbf{F}\cdot\mathbf{v} = q\mathbf{E}\cdot\mathbf{v}, \tag{6.4.3}$$

from which we obtain the derivative with respect to the proper time as

$$\frac{dE}{d\tau} = \gamma\frac{dE}{dt} = q\mathbf{E}\cdot\mathbf{u}. \tag{6.4.4}$$

We can now combine Eqs. (6.4.2) and (6.4.4) into the proper time derivative of the four-momentum $p^\mu = (E/c, \mathbf{p})$,

$$\frac{dp^\mu}{d\tau} = \frac{d}{d\tau}\begin{pmatrix} E/c \\ p_x \\ p_y \\ p_z \end{pmatrix} = q\begin{pmatrix} 0 & E_x/c & E_y/c & E_z/c \\ E_x/c & 0 & B_z & -B_y \\ E_y/c & -B_z & 0 & B_x \\ E_z/c & B_y & -B_x & 0 \end{pmatrix}\begin{pmatrix} u^0 \\ u^1 \\ u^2 \\ u^3 \end{pmatrix}. \tag{6.4.5}$$

The matrix appearing in this expression is called the *Faraday tensor* or the field-strength tensor and denoted by $F^\mu{}_\nu$, so we can write more compactly the relativistic Lorentz force equation as

$$\frac{dp^\mu}{d\tau} = qF^\mu{}_\nu u^\nu. \tag{6.4.6}$$

The Faraday tensor is often written with two contravariant indices as

$$F^{\mu\nu} = F^\mu{}_\rho g^{\rho\nu} = \begin{pmatrix} 0 & -E_x/c & -E_y/c & -E_z/c \\ E_x/c & 0 & -B_z & B_y \\ E_y/c & B_z & 0 & -B_x \\ E_z/c & -B_y & B_x & 0 \end{pmatrix}. \tag{6.4.7}$$

The Lorentz force equation (6.4.6) then becomes

$$\frac{dp^\mu}{d\tau} = qF^{\mu\nu}u_\nu. \tag{6.4.8}$$

Note that this tensor is antisymmetric, $F^{\nu\mu} = -F^{\mu\nu}$.

In order for the right-hand-side of Eq. (6.4.6) to be Lorentz contravariant, $F^{\mu\nu}$ has to transform as a contravariant rank 2 Lorentz tensor,

$$F'^{\mu\nu} = \Lambda^\mu{}_\rho\Lambda^\nu{}_\sigma F^{\rho\sigma}. \tag{6.4.9}$$

This tells us how the electric and magnetic fields must transform. For example, considering a boost in z direction,

$$\Lambda^\mu{}_\rho = \begin{pmatrix} \gamma & 0 & 0 & -\beta\gamma \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -\beta\gamma & 0 & 0 & \gamma \end{pmatrix}, \tag{6.4.10}$$

we find that the electric and magnetic fields transform as

$$\begin{aligned} E'_x &= \gamma\left(E_x - vB_y\right), \\ E'_y &= \gamma\left(E_y + vB_x\right), \\ E'_z &= E_z\,, \end{aligned} \tag{6.4.11}$$

$$\begin{aligned} B'_x &= \gamma\left(B_x + vE_y/c^2\right), \\ B'_y &= \gamma\left(B_y - vE_x/c^2\right), \\ B'_z &= B_z\,. \end{aligned} \tag{6.4.12}$$

More generally, we can write

$$\begin{aligned} \mathbf{E}'_\parallel &= \mathbf{E}_\parallel, \\ \mathbf{B}'_\parallel &= \mathbf{B}_\parallel, \\ \mathbf{E}'_\perp &= \gamma\left(\mathbf{E}_\perp + \mathbf{v}\times\mathbf{B}\right), \\ \mathbf{B}'_\perp &= \gamma\left(\mathbf{B}_\perp - \mathbf{v}\times\mathbf{E}/c^2\right) = \gamma\left(\mathbf{B}_\perp - \mu_0\epsilon_0\mathbf{v}\times\mathbf{E}\right)\,, \end{aligned} \tag{6.4.13}$$

where $\parallel$ refers to the component parallel to the boost velocity, and $\perp$ to the perpendicular components. If $\mathbf{B} = 0$, then

$$\mathbf{B}' = -\frac{\mathbf{v}\times\mathbf{E}'}{c^2}, \tag{6.4.14}$$

in agreement with Eq. (**??**) (note the opposite sign of $\mathbf{v}$).

6.5 Maxwell's Equations

Now that we have combined the electric and magnetic fields into one Lorentz tensor $F^{\mu\nu}$, we want to write Maxwell's equations (6.2.1) in terms of it. We start by noting that, according to Eq. (6.4.7), the electric and magnetic fields are given by

$$\begin{aligned} \mathbf{E} &= (cF^{10}, cF^{20}, cF^{30}), \\ \mathbf{B} &= (F^{32}, F^{13}, F^{21}). \end{aligned} \tag{6.5.1}$$

We now write Gauss's law as

$$\begin{aligned} \boldsymbol{\nabla}\cdot\mathbf{E} - \frac{\rho}{\epsilon_0} &= \frac{\partial E_x}{\partial x} + \frac{\partial E_y}{\partial y} + \frac{\partial E_z}{\partial z} - \frac{\rho}{\epsilon_0} = c\left(\frac{\partial F^{10}}{\partial x} + \frac{\partial F^{20}}{\partial y} + \frac{\partial F^{30}}{\partial z} - \frac{\rho}{c\epsilon_0}\right) \\ &= c\left(\partial_1 F^{10} + \partial_2 F^{20} + \partial_3 F^{30} - \frac{\rho}{c\epsilon_0}\right) = c\left(\partial_\mu F^{\mu 0} - \frac{\rho}{c\epsilon_0}\right). \end{aligned} \tag{6.5.2}$$

Therefore we have

$$\partial_\mu F^{\mu 0} = \frac{\rho}{c\epsilon_0} = \mu_0 c\rho. \tag{6.5.3}$$

To deal with Ampère's law,

$$\boldsymbol{\nabla}\times\mathbf{B} - \mu_0\mathbf{J} - \mu_0\epsilon_0\frac{\partial\mathbf{E}}{\partial t} = 0 \tag{6.5.4}$$

we write the x component

$$\begin{aligned}\frac{\partial B_z}{\partial y}-\frac{\partial B_y}{\partial z}-\mu_0 j_x-\mu_0\epsilon_0\frac{\partial E_x}{\partial t} &= \frac{\partial F^{21}}{\partial y}-\frac{\partial F^{13}}{\partial z}-\mu_0 j_x-\frac{1}{c}\frac{\partial F^{10}}{\partial t}\\ &= \partial_0 F^{01}+\partial_2 F^{21}+\partial_3 F^{31}-\mu_0 j_x\\ &= \partial_\mu F^{\mu 1}-\mu_0 j_x=0,\end{aligned} \tag{6.5.5}$$

so we have

$$\partial_\mu F^{\mu 1}=\mu_0 j_x, \tag{6.5.6}$$

and similarly for the other components. We can now write Eqs. (6.5.3) and (6.5.6) as one equation in terms of the four-current $j^\mu=(c\rho,\mathbf{j})$,

$$\partial_\mu F^{\mu\nu}=\mu_0 j^\nu. \tag{6.5.7}$$

The magnetic Gauss's law reads

$$\begin{aligned}\boldsymbol{\nabla}\cdot\mathbf{B} &= \frac{\partial B_x}{\partial x}+\frac{\partial B_y}{\partial y}+\frac{\partial B_z}{\partial z}=\frac{\partial F^{32}}{\partial x}+\frac{\partial F^{13}}{\partial y}+\frac{\partial F^{21}}{\partial z}=\partial_1 F^{32}+\partial_2 F^{13}+\partial_3 F^{21}\\ &= \partial^1 F^{23}+\partial^2 F^{31}+\partial^3 F^{12}=0.\end{aligned} \tag{6.5.8}$$

For Faraday's law

$$\boldsymbol{\nabla}\times\mathbf{E}+\frac{\partial\mathbf{B}}{\partial t}=0, \tag{6.5.9}$$

we again take the x component,

$$\begin{aligned}\frac{\partial B_x}{\partial t}+\frac{\partial E_z}{\partial y}-\frac{\partial E_y}{\partial z} &= \frac{\partial F^{32}}{\partial t}+c\frac{\partial F^{30}}{\partial y}-c\frac{\partial F^{20}}{\partial z}=c\left(\partial_0 F^{32}-\partial_2 F^{03}-\partial_3 F^{20}\right)\\ &= c\left(\partial^0 F^{32}+\partial^2 F^{03}+\partial^3 F^{20}\right)=0\end{aligned} \tag{6.5.10}$$

By comparing Eqs. (6.5.8) and (6.5.10), we note that we can combine them into one equation

$$\partial^\mu F^{\nu\rho}+\partial^\nu F^{\rho\mu}+\partial^\rho F^{\mu\nu}=0. \tag{6.5.11}$$

Thus, we have found that in the four-vector notation, the four Maxwell's equations (6.2.1) can be expressed by just two equations (6.5.7) and (6.5.11).

6.6 Four-vector Potential

In Section 6.2, we saw that we can express the electric and magnetic fields using the scalar and vector potentials, and that reduced the number of non-trivial Maxwell's equations from four to two. We will now see that we can do the same to the Faraday tensor $F^{\mu\nu}$, and that way reduce Maxwell's equations to just one.

Using Eq. (6.2.6), we can write

$$F^{10}=\frac{E_x}{c}=-\frac{1}{c}\left(\frac{\partial A_x}{\partial t}+\frac{\partial\phi}{\partial x}\right)=-\partial_0 A_x-\partial_1\left(\frac{\phi}{c}\right)=\partial^1\left(\frac{\phi}{c}\right)-\partial^0 A_x, \tag{6.6.1}$$

and
$$F^{21} = B_z = \frac{\partial A_y}{\partial x} - \frac{\partial A_x}{\partial y} = \partial_1 A_y - \partial_2 A_x = \partial^2 A_x - \partial^1 A_y. \tag{6.6.2}$$

By defining the *four-vector potential* $A^\mu = (\phi/c, \mathbf{A})$, we can write these two equations as

$$\begin{aligned} F^{10} &= \partial^1 A^0 - \partial^0 A^1 \\ F^{21} &= \partial^2 A^1 - \partial^1 A^2. \end{aligned} \tag{6.6.3}$$

Similar relations apply to other components of $F^{\mu\nu}$, so we can combine them into one equation

$$F^{\mu\nu} = \partial^\mu A^\nu - \partial^\nu A^\mu. \tag{6.6.4}$$

Using this equation, we can write the Faraday tensor in terms of the four-vector potential. Finally, we want to express Maxwell's equations (6.5.7) and (6.5.11) in terms of A^μ. Eq. (6.5.7) becomes

$$\partial_\mu F^{\mu\nu} = \partial_\mu \partial^\mu A^\nu - \partial^\nu \partial_\mu A^\mu = \mu_0 j^\nu. \tag{6.6.5}$$

For Eq. (6.5.11), we find that

$$\partial^\mu F^{\nu\rho} + \partial^\nu F^{\rho\mu} + \partial^\rho F^{\mu\nu} = \partial^\mu\partial^\nu A^\rho - \partial^\mu\partial^\rho A^\nu + \partial^\nu\partial^\rho A^\mu - \partial^\nu\partial^\mu A^\rho + \partial^\rho\partial^\mu A^\nu - \partial^\rho\partial^\nu A^\mu = 0 \tag{6.6.6}$$

identically, because each term appears twice with opposite signs. Therefore, Eq. (6.5.11) is automatically satisfied when the Faraday tensor is expressed in terms of the four-vector potential. The only non-trivial equation is therefore Eq. (6.6.5).

6.7 Lagrangian for Electrodynamics

Finally, let us see how we can describe electrodynamics in the Lagrangian formulation. Because the electromagnetic fields are continuous fields, we need to find a Lagrangian density $\mathcal{L}$ as defined in Section 3.6, and we will express it in terms of the four-vector potential A^μ.

We know that electrodynamics is invariant under both gauge and Lorentz transformations. Therefore the Lagrangian density $\mathcal{L}$ has to be a Lorentz scalar, and to be gauge invariant it can only depend on the four-vector potential through the Faraday tensor $F^{\mu\nu}$. It also makes sense to demand that the Lagrangian density should contain only first time derivatives and they should appear only in quadratic form. This corresponds to a "natural" system as defined in Section 3.5.1, and it ensures that the Euler-Lagrange equations have the familiar form. The expression that satisfies these requirements is $F^{\mu\nu}F_{\mu\nu}$, and therefore we are led to consider a Lagrangian density of the form

$$\mathcal{L} = aF^{\mu\nu}F_{\mu\nu}, \tag{6.7.1}$$

where a is some constant. Actually, the numerical value of a does not matter because it will drop out the Euler-Lagrange equation, but we will see later that the sign should be negative. It is conventional to choose $a = -1/4$, so that we have

$$\mathcal{L} = -\frac{1}{4}F^{\mu\nu}F_{\mu\nu}. \tag{6.7.2}$$

In terms of the four-vector potential, this becomes

$$\mathcal{L} = -\frac{1}{4}(\partial^\mu A^\nu - \partial^\nu A^\mu)(\partial_\mu A_\nu - \partial_\nu A_\mu) = -\frac{1}{2}\partial^\mu A^\nu(\partial_\mu A_\nu - \partial_\nu A_\mu). \tag{6.7.3}$$

In order to derive the Euler-Lagrange equation, we first write Eq. (3.6.12) in a four-vector form,

$$\partial_\mu \frac{\partial \mathcal{L}}{\partial(\partial_\mu y)} - \frac{\partial \mathcal{L}}{\partial y} = 0, \tag{6.7.4}$$

and generalise it to the current case by replacing y with A_ν. Because the Lagrangian $\mathcal{L}$ depends only on its derivatives, we find the equation

$$\partial_\mu \frac{\partial \mathcal{L}}{\partial(\partial_\mu A_\nu)} = 0. \tag{6.7.5}$$

To avoid any confusion about the derivative in this expression, it is best to lower all the indices in Eq. (6.7.3) and write it in the form

$$\mathcal{L} = -\frac{1}{2}g^{\kappa\rho}g^{\lambda\sigma}\partial_\kappa A_\lambda(\partial_\rho A_\sigma - \partial_\sigma A_\rho). \tag{6.7.6}$$

Then the derivative is easy to take by noting that it is non-zero only if the Lorentz indices match, that is,

$$\frac{\partial\left(\partial_\kappa A_\lambda\right)}{\partial\left(\partial_\mu A_\nu\right)} = \delta^\mu_\kappa \delta^\nu_\lambda. \tag{6.7.7}$$

We find

$$\begin{aligned}
\frac{\partial \mathcal{L}}{\partial(\partial_\mu A_\nu)} &= -\frac{1}{2}g^{\kappa\rho}g^{\lambda\sigma}\left[\delta^\mu_\kappa\delta^\nu_\lambda(\partial_\rho A_\sigma - \partial_\sigma A_\rho) + \partial_\kappa A_\lambda\left(\delta^\mu_\rho\delta^\nu_\sigma - \delta^\mu_\sigma\delta^\nu_\rho\right)\right] \\
&= -\frac{1}{2}\left[g^{\mu\rho}g^{\nu\sigma}(\partial_\rho A_\sigma - \partial_\sigma A_\rho) + \partial_\kappa A_\lambda\left(g^{\mu\kappa}g^{\nu\lambda} - g^{\mu\lambda}g^{\nu\kappa}\right)\right] \\
&= -\frac{1}{2}\left[\partial^\mu A^\nu - \partial^\nu A^\mu + \partial^\mu A^\nu - \partial^\nu A^\mu\right] = -F^{\mu\nu},
\end{aligned} \tag{6.7.8}$$

and therefore the Euler-Lagrange equation (6.7.5) is

$$\partial_\mu \frac{\partial \mathcal{L}}{\partial(\partial_\mu A_\nu)} = -\partial_\mu F^{\mu\nu} = 0. \tag{6.7.9}$$

Which is exactly the Maxwell equation (6.5.7) in vacuum, i.e., with $j^\mu = 0$. Because the other Maxwell equation (6.5.11) is satisfied identically when using the four-vector potential A^μ, we have shown that the laws of electrodynamics in vacuum are correctly described by the Lagrangian (6.7.3) which we obtained by assuming essentially only gauge and Lorentz invariance. This demonstrates how powerful symmetry considerations can be in physics, and in fact the properties of the other fundamental interactions (strong and weak nuclear force, and gravity) are also determined by their corresponding gauge invariances.