Maxwell's equations are:

$$\nabla \cdot \mathbf{D} = \rho$$
$$\nabla \cdot \mathbf{B} = 0$$
$$\nabla \wedge \mathbf{E} + \frac{\partial \mathbf{B}}{\partial t} = 0$$
$$\nabla \wedge \mathbf{H} - \frac{\partial \mathbf{D}}{\partial t} = \mathbf{J}$$

In these equations:

ρ and $\mathbf{J}$ are the density and flux of *free charge;*
$\mathbf{E}$ and $\mathbf{B}$ are the fields that exert forces on charges and currents ($\mathbf{f}$ is force per unit volume):

$$\mathbf{f} = \rho\mathbf{E} + \mathbf{J} \wedge \mathbf{B}; \qquad \text{or for a point charge} \qquad \mathbf{F} = q(\mathbf{E} + \mathbf{v} \wedge \mathbf{B});$$

$\mathbf{D}$ and $\mathbf{H}$ are related to $\mathbf{E}$ and $\mathbf{B}$ but include contributions from charges and currents bound within atoms:

$$\mathbf{D} = \epsilon_0\mathbf{E} + \mathbf{P} \qquad \mathbf{H} = \mathbf{B}/\mu_0 - \mathbf{M};$$

$\mathbf{P}$ is the *polarization* and $\mathbf{M}$ the *magnetization* of the matter.
μ_0 has a defined value of $4\pi\, 10^{-7}\,\mathrm{kg\,m\,C^{-2}}$.

Maxwell's equations in this form apply to spatial averages (over regions of atomic size) of the fundamental charges, currents and fields. This averaging generates a division of the charges and currents into two classes: the free charges, represented by ρ and $\mathbf{J}$, and charges and currents in atoms, whose averaged effects are represented by $\mathbf{P}$ and $\mathbf{M}$. We can see what these effects are by substituting for $\mathbf{D}$ and $\mathbf{H}$:

$$\nabla \cdot \mathbf{E} = (\rho - \nabla \cdot \mathbf{P})/\epsilon_0$$
$$\nabla \cdot \mathbf{B} = 0$$
$$\nabla \wedge \mathbf{E} + \frac{\partial \mathbf{B}}{\partial t} = 0$$
$$\nabla \wedge \mathbf{B} - \epsilon_0\mu_0\frac{\partial \mathbf{E}}{\partial t} = \mu_0\left(\mathbf{J} + \frac{\partial \mathbf{P}}{\partial t} + \nabla \wedge \mathbf{M}\right).$$

We see that the divergence of $\mathbf{P}$ generates a charge density: $\rho_b = -\nabla \cdot \mathbf{P}$
and the curl of $\mathbf{M}$ and temporal changes in $\mathbf{P}$ generate current: $\mathbf{J}_b = \frac{\partial \mathbf{P}}{\partial t} + \nabla \wedge \mathbf{M}$.
Any physical model of the atomic charges and currents will produce these spatially averaged effects (see problems).

This suggests that if we do not make the spatial average we can treat *all* charges and currents on the same basis, and obtain the fundamental classical equations relating charges, currents and electromagnetic fields, *the Microscopic Maxwell Equations:*

$$\nabla \cdot \mathbf{E} = \rho/\epsilon_0 \qquad \text{M1}$$

$$\nabla \cdot \mathbf{B} = 0 \qquad \text{M2}$$

$$\nabla \wedge \mathbf{E} + \frac{\partial \mathbf{B}}{\partial t} = 0 \qquad \text{M3}$$

$$\nabla \wedge \mathbf{B} - \epsilon_0 \mu_0 \frac{\partial \mathbf{E}}{\partial t} = \mu_0 \mathbf{J} \qquad \text{M4}$$

These are the equations we shall be working with from now on, and we shall make no further reference to **D**, **H**, **P** or **M**.

The physical interpretation is provided by the force density:

$$\boxed{\mathbf{f} = \rho \mathbf{E} + \mathbf{J} \wedge \mathbf{B}.}$$

We next look at three simple consequences.

Local Conservation of Charge

Take $1/\mu_0$ div M4+ϵ_0 $\partial/\partial t$ M1 $\rightarrow \nabla \cdot \mathbf{J} + \partial\rho/\partial t = 0$.

This equation is known as the continuity equation and implies *local conservation of charge.*

Consider an arbitrary volume V in a current flow. The total charge inside V at time t is

$$Q(t) = \int_V \rho(\mathbf{r}, t)\, d^3\mathbf{r}$$

If charge is locally conserved then the only way in which charge can leave V is by flowing through the bounding surface S:

$$-\frac{dQ}{dt} = \int_V -\frac{\partial \rho}{\partial t}\, d^3\mathbf{r} = \int_S \mathbf{J} \cdot d\mathbf{S} = \int_V \nabla \cdot \mathbf{J}\, d^3\mathbf{r}.$$

This must hold for an arbitrary volume, so the integrand must vanish:

$$\boxed{\frac{\partial \rho}{\partial t} + \nabla \cdot \mathbf{J} = 0}$$

This ensures the local conservation of charge, and is the prototype for other local conservation laws, which have a density term, a flux term, and in general a loss or source term as well if the quantity is not absolutely conserved but can transform into something else.

Conservation of Energy (Poynting's Theorem)

We can find a local conservation law for energy in the electromagnetic field.
Take $\mathbf{B}\cdot$ (M3) and subtract $\mathbf{E}\cdot$ (M4):

$$\Big(\mathbf{B}\cdot(\nabla\wedge\mathbf{E})-\mathbf{E}\cdot(\nabla\wedge\mathbf{B})\Big)+\left(\mathbf{B}\cdot\frac{\partial\mathbf{B}}{\partial t}+\epsilon_0\mu_0\mathbf{E}\cdot\frac{\partial\mathbf{E}}{\partial t}\right)+\mu_0\mathbf{E}\cdot\mathbf{J}=0$$

We can combine the terms in the first bracket as $\nabla\cdot(\mathbf{E}\wedge\mathbf{B})$.
We can rewrite the partial derivatives using $\mathbf{E}\cdot\dfrac{\partial\mathbf{E}}{\partial t}=\dfrac{1}{2}\dfrac{\partial E^2}{\partial t}$ and similarly for the $\mathbf{B}$ terms.
Thus we have a conservation law with a loss/source term:

$$\boxed{\nabla\cdot\mathbf{S}+\frac{\partial u}{\partial t}+\mathbf{E}\cdot\mathbf{J}=0}$$

where:

$\mathbf{S}=\dfrac{1}{\mu_0}\mathbf{E}\wedge\mathbf{B}$ is Poynting's vector and represents the energy flux in the field;

$u=\dfrac{\epsilon_0}{2}E^2+\dfrac{1}{2\mu_0}B^2$ is the energy density in the field;

$\mathbf{E}\cdot\mathbf{J}$ represents the rate per unit volume of energy loss from the EM field to the matter, represented by the current (or from matter to field if negative).

Wave Equations for E and B

Eliminate $\mathbf{B}$ by taking $\partial/\partial t$ of M4 and subtracting the curl of M3:

$$-\epsilon_0\mu_0\frac{\partial^2\mathbf{E}}{\partial t^2}-\nabla\wedge\nabla\wedge\mathbf{E}=\mu_0\frac{\partial\mathbf{J}}{\partial t}.$$

Using the vector identity curl curl = grad div $-\nabla^2$, and substituting from M1:

$$\nabla^2\mathbf{E}-\epsilon_0\mu_0\frac{\partial^2\mathbf{E}}{\partial t^2}=\mu_0\frac{\partial\mathbf{J}}{\partial t}+\nabla\rho/\epsilon_0$$

which is the inhomogeneous wave equation with wave speed c where

$$\boxed{c^2=\frac{1}{\epsilon_0\mu_0}}$$

and the source term is $\mu_0\dfrac{\partial\mathbf{J}}{\partial t}+\nabla\rho/\epsilon_0$.

Similarly we can eliminate $\mathbf{E}$: $\quad\nabla^2\mathbf{B}-\epsilon_0\mu_0\dfrac{\partial^2\mathbf{B}}{\partial t^2}=-\mu_0\nabla\wedge\mathbf{J}.$

The word *COVARIANT,* in this context, means the same as *form invariant:*
some transformation of the quantities in the equations produces *no change* in the *form* of the equations.

Some of the transformations are induced by physical operations (rotations, changes of velocity) that can be understood in either an *active* or *passive* sense; for example

the system is rotated in space (ACTIVE)

or *the same system is described relative to two different co-ordinate systems.* (PASSIVE)

I shall take ALL TRANSFORMATIONS IN A PASSIVE SENSE.

We shall discuss the covariance of Maxwell's equations under the following transformations:

Change of Units (!)

Rotation (including Reflection)

Duality Transformation

Gauge Transformation

Lorentz Transformation

In Maxwell's equations there are four places we can put constants. However, to understand all the historical arguments we must assume the existence of magnetic charge and current, which increases the number of constants to six:

$$\nabla \cdot \mathbf{E} = \alpha\rho$$
$$\nabla \cdot \mathbf{B} = \kappa\rho_m$$
$$\nabla \wedge \mathbf{E} + \beta\frac{\partial \mathbf{B}}{\partial t} = -\lambda\mathbf{J}_m$$
$$\nabla \wedge \mathbf{B} - \gamma\frac{\partial \mathbf{E}}{\partial t} = \delta\mathbf{J}$$

The six constants α, β, γ, δ, κ and λ are not all independently variable.

If we re-work the derivation of the wave equations we find the coefficient of the time derivative is $\beta\gamma$:

$$\boxed{\beta\gamma = \frac{1}{c^2}}$$

This is an absolute requirement.

If we re-work the conservation of electric charge we find

$$\alpha\gamma\frac{\partial\rho}{\partial t}+\delta\nabla\cdot\mathbf{J}=0$$

so *if the unit of charge in* $\mathbf{J}$ *is the same as in* ρ then

$\boxed{\delta=\alpha\gamma}$ and similarly for magnetic charge $\boxed{\lambda=\kappa\beta.}$

Most systems of units assume that $\mathbf{E}$ ($\mathbf{B}$) is the force on a unit electric (magnetic) charge. However, if we re-work Poynting's theorem we find

$$\nabla\cdot(\mathbf{E}\wedge\mathbf{B})+\frac{\gamma}{2}\frac{\partial\mathbf{E}^2}{\partial t}+\frac{\beta}{2}\frac{\partial\mathbf{B}^2}{\partial t}+\delta\mathbf{E}\cdot\mathbf{J}+\lambda\mathbf{B}\cdot\mathbf{J}_m=0$$

so we require the coefficients of $\mathbf{E}\cdot\mathbf{J}$ and $\mathbf{B}\cdot\mathbf{J}_m$ to be the same:

$$\boxed{\delta=\lambda.}$$

This leaves two degrees of freedom:

The value of α

The way that $1/c^2$ is factored between β and γ.

Various considerations have seemed relevant at different times.

Rationalized and Unrationalized Units

If we use M1 to derive Coulomb's law we find

$$\mathbf{F}=\frac{\alpha q_1 q_2}{4\pi r^2}\hat{\mathbf{r}}$$

so we have a choice as to whether the 4π apears in the equation, or in the solution.

Maxwell, following the example of Newton's law of gravitation $F=\dfrac{Gm_1m_2}{r^2}$ chose to put a factor of 4π in α. (But note he did *not* use an analogue of G — see below.)

It was later recognized (Heaviside) that the factor of 4π appears in solutions with spherical symmetry, and a factor of 2π in problems with cylindrical symmetry (like the $\mathbf{B}$ field of a straight wire carrying a current), so it seemed more 'rational' to put the factors of 4π in the solution.

On the whole this idea has found favour, and the modern fashion is for 'rationalized' rather than the older 'unrationalized' units.

Dimensions

In mechanics it is normally assumed (Gauss) that there are three physical dimensions, and thus base units, of mass, length and time.
Units for all other quantities, such as velocity or force, are *derived* from these.
In any physical equation the dimensions must be equal in all terms.

How to fit Electromagnetism into this scheme has caused the most intense debate.
In fact the number of dimensions used in describing a physical situation *is a matter of choice, convenience and personal preference.*

We can choose to increase the number of dimensions:
for example an atmospheric model may have different units for height and distance,
but that will introduce a new physical constant with dimensions height/distance.

We can reduce the number of dimensions, and eliminate physical constants:
for example temperature is a measure of energy —
there is a conversion constant $k = 1.380\,6503(24) \times 10^{-23}\,\mathrm{J\,K^{-1}}$.
We can view the speed of light ($\mathrm{m\,s^{-1}}$), Planck's constant ($\mathrm{J\,Hz^{-1}}$) or the Avogadro constant ($\mathrm{mol^{-1}}$) in the same way.
We could even choose to eliminate mass as a dimension by setting $G = 1$.

Maxwell's Choices

Maxwell used unrationalized units, and in his development of the theory it seemed natural to choose $\beta = 1$, and hence $\gamma = 1/c^2$.
He also took the view that the number of physical dimensions was naturally three:

> In all dynamical sciences it is possible to define . . . units
> in terms of the three fundamental units of Length, Time and Mass.

This led to two possible systems of units:

The electrostatic units, based on forces between electric charges, in which $\alpha = 4\pi$
The electromagnetic units, based on forces between magnetic charges, in which $\kappa = 4\pi$.

The unit sizes are in the ratio of c, (emu larger), and the dimensions differ correspondingly:

$$[Q_{\mathrm{esu}}] = [L^{3/2}][M^{1/2}][T^{-1}] \qquad [Q_{\mathrm{emu}}] = [L^{1/2}][M^{1/2}].$$

Gaussian Units

Maxwell never chose between these systems. A later compromise was to use esu for $\mathbf{E}$, ρ and $\mathbf{J}$ and emu for $\mathbf{B}$, ρ_m and $\mathbf{J}_m$. This is the standard system used in Jackson, and Landau and Lifshitz.

System:	α	β	γ	δ	κ	λ
esu	4π	1	$1/c^2$	$4\pi/c^2$	$4\pi/c^2$	$4\pi/c^2$
emu	$4\pi c^2$	1	$1/c^2$	4π	4π	4π
Gaussian	4π	$1/c$	$1/c$	$4\pi/c$	4π	$4\pi/c$
SI	$\mu_0 c^2$	1	$1/c^2$	μ_0	μ_0	μ_0
Heaviside	1	$1/c$	$1/c$	$1/c$	1	$1/c$

Pros and Cons

Non-obvious and unphysical fractional powers in dimensions in Gaussian units

Unnecessary physical constant in SI units

Both weak arguments, and cancel each other out

Inconvenient unit sizes in Gaussian units

Extraordinarily weak argument!

Electric and magnetic fields have different dimensions in SI units

The only good argument, especially in a relativistic context, as we shall see.

'The CGS system, still used by certain physicists, is absolutely inadequate for an electrical engineer who designs a turbine generator or an electric power transmission system.' (IEC)

'All systems of units are absurd' (Binney)

Vectors under Rotation

The cartesian co-ordinates $x_1,\ x_2,\ x_3$ assigned to a point in space will change when the axes are rotated according to

$$\mathbf{x}' = \mathbf{R}\mathbf{x} \qquad \text{where} \qquad \mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \qquad \mathbf{R} = \begin{pmatrix} R_{11} & R_{12} & R_{13} \\ R_{21} & R_{22} & R_{23} \\ R_{31} & R_{32} & R_{33} \end{pmatrix}$$

$\mathbf{R}$ is a rotation matrix, defined by $\mathbf{R}^T\mathbf{R} = \mathbf{I}$.

It is useful to be able to use suffix notation and the Einstein summation convention:

$$x_i' = \sum_{j=1}^{3} R_{ij}x_j = R_{ij}x_j \qquad \text{with} \qquad R_{ji}R_{jk} = \delta_{ik}. \qquad (\text{Note that } \frac{\partial x_i'}{\partial x_j} = R_{ij}.)$$

This implies $\det \mathbf{R} = \pm 1$;

We shall for the moment consider only *proper rotations,* for which $\det \mathbf{R} = +1$.

Then *vectors under rotation* are defined as quantities which transform in this way, such as velocity $\mathbf{v}$, momentum $\mathbf{p}$ and force $\mathbf{F}$. A vector is also a *tensor of rank one.*

A *tensor of rank zero,* also known as a scalar, is an invariant under rotation, such as mass, charge, energy.

Tensors of higher rank.

Tensors of rank two occur in linear relations between vectors.
For example, in a crystalline material the electrical conductivity will generally depend on the direction of the electric field, and the direction of current flow will not be parallel to the field. The coefficient of the linear relationship is the *conductivity tensor* $\boldsymbol{\sigma}$ with components σ_{ij}:

$$J_i = \sigma_{ij} E_j \qquad \begin{pmatrix} J_1 \\ J_2 \\ J_3 \end{pmatrix} = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{pmatrix} \begin{pmatrix} E_1 \\ E_2 \\ E_3 \end{pmatrix} \qquad \mathbf{J} = \boldsymbol{\sigma}\mathbf{E}$$

and is a tensor of rank two, with $9 = 3^2$ components.

We can deduce the transformation property of $\boldsymbol{\sigma}$ from that of the vectors

$$\mathbf{J}' = \mathbf{R}\mathbf{J} = \mathbf{R}\boldsymbol{\sigma}\mathbf{E} = \mathbf{R}\boldsymbol{\sigma}\mathbf{R}^T\mathbf{R}\mathbf{E} = \boldsymbol{\sigma}'\mathbf{E}'$$

and hence $\boldsymbol{\sigma}' = \mathbf{R}\boldsymbol{\sigma}\mathbf{R}^T$. This is clearer in suffix notation: $\boxed{T'_{i'j'} = R_{i'i} R_{j'j} T_{ij}}$
for any tensor $\mathbf{T}$: each suffix is transformed in an identical way. We can generalize to tensors of arbitrary rank.

Note that tensors of rank higher than two can *only* be handled with suffix notation, as there is no three-dimensional generalization of the matrix.

Unit Tensor

The tensor $\mathbf{I}$ with components $I_{ij} = \delta_{ij}$ thus has the same components in all rotated coordinate systems (an *invariant tensor*):

$$\mathbf{I}' = \mathbf{R}\mathbf{I}\mathbf{R}^T = \mathbf{R}\mathbf{R}^T = \mathbf{I}.$$

Contraction

How do scalar products become scalar?
Consider the rate P at which a force does work on a particle: $P = \mathbf{F} \cdot \mathbf{v}$:

$$P = \mathbf{F} \cdot \mathbf{v} = \begin{pmatrix} F_1 & F_2 & F_3 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} = F_i v_i = \delta_{ij} F_i v_j.$$

Evaluate this in the rotated system, where $\mathbf{F}' = \mathbf{R}\mathbf{F}$ and $\mathbf{v}' = \mathbf{R}\mathbf{v}$:

$$P' = \mathbf{F}' \cdot \mathbf{v}' = \begin{pmatrix} F_1 & F_2 & F_3 \end{pmatrix} \mathbf{R}^T\mathbf{R} \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} = R_{ki} R_{kj} F_i v_j = \delta_{ij} F_i v_j = F_i v_i = \mathbf{F} \cdot \mathbf{v} = P.$$

The process of setting two indices equal and summing over them is called contraction, and reduces the tensor rank by two. Thus $F_i v_i$, $x^2 = x_i x_i$, and σ_{ii} are all scalars.

Vector Products

To write a vector product in suffix notation we need to define the *alternating tensor* or *Levi-Civita density* ϵ_{ijk}:

$$\epsilon_{ijk} = \begin{cases} 1 & \text{if } ijk \text{ is a even permutation of 123} \\ -1 & \text{if } ijk \text{ is an odd permutation of 123} \\ 0 & \text{otherwise} \end{cases}$$

Then we find that $(\mathbf{A} \wedge \mathbf{B})_i = \epsilon_{ijk} A_j B_k$.

Thus, since $\mathbf{A} \wedge \mathbf{B}$ is obtained by contracting vectors with ϵ_{ijk}, it is vector provided ϵ_{ijk} is a third-rank tensor.
We can use the standard rule for a third rank tensor:

$$\epsilon'_{i'j'k'} = R_{i'i} R_{j'j} R_{k'k} \epsilon_{ijk}.$$

This expression is in fact related to $\det \mathbf{R}$; (this follows from the connection between ϵ_{ijk} and permutations, and the definition of a determinant):

$$\epsilon'_{ijk} = \epsilon_{ijk} \det \mathbf{R}$$

and so ϵ_{ijk} is indeed a tensor under *proper* rotations.

Combinations of ϵ_{ijk}

Two ϵ_{ijk} symbols with one index contracted can be expanded:

$$\epsilon_{ijk}\epsilon_{ij'k'} = \delta_{jj'}\delta_{kk'} - \delta_{jk'}\delta_{kj'} \qquad \text{and hence} \qquad \epsilon_{ijk}\epsilon_{ijk'} = 2\delta_{kk'}$$

Scalar and Vector Fields

Applying these ideas to the transformation of fields (i.e. functions of position with scalar or vector or tensor values) introduces two new problems: the transformation of the argument, and differentiation with respect to the argument.

Consider a scalar field such as charge density ρ: the value at a point is unchanged by rotation, but the label of the point has changed:

$$\rho'(\mathbf{x}') = \rho(\mathbf{x}) \qquad \text{where} \qquad \mathbf{x}' = \mathbf{R}\mathbf{x} \qquad \text{or simply} \qquad \rho'(\mathbf{x}') = \rho(\mathbf{R}^T\mathbf{x}').$$

Vector fields work in exactly the same way with respect to their arguments, but in addition we have to rotate the components of the vector field:

$$\mathbf{E}'(\mathbf{x}') = \mathbf{R}\mathbf{E}(\mathbf{R}^T\mathbf{x}').$$

Differentiating Fields

Differentiation with respect to position is a vector operation:
Consider differentiating a scalar field such as ρ:

$$\frac{\partial \rho'}{\partial x'_i} = \left(\frac{\partial x_j}{\partial x'_i}\right) \frac{\partial \rho'}{\partial x_j} = \left(\mathbf{R}^{-1}\right)_{ji} \frac{\partial \rho}{\partial x_j}.$$

But this is just $(\mathbf{R}^{-1})^T \nabla\rho = \mathbf{R}\nabla\rho$.
Thus $\nabla\rho$ is a vector. Similarly differentiating a vector field gives a second rank tensor

$$\frac{\partial E_j}{\partial x_i}.$$

By contraction we can obtain a scalar, the divergence:

$$\nabla \cdot \mathbf{E} = \frac{\partial E_i}{\partial x_i}$$

and using ϵ_{ijk} we can obtain a vector, the curl:

$$(\nabla \wedge \mathbf{E})_i = \epsilon_{ijk} \frac{\partial E_k}{\partial x_j}.$$

Two of Maxwell's equations are thus of the form

$$\text{Scalar} = \text{Scalar:} \qquad \nabla \cdot \mathbf{E} = \rho/\epsilon_0 \qquad \nabla \cdot \mathbf{B} = 0$$

and two are of the form

$$\text{Vector} = \text{Vector:} \qquad \nabla \wedge \mathbf{E} + \frac{\partial \mathbf{B}}{\partial t} = \mathbf{0} \qquad \nabla \wedge \mathbf{B} - \epsilon_0\mu_0 \frac{\partial \mathbf{E}}{\partial t} = \mu_0 \mathbf{J}.$$

(Note that the **0** in M3 is not the same as the 0 in M2: it is the *null vector* not the number zero. It is three zeros.)

Thus the very way we write Maxwell's equations, as vector equations, ensures that they are covariant under proper rotations.

Any rotation $\mathbf{R}$ with $\det \mathbf{R} = -1$ is called an improper rotation. Obviously if we change the sign of all the components of $\mathbf{R}$ we obtain a proper rotation $\mathbf{R}'$ (because $\mathbf{R}$ is a 3×3 matrix and 3 is odd):

$$\mathbf{R} = -\mathbf{R}' = \mathbf{P}\mathbf{R}' \qquad \text{where} \qquad \mathbf{P} = \begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix} = -1 \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Thus any improper rotation $\mathbf{R}$ is a combination of a proper rotation $\mathbf{R}'$ and the *Parity operation* $\mathbf{P}$. The passive interpretation of $\mathbf{P}$ is that it involves the use of a reversed, and therefore left-handed, co-ordinate set.

Our discussion of transformation of vectors, and the scalar nature of dot products, is valid for arbitrary rotations. Under parity $\mathbf{P}$:

Scalars (and even-rank tensors) are *unchanged*: $\rho'(\mathbf{r}') = \rho(\mathbf{P}\mathbf{r}) = \rho(-\mathbf{r})$;

Vectors (and odd-rank tensors) *change sign*: $\mathbf{E}'(\mathbf{r}') = \mathbf{P}\mathbf{E}(\mathbf{P}\mathbf{r}) = -\mathbf{E}(-\mathbf{r})$.
(But note the change in argument, which can undo the sign change!)

But for the alternating tensor this is not so:
ϵ_{ijk} is *defined* to be an invariant tensor ($\epsilon_{123} = 1$ etc) — but under $\mathbf{P}$ it ought to change sign. *Thus it is NOT a tensor under* $\mathbf{P}$. It is known as a *PSEUDOTENSOR*.

Polar and Axial Vectors (True Vectors and PseudoVectors)

Vectors can be classified by their behaviour under $\mathbf{P}$:

POLAR vectors change sign, in the same way that the co-ordinates do;

AXIAL vectors or *PSEUDOVECTORS* do not change sign.

Our ordinary vector notation deos not distinguish them and so we cannot by simple inspection determine whether Maxwell's equations are covariant under $\mathbf{P}$.

We assume that charge and charge density ρ are scalars.

Then $\mathbf{E}$ (force per unit charge) is a polar vector, and M1 is scalar = scalar.

$\nabla \wedge \mathbf{E}$ is then a pseudovector, and hence $\mathbf{B}$ is a pseudovector.

M2 is thus pseudoscalar = zero, M3 is an equality between pseudovectors, and M4 an equality between vectors.

Thus Maxwell's equations are covariant under $\mathbf{P}$, and $\mathbf{B}$ is an axial vector.

Manifest Covariance

Can we write Maxwell's equations so that it is obvious that they are covariant under **P**?

We define the dual tensor $\mathcal{B}_{ij} = \epsilon_{ijk} B_k$:

$$\mathcal{B} = \begin{pmatrix} 0 & B_z & -B_y \\ -B_z & 0 & B_x \\ B_y & -B_x & 0 \end{pmatrix}.$$

Any axial vector can be associated with a true antisymmetric second rank tensor in this way (or a polar vector with an axial second rank tansor).
M4 can then be written

$$\frac{\partial \mathcal{B}_{ij}}{\partial x_j} - \frac{1}{c^2}\frac{\partial E_i}{\partial t} = \mu_0 J_i.$$

M2 and M3 can also be written in a way which involves only true tensors, and so they are also covariant under **P** (see problems).

Thus in this notation the parity transformation is just a special case of a rotation, and all quantities transform in the appropriate way: scalars unchanged, vectors and tensors of odd rank acquire a minus sign, even rank tensors unchanged:
Electromagnetism, like Mechanics, is covariant under **P**.

Maxwell's Equations have a fairly obvious $\mathbf{E} \rightarrow \mathbf{B}$, $\mathbf{B} \rightarrow -\mathbf{E}$ symmetry. This is more explicit if we re-introduce magnetic charge density ρ_m and current density $\mathbf{J}_m$:

$$\nabla \cdot \mathbf{E} = \mu_0 c^2 \rho \qquad \frac{\partial \mathbf{E}}{\partial t} - c^2 \nabla \wedge \mathbf{B} = -\mu_0 c^2 \mathbf{J}$$

$$\nabla \cdot \mathbf{B} = \mu_0 \quad \rho_m \qquad \frac{\partial \mathbf{B}}{\partial t} + \quad \nabla \wedge \mathbf{E} = -\mu_0 \quad \mathbf{J}_m$$

We define the duality transformation $\mathcal{D}$ by

$$\begin{pmatrix} \mathbf{E}' \\ c\mathbf{B}' \end{pmatrix} = \mathcal{D} \begin{pmatrix} \mathbf{E} \\ c\mathbf{B} \end{pmatrix} \qquad \begin{pmatrix} \rho' \\ \rho'_m/c \end{pmatrix} = \mathcal{D} \begin{pmatrix} \rho \\ \rho_m/c \end{pmatrix} \qquad \begin{pmatrix} \mathbf{J}' \\ \mathbf{J}'_m/c \end{pmatrix} = \mathcal{D} \begin{pmatrix} \mathbf{J} \\ \mathbf{J}_m/c \end{pmatrix}$$

where the matrix $\mathcal{D}$ is given by $\mathcal{D} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$.
This is is obviously a rotation matrix in the abstract $(\mathbf{E}, c\mathbf{B})$ space, and this transformation is in fact the 90° example of a continuous transformation. The equations at the top can be written

$$\nabla \cdot \begin{pmatrix} \mathbf{E} \\ c\mathbf{B} \end{pmatrix} = \mu_0 c^2 \begin{pmatrix} \rho \\ \rho_m/c \end{pmatrix} \qquad \frac{\partial}{\partial t} \begin{pmatrix} \mathbf{E} \\ c\mathbf{B} \end{pmatrix} - c\nabla \wedge \mathcal{D} \begin{pmatrix} \mathbf{E} \\ c\mathbf{B} \end{pmatrix} = \mu_0 c^2 \begin{pmatrix} \mathbf{J} \\ \mathbf{J}_m/c \end{pmatrix}$$

which are obviously covariant under any rotation in this abstract space, since the rotation matrix commutes with $\mathcal{D}$.

This transformation has several puzzling features:

> We have just seen that M2 is pseudoscalar, and M3 is pseudovector. Hence *magnetic charge is pseudoscalar.*
>
> Alternatively, magnetic charge is scalar, and the covariance under **P** is lost.

As far as we know ρ_m and $\mathbf{J}_m$ are both zero.

Passive View: this is a matter of convention: the most we can really say is that if we define the electron to be *electrically* charged, then so are all other particles. We could still change the description and make all charges mixed or magnetic.

Active View: the transformation relates distinct situations. So dual solutions may exist in source-free regions, provided we can find sources for them. For example the duals

$$\mathbf{E} = \frac{1}{4\pi\epsilon_0 r^3}(3\mathbf{p}\cdot\hat{\mathbf{r}}\,\hat{\mathbf{r}} - \mathbf{p}) \quad \longrightarrow \quad \mathbf{B}' = \frac{\mu_0}{4\pi r^3}(3\mathbf{m}\cdot\hat{\mathbf{r}}\,\hat{\mathbf{r}} - \mathbf{m})$$

are electric and magnetic dipole fields, both of which have sources. (Note that $\mathbf{p} \to \mathbf{m}/c$.) But there is no source for a magnetic monopole field dual to the field of a point charge.

Note that $\mathcal{D}^2 = -1$, corresponding to reversing the definition of positive/negative charge. This operation is *charge conjugation* $\mathcal{C}$, and Maxwell's equations are covariant under $\mathcal{C}$.

Since $\nabla\cdot\mathbf{B} = 0$ we can define a vector potential $\mathbf{A}$ such that $\mathbf{B} = \nabla\wedge\mathbf{A}$.

In electrostatics we can define an electrostatic potential Φ such that $\mathbf{E} = -\nabla\Phi$.
However this implies $\nabla\wedge\mathbf{E} = 0$, which is not true in general (see M3):

$$\nabla\wedge\mathbf{E} = -\frac{\partial}{\partial t}(\nabla\wedge\mathbf{A}) = -\nabla\wedge\left(\frac{\partial\mathbf{A}}{\partial t}\right) \quad \to \quad \nabla\wedge\left(\mathbf{E} + \frac{\partial\mathbf{A}}{\partial t}\right) = 0$$

This implies that we can always find $\mathbf{A}$ and Φ such that:

$$\boxed{\begin{aligned}\mathbf{B} &= \nabla\wedge\mathbf{A}\\ \mathbf{E} &= -\nabla\Phi - \frac{\partial\mathbf{A}}{\partial t}\end{aligned}}$$

Writing the **E** and **B** fields in this form ensures that the homogeneous Maxwell equations (the ones with no source terms: M2 and M3) are identically satisfied.
The *six* components of **E** and **B** can thus be defined by the *four* components of the potentials **A** and Φ.

A and Φ are not uniquely defined by the requirement that the above differentials give the **E** and **B** fields. Consider the following transformation ($\chi(\mathbf{r},t)$ is any differentiable function):

$$\boxed{\mathbf{A}' = \mathbf{A} + \nabla\chi \qquad \Phi' = \Phi - \frac{\partial\chi}{\partial t}}$$

If we use $\mathbf{A}'$ and Φ' to derive the fields we obtain

$$\mathbf{B}' = \nabla \wedge \mathbf{A} + \nabla \wedge (\nabla\chi) = \mathbf{B} \qquad \text{since curl grad} = 0$$
$$\mathbf{E}' = \mathbf{E} + \nabla\left(\frac{\partial\chi}{\partial t}\right) - \frac{\partial\nabla\chi}{\partial t} = \mathbf{E}.$$

This is an example of a *gauge transformation,* and indicates that the potentials as defined have an unphysical degree of freedom.
For example, it is well-known that we can change the electrostatic potential by a constant: $\Phi' = \Phi + V_0$. (V_0 is normally fixed by requiring $\Phi \to 0$ as $\mathbf{r} \to \infty$.)
But this is just a gauge transformation with $\chi = -V_0 t$.
Similarly $\chi = \chi(\mathbf{r})$ generates a transformation of **A** only; in magnetostatics we normally require $\nabla \cdot \mathbf{A} = 0$ and $\mathbf{A} \to 0$ as $\mathbf{r} \to \infty$.

We are free to specify an extra relation between **A** and Φ; this is called choosing a gauge.
One useful choice (of several) is the Lorenz gauge:

$$\boxed{\nabla \cdot \mathbf{A} + \frac{1}{c^2}\frac{\partial\Phi}{\partial t} = 0.}$$

We can always choose χ to bring **A** and Φ to this form:
Suppose we have potentials **A** and Φ not in the Lorenz gauge, but with

$$\nabla \cdot \mathbf{A} + \frac{1}{c^2}\frac{\partial\Phi}{\partial t} = f(\mathbf{r},t).$$

Now do a gauge transformation with χ:

$$\nabla \cdot \mathbf{A}' + \frac{1}{c^2}\frac{\partial\Phi'}{\partial t} = \nabla^2\chi - \frac{1}{c^2}\frac{\partial^2\chi}{\partial t^2} + f(\mathbf{r},t)$$

so we must choose χ such that

$$\frac{1}{c^2}\frac{\partial^2\chi}{\partial t^2} - \nabla^2\chi = f(\mathbf{r},t)$$

which always has a solution. Indeed this χ is not unique, so that there are still *restricted gauge transformations* within the Lorenz gauge.

The use of the potentials $\mathbf{A}$ and Φ entails that the two homogeneous Maxwell equations are automatically satisfied. The remaining two become:

$$\nabla \cdot \mathbf{E} = \rho/\epsilon_0 \quad \longrightarrow \quad -\nabla \cdot \left(\frac{\partial \mathbf{A}}{\partial t}\right) - \nabla^2 \Phi = \rho/\epsilon_0$$

$$\nabla \wedge \mathbf{B} - \frac{1}{c^2}\frac{\partial \mathbf{E}}{\partial t} = \mu_0 \mathbf{J} \quad \longrightarrow \quad \nabla(\nabla \cdot \mathbf{A}) - \nabla^2 \mathbf{A} + \frac{1}{c^2}\frac{\partial^2 \mathbf{A}}{\partial t^2} + \frac{1}{c^2}\frac{\partial \nabla \Phi}{\partial t} = \mu_0 \mathbf{J}$$

In the Lorenz gauge these uncouple and simplify:

$$\boxed{\begin{aligned} \frac{1}{c^2}\frac{\partial^2 \Phi}{\partial t^2} - \nabla^2 \Phi &= \rho/\epsilon_0 \\ \frac{1}{c^2}\frac{\partial^2 \mathbf{A}}{\partial t^2} - \nabla^2 \mathbf{A} &= \mu_0 \mathbf{J}. \end{aligned}}$$

These are both examples of the inhomogeneous wave equation, and we next look at its solution.

$$\frac{1}{c^2}\frac{\partial^2 g}{\partial t^2} - \nabla^2 g = f$$

The *linearity* of the equation means that the solutions for two different f's can be added together. Thus we can split up $f(\mathbf{r}, t)$ into functions that are non-zero in different spacetime regions. The end-point of this process is a δ-function source:

$$\frac{1}{c^2}\frac{\partial^2 G}{\partial t^2} - \nabla^2 G = \delta(\mathbf{r} - \mathbf{r}')\delta(t - t')$$

which is non-zero only at $\mathbf{r}'$ for an instant at t'. The complete solution is then

$$g(\mathbf{r}, t) = \int G(\mathbf{r}, \mathbf{r}', t, t') f(\mathbf{r}', t')\, d^3\mathbf{r}'\, dt'$$

The function G is called the Green function for the equation.

Compare the solution of Poisson's equation as an integral over the charge distribution:

$$\Phi(\mathbf{r}) = \int \frac{\rho(\mathbf{r}')}{4\pi\epsilon_0 |\mathbf{r} - \mathbf{r}'|} d^3\mathbf{r}'. \qquad \text{The Green function is} \quad -\frac{1}{4\pi|\mathbf{r} - \mathbf{r}'|}.$$

In fact G only depends on co-ordinate differences $G(\mathbf{r}-\mathbf{r}', t-t')$, so without loss of generality we can take $\mathbf{r}' = 0$ and $t' = 0$:

$$\frac{1}{c^2}\frac{\partial^2 G}{\partial t^2} - \nabla^2 G = \delta(\mathbf{r})\delta(t)$$

There are several ways to find a solution to this equation — see

Jackson section 6.6(2nd ed.) or 6.4 (3rd ed.)

Jackson section 12.11

Landau & Lifshitz §62

All of these very different mathematical approaches agree in finding not one Green function but two, which I will call G^+ and G^-. (This is because the Green function gives a particular integral to a differential equation, to which we can add any multiple of the homogeneous equation. The two Green functions differ by a solution of the homogeneous equation.)

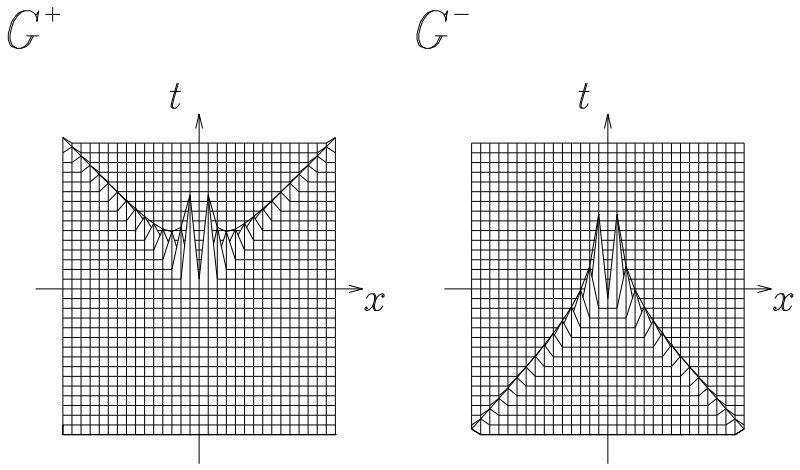
$$G^+(\mathbf{r}, t) = \frac{1}{4\pi r}\delta(t - r/c) \qquad \text{or} \qquad G^+(\mathbf{r}-\mathbf{r}', t-t') = \frac{1}{4\pi|\mathbf{r}-\mathbf{r}'|}\delta(t - t' - |\mathbf{r}-\mathbf{r}'|/c)$$

Correspondingly the other solution is $G^-(\mathbf{r}-\mathbf{r}', t-t') = \dfrac{1}{4\pi|\mathbf{r}-\mathbf{r}'|}\delta(t - t' + |\mathbf{r}-\mathbf{r}'|/c)$

Advanced and Retarded Green Functions

G^+ and G^- give solutions appropriate for different boundary conditions:

G^+ represents an outgoing spherical pulse emitted from $\mathbf{r}'$ at time t', and is known as the *retarded* or *causal* Green function, because the wave is determined by the location and strength of the source at *earlier* times; G^- represents an incoming spherical pulse *absorbed* by the 'source' (or absorber), and is known as the *advanced* or *acausal* Green function, because the wave is determined by the location and strength of the absorber at *future* times.



The G^+ plot shows the source at the spacetime origin radiating a pulse, at speed c, forwards in time, weakening as it expands.

The G^- plot shows an incoming pulse being absorbed by the 'source' at the origin.

Advanced and Retarded Potentials

We can use the Green functions to solve the wave equations.
To preserve causality we choose the retarded Green function G^+. (Note that in so doing we choose a time-asymmetric solution to the time-symmetric Maxwell equations)

$$\Phi(\mathbf{r},t) = \int \frac{\rho(\mathbf{r}',t')}{4\pi\epsilon_0|\mathbf{r}-\mathbf{r}'|}\delta(t-t'-|\mathbf{r}-\mathbf{r}'|/c)\,d^3\mathbf{r}'\,dt'$$

Carrying out the t integral:

$$\Phi(\mathbf{r},t) = \frac{1}{4\pi\epsilon_0}\int \frac{[\rho(\mathbf{r}',t')]_{t'=t-|\mathbf{r}-\mathbf{r}'|/c}}{|\mathbf{r}-\mathbf{r}'|}\,d^3\mathbf{r}'$$

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

These are known as the retarded potentials.
These are the formal solutions to a class of electromagnetic problems in which the source distributions of ρ and $\mathbf{J}$ are specified. (However these problems are usually incompletely specified, as they neglect the effect of the radiated fields on the source distributions.)

Liénard-Wiechert Potentials

For a point charge q in arbitrary motion we can obtain an explicit form for the potentials.
The result is not quite what you might expect . . .

At time t the charge is at $\mathbf{r}_q(t)$; thus we have

$$\rho(\mathbf{r}',t') = q\delta(\mathbf{r}'-\mathbf{r}_q(t')) \qquad \text{and} \qquad \mathbf{J}(\mathbf{r}',t') = q\mathbf{v}\delta(\mathbf{r}'-\mathbf{r}_q(t')).$$

Putting ρ into the integral with G^+ we obtain

$$\Phi(\mathbf{r},t) = \frac{q}{4\pi\epsilon_0}\int \frac{\delta(\mathbf{r}'-\mathbf{r}_q(t'))\,\delta\left(t'-t+\frac{|\mathbf{r}-\mathbf{r}'|}{c}\right)}{|\mathbf{r}-\mathbf{r}'|}\,d^3\mathbf{r}'\,dt'$$

We do the integrals in the reverse order as compared with the derivation of the retarded potentials — we do the spatial integrals first. This just has the effect of replacing $\mathbf{r}'$ with $\mathbf{r}_q(t')$:

$$\Phi(\mathbf{r},t) = \frac{q}{4\pi\epsilon_0}\int \frac{\delta\left(t'-t+\frac{|\mathbf{r}-\mathbf{r}_q(t')|}{c}\right)}{|\mathbf{r}-\mathbf{r}_q(t')|}\,dt'$$

The t' integral is of the form

$$\int h(t')\delta(f(t'))\,dt' = \left.\frac{h(t')}{|df/dt'|}\right|_{f(t')=0}$$

The derivative of the δ-function argument is

$$\frac{d}{dt'}\left(t' - t + \frac{\sqrt{(\mathbf{r}-\mathbf{r}_q(t'))\cdot(\mathbf{r}-\mathbf{r}_q(t'))}}{c}\right) = 1 - \frac{(\mathbf{r}-\mathbf{r}_q(t'))\cdot\mathbf{v}(t')}{c|\mathbf{r}-\mathbf{r}_q(t')|}.$$

and we must evaluate this at the retarded time $t' = t - \frac{|\mathbf{r}-\mathbf{r}_q(t')|}{c}$.

Introducing the unit vector from the retarded position of the charge towards $\mathbf{r}$ $\mathbf{n}(t') = (\mathbf{r}-\mathbf{r}_q(t'))/|\mathbf{r}-\mathbf{r}_q(t')|$ the potentials can be written

$$\Phi(\mathbf{r},t) = \frac{1}{4\pi\epsilon_0}\frac{q}{|\mathbf{r}-\mathbf{r}_q(t')|(1-\mathbf{v}(t')\cdot\mathbf{n}/c)}$$
$$\mathbf{A}(\mathbf{r},t) = \frac{\mu_0}{4\pi}\frac{q\mathbf{v}(t')}{|\mathbf{r}-\mathbf{r}_q(t')|(1-\mathbf{v}(t')\cdot\mathbf{n}/c)}$$

This differs from our expectation by the $(1-\mathbf{v}\cdot\mathbf{n}/c)$ factor, which we can understand in terms of the time taken for the retarded Green function to sweep over the charge.

Feynman gives a very elegant form for the derived fields (*Lectures* I 28-2 and II 21-1 – 21-11).

Dual Sets of Vectors

Consider three linearly independent but ortherwise arbitrary vectors which span a three-dimesional space: $\mathbf{u}_1$, $\mathbf{u}_2$, $\mathbf{u}_3$. We can then write any other vector in terms of the $\mathbf{u}$ basis:

$$\mathbf{A} = A^1\mathbf{u}_1 + A^2\mathbf{u}_2 + A^3\mathbf{u}_3 = A^i\mathbf{u}_i.$$

We now consider a dual basis $\mathbf{d}^1$, $\mathbf{d}^2$, $\mathbf{d}^3$, which is defined by $\boxed{\mathbf{u}_i\cdot\mathbf{d}^j = \delta_{ij}.}$

Explicitly $\mathbf{d}_1 = \mathbf{u}_2\wedge\mathbf{u}_3/(\mathbf{u}_1\cdot\mathbf{u}_2\wedge\mathbf{u}_3)$ and cyclically.
We can also write $\mathbf{A}$ in the $\mathbf{d}$ basis: $\mathbf{A} = A_1\mathbf{d}^1 + A_2\mathbf{d}^2 + A_3\mathbf{d}^3 = A_i\mathbf{d}^i$.
(The A-component index is up in the $\mathbf{u}$-vector basis, and down in the $\mathbf{d}$-vector basis.)

A dot product between vectors expressed in the same basis is messy:

$$\mathbf{A}\cdot\mathbf{B} = A^i\mathbf{u}_i\cdot\mathbf{u}_jB^j = A^ig_{ij}B^j$$

but a dot product between a vectors expressed in the $\mathbf{u}$-basis and the $\mathbf{d}$-basis is easy:

$$\mathbf{A}\cdot\mathbf{B} = A^i\mathbf{u}_i\cdot\mathbf{d}^jB_j = A^iB_i$$

The quantities $g_{ij} = \mathbf{u}_i\cdot\mathbf{u}_j$ are the components of the metric tensor.

Under linear transformations other than rotations the 'invariant tensor' δ_{ij} is either not a tensor, or not invariant, since the condition for invariance

$$\mathbf{R}\,\mathbf{I}\,\mathbf{R}^T = \mathbf{I}$$

specifically defines a rotation.

Both the invariant concept and the tensor are useful, and so both are used:

The Kronecker delta δ_{ij} is defined to be the invariant (1 if $i = j$, 0 if $i \neq j$), and in general is not a tensor.

The tensor, that is the quantity that transforms under more general transformations in a way linked to the co-ordinate transformation, is called the *metric tensor* g_{ij}.

Consider the scalar quantity $D^2 = x^i g_{ij} x^j = \mathbf{x}^T \mathbf{g} \mathbf{x} = \begin{pmatrix} x_1 & x_2 & x_3 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$ where x^i are Cartesian co-ordinates. Clearly this represents the distance-squared from the origin to the point with co-ordinates (x_1, x_2, x_3), and similarly $(x^i - y^i) g_{ij} (x^j - y^j)$ is the distance-squared between the points with co-ordinates (x_1, x_2, x_3) and (y_1, y_2, y_3).

The metric tensor allows co-ordinate separations to be converted to distances, and it coincides with δ_{ij} only in Cartesian co-ordinates. (In general co-ordinate systems g_{ij} will vary from place to place, and it defines distances only between infinitesimally close points. By restricting the co-ordinate transformations to be *linear* we avoid these complications, as g_{ij} does not depend on position.)

Define a linear and invertible transformation of co-ordinates:

$$(x')^i = L^i{}_j x^j \qquad \text{or} \qquad \mathbf{x}' = \mathbf{L}\,\mathbf{x} \qquad \mathbf{x} = \mathbf{L}^{-1}\,\mathbf{x}'$$

Any set of three quantities that transform in the same way as the co-ordinates $(x')^i = L^i{}_j x^j$, are defined to be the components of a *contravariant vector,* and are written with *superscripts.* The basis vectors for these components are obviously the unit co-ordinate displacement vectors, such as from $(0, 0, 0)$ to $(1, 0, 0)$.

(Note that upper and lower indices balance on the two sides of the equation, and that contraction occurs between upper and lower indices.)

But distance is a physical quantity not depending on the co-ordinate system. This allows us to deduce the transformation law for the metric tensor: (compare the discussion of $\boldsymbol{\sigma}$, p.15)

$$D^2 = \mathbf{x}^T\mathbf{g}\mathbf{x} = (\mathbf{x}')^T\mathbf{g}'\mathbf{x}' = \mathbf{x}^T\mathbf{L}^T\mathbf{g}'\mathbf{L}\mathbf{x}$$

and hence

$$\mathbf{g} = \mathbf{L}^T\mathbf{g}'\mathbf{L} \qquad \text{or} \qquad \mathbf{g}' = (\mathbf{L}^{-1})^T\mathbf{g}\mathbf{L}^{-1}$$

which is what you would expect for a tensor except $\mathbf{L} \to (\mathbf{L}^{-1})^T$.

Any set of three quantities that transform as $f_i' = (\mathbf{L}^{-1})^j{}_i f_j$ are defined to be the components of a *covariant vector,* and are written with *subscripts.*
Tensors of higher rank transform in the same way for each index.

Note that for rotations $(\mathbf{R}^{-1})^T = \mathbf{R}$ and so the distinction does not apply.

Some vectors are naturally contravariant, like the co-ordinates.
The metric tensor, as we have seen, is naturally covariant.
So is the gradient of a scalar field Φ (compare the discussion of transformation of ρ p. 19):

$$G_i' = \frac{\partial\Phi}{\partial(x')^i} = \frac{\partial x^j}{\partial(x')^i}\,\frac{\partial\Phi}{\partial x^j} = (L^{-1})^j{}_i G_j.$$

The basis vectors for G_i are vectors orthogonal to surfaces of constant x^i.

Invariants

An invariant is formed by contracting (i.e. setting equal and summing over) a covariant and a contravariant index: consider the change in a scalar function Φ over a small displacement δx^i:

$$\delta\Phi = \frac{\partial\Phi}{\partial x^i}\delta x^i = G_i\,\delta x^i = \text{invariant}$$

Since $\delta\mathbf{x}$ transforms with $\mathbf{L}$ and $\mathbf{G}$ with $(\mathbf{L}^{-1})^T$ the invariance of $\mathbf{G}^T\,\delta\mathbf{x}$ is obvious.
The distance-squared function D^2 provides another example of an invariant formed by contraction in this way.

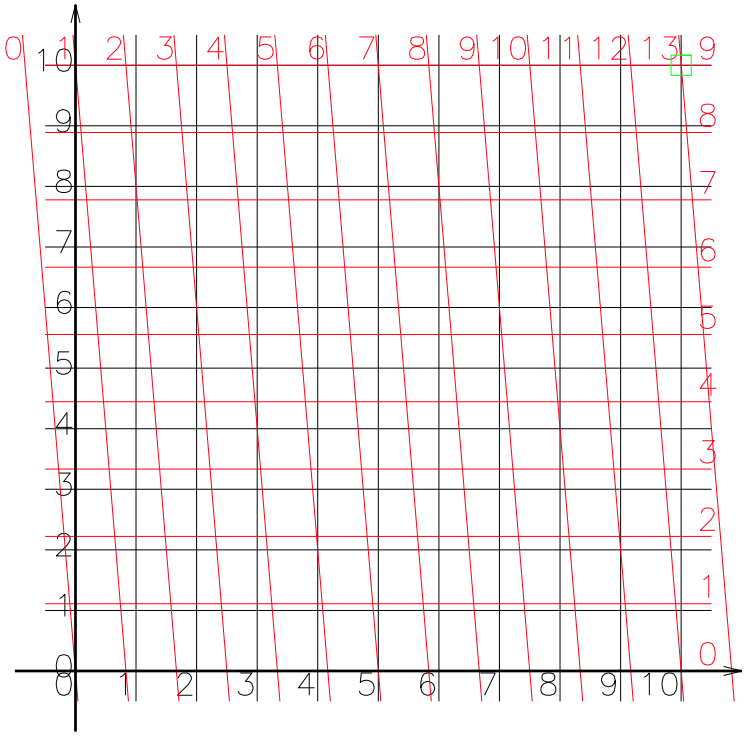
A few minor points:

It is this requirement to contract upper and lower indices that implies that $\mathbf{L}$ should be written with one upper and one lower index.

The use of vector and matrix notation, as above, is strictly ambiguous because it does not indicate the covariant or contravariant nature of the index. The only safe notation is index notation.

The words covariant and contravariant are hallowed by usage but don't really imply anything. In particular, there is no real connection with the other use of covariant.

An Two-Dimensional Example: Oblique Co-ordinates



$$\begin{pmatrix} x_1' \\ x_2' \end{pmatrix} = \begin{pmatrix} \alpha & \gamma \\ 0 & \beta \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

where:

$\alpha = 1.2$

$\beta = 0.9$

$\gamma = 0.1$

so

$$\mathbf{L} = \begin{pmatrix} \alpha & \gamma \\ 0 & \beta \end{pmatrix}$$

and

$$\mathbf{L}^{-1} = \begin{pmatrix} 1/\alpha & -\gamma/\alpha\beta \\ 0 & 1/\beta \end{pmatrix}$$

Black lines: constant x_i.

Red lines: constant x_i'.

Note that the point with $(x_1, x_2) = (10, 10)$ has $(x_1', x_2') = (13, 9)$.

Spacetime Metric

Under a Lorentz Transformation the *interval* between two events s is invariant where

$$s^2 = c^2\Delta t^2 - \Delta x^2 - \Delta y^2 - \Delta z^2$$

and we can use this to make the mathematics of Lorentz transformations look very like three-dimensional linear transformations.

We define the spacetime position four-vector

$$\mathsf{x} = x^\mu = \begin{pmatrix} x^0 \\ x^1 \\ x^2 \\ x^3 \end{pmatrix} = \begin{pmatrix} ct \\ x \\ y \\ z \end{pmatrix} \qquad \text{(Greek indices run } 0 \to 3\text{, Latin run } 1 \to 3\text{).}$$

We define the group of Lorentz transformations Λ as the set of linear transformations on the x^μ that leave invariant the quadratic form

$$s^2 = x^\mu g_{\mu\nu} x^\nu = (x')^\mu g'_{\mu\nu} (x')^\nu \qquad \text{where} \qquad (x')^{\mu'} = \Lambda^{\mu'}{}_{\mu} x^\mu \qquad \text{and} \qquad g'_{\mu\nu} = g_{\mu\nu}$$

which implies that the defining property of Λ is (compare the discussion of metric tensor transformation, p. 38):

$$\Lambda^{\mu'}{}_{\mu} g_{\mu'\nu'} \Lambda^{\nu'}{}_{\nu} = g_{\mu\nu} \qquad \text{or} \qquad \Lambda^T \mathbf{g} \Lambda = \mathbf{g} \qquad \text{or} \qquad \mathbf{g}\Lambda\mathbf{g}^{-1} = \left(\Lambda^{-1}\right)^T$$

Spacetime Metric

$g_{\mu\nu}$ is called the spacetime metric or metric tensor.

$$\mathsf{g} = g_{\mu\nu} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}$$

(Note that we could have written s^2 with the other sign, which would change the sign of g.)

We can now define vectors and tensors just as for general linear transformations:

A *contravariant four-vector* V or V^μ transforms in the same way as x^μ;

A *contravariant second-rank tensor* T or $T^{\mu\nu}$ has 16 components and transforms as $T^{\mu'\nu'} = \Lambda^{\mu'}{}_{\mu}\Lambda^{\nu'}{}_{\nu}T^{\mu\nu}$;

The metric tensor is a *covariant tensor*, and the defining equation for Λ ensures that it is an invariant tensor;

The metric tensor can be used to lower indices: $g_{\mu\nu}x^\nu = x_\mu$ where x_μ are the covariant elements of x, $(ct, -\mathbf{x})$. (The inverse $g^{\mu\nu}$ can be used to raise indices — note it is numerically identical.)

The gradient $\partial/\partial x^\mu$ is naturally covariant, as indicated by the notation $\partial_\mu = \partial/\partial x^\mu$

Space-like and Time-like Intervals

Although we have written the invariant squared interval as s^2, s^2 is not positive-definite, because the metric is indefinite. In fact there are three possibilities:

$s^2 > 0$: Time-like interval, when $c^2\Delta t^2 > \Delta x^2 + \Delta y^2 + \Delta z^2$; there is a reference frame in which events occur *in same place* with time separation $\tau = s/c$ (the *proper time*);

$s^2 < 0$: Space-like interval, when $c^2\Delta t^2 < \Delta x^2 + \Delta y^2 + \Delta z^2$; there is a reference frame in which events occur *at same time* with spatial separation $l = \sqrt{-s^2}$ (the *proper length*);

$s^2 = 0$: Light-like interval, when $c^2\Delta t^2 = \Delta x^2 + \Delta y^2 + \Delta z^2$; the events can be connected by a light wave; there is no special reference frame.

Classes of Lorentz Transformations

Another consequence of the indefinite metric is that we cannot use just the determinant $\det\Lambda$ to classify Lorentz transformations; we need also the sign of $\Lambda^0{}_0$:

$\det\Lambda = 1$ and $\Lambda^0{}_0 > 0$ defines a *proper orthchronous Lorentz transformation;*
The other three classes have spatial inversion, time reversal or both.

Examples of Lorentz transformations

The standard Lorentz transformation from a reference $\mathcal{S}$ to a reference frame $\mathcal{S}'$ moving with speed v along the x-axis is

$$\Lambda = \begin{pmatrix} \gamma & -\gamma\beta & 0 & 0 \\ -\gamma\beta & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

where

$$\beta = \frac{v}{c} \qquad \text{and} \qquad \gamma = \frac{1}{\sqrt{1-\beta^2}}$$

and this satisfies $\Lambda^T \mathsf{g} \Lambda = \mathsf{g}$ as required.

Similarly, if $\mathcal{S}'$ moves along the z-axis:

$$\Lambda = \begin{pmatrix} \gamma & 0 & 0 & -\gamma\beta \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -\gamma\beta & 0 & 0 & \gamma \end{pmatrix}$$

Four-Velocity and Four-Momentum

Consider the world-line of a particle in spacetime, and nearby events on the world-line with co-ordinate separations dx^μ;

a) The velocity is $\dfrac{\mathbf{v}}{c} = \left(\dfrac{dx^1}{dx^0}, \dfrac{dx^2}{dx^0}, \dfrac{dx^3}{dx^0} \right)$ and $v < c$

b) The separation $ds^2 = dx^\mu dx_\mu$ is invariant

c) It follows from a) that the interval between nearby events is time-like, $ds^2 = c^2\, d\tau^2$;

d) By summing the intervals between events we can define an invariant proper time τ for any point on the world-line relative to a chosen event on the world-line as origin;

e) There is thus a covariant desciption of the trajectory $x^\mu(\tau)$;

f) The *four-velocity* is defined as the derivative with respect to τ:

$$u^\mu = \frac{dx^\mu}{d\tau} = (\gamma c,\ \gamma \mathbf{v})$$

and u^μ is a four-vector. (Consider the limiting process in the definition of the derivative: $dx^\mu/d\tau$ = four-vector/invariant.) $u^\mu u_\mu = \mathbf{u} \cdot \mathbf{u} = c^2$.

g) The *four-momentum* is defined by $p^\mu = mu^\mu = (E/c,\ \mathbf{p})$, $\mathbf{p} \cdot \mathbf{p} = m^2c^2$ where m is rest-mass.

Density and Flux

The density and flux of an invariant quantity, like a number, form a four-vector, such as the number density n and flux $\mathbf{f}$ of molecules in a gas (nc, $\mathbf{f}$):

Consider a reference frame in which the flux is zero and the number density is n_0:

$$f^\mu = (n_0 c,\, \mathbf{0}).$$

Consider a reference frame moving with velocity $\mathbf{v}$; the number density increases by a factor of γ because of the Lorentz contraction in the direction of $\mathbf{v}$, $n = \gamma n_0$, whereas the flux is now the rate at which the number density crosses unit area, $\mathbf{f} = -n\mathbf{v} = -\gamma n_0 \mathbf{v}$.

In this reference frame the components are thus $(\gamma n_0 c,\, -\gamma n_0 \mathbf{v})$; but this is just the Lorentz transform of the rest frame f^μ. Thus f^μ is a four-vector.

By a similar argument, the density and flux of a four-vector form a second-rank tensor.

Charge and Current Four-vector

Charge is a Lorentz invariant (see Jackson §11.9 for experimental references). Thus charge density and current density will form a four-vector:

$$\mathbf{j}^\mu = (c\rho,\, \mathbf{J}).$$

We can write the equation of continuity as

$$\frac{\partial \rho}{\partial t} + \nabla \cdot \mathbf{J} = \frac{\partial c\rho}{\partial x^0} + \frac{\partial j^1}{\partial x^1} + \frac{\partial j^2}{\partial x^2} + \frac{\partial j^3}{\partial x^3} = 0$$

which gives an explicitly covariant expression of the conservation of charge:

$$\boxed{\partial_\mu j^\mu = 0.}$$

d'Alembertian

The four-dimensional version of the Laplacian is named after a different Frenchman:

$$\frac{1}{c^2}\frac{\partial^2}{\partial t^2} - \nabla^2 = g^{\mu\nu}\partial_\mu \partial_\nu = \partial_\mu \partial^\mu \qquad \text{where} \qquad \partial^\mu = g^{\mu\nu}\frac{\partial}{\partial x^\nu} = \left(\frac{1}{c}\frac{\partial}{\partial t}, -\nabla\right)$$

The second form makes it clear that this is invariant under a Lorentz transformation.

Potential Four-vector

In the Lorenz gauge, the elements of the current four-vector are the sources in the wave equations for the potentials:

$$\partial_\mu \partial^\mu \Phi = \frac{\rho c}{\epsilon_0 c} \qquad \partial_\mu \partial^\mu \mathbf{A} = \mu_0 \mathbf{J}.$$

If we divide the first by c (remember $\mu_0 \epsilon_0 c^2 = 1$) we obtain the manifestly covariant form

$$\boxed{\partial_\mu \partial^\mu A^\nu = \mu_0 j^\nu}$$

where the potential four-vector A^μ is given by $\boxed{A^\mu = \left(\frac{\Phi}{c},\ \mathbf{A}\right)}$

The Lorenz gauge condition can then be written covariantly: $\boxed{\partial_\mu A^\mu = 0}$

which implies that this is a covariant condition.

Finally we can also write the gauge transformation covariantly: $\boxed{(A')^\mu = A^\mu - \partial^\mu \chi.}$

Field Tensor

Knowing how the potentials transform we can deduce how the fields transform. Consider the equation for **E**:

$$\mathbf{E} = -\nabla\Phi - \frac{\partial \mathbf{A}}{\partial t} = -c\left(\frac{\partial A^0}{\partial x^i} + \frac{\partial A^i}{\partial x^0}\right) = -c\left(\frac{\partial A^i}{\partial x_0} - \frac{\partial A^0}{\partial x_i}\right).$$

Each term looks like the $0i$ or $i0$ component of a tensor, which we can choose to write as upper index, lower index or mixed. Choose the upper index version and define

$$\boxed{F^{\mu\nu} = \partial^\mu A^\nu - \partial^\nu A^\mu} \qquad \text{and then} \qquad F^{0i} = -\frac{E_i}{c}.$$

The elements of F— contravariant: mixed:

$$F^{\mu\nu} = \begin{pmatrix} 0 & -E_x/c & -E_y/c & -E_z/c \\ E_x/c & 0 & -B_z & B_y \\ E_y/c & B_z & 0 & -B_x \\ E_z/c & -B_y & B_x & 0 \end{pmatrix} \quad F^\mu{}_\nu = F^{\mu\lambda} g_{\lambda\nu} = \begin{pmatrix} 0 & E_x/c & E_y/c & E_z/c \\ E_x/c & 0 & B_z & -B_y \\ E_y/c & -B_z & 0 & B_x \\ E_z/c & B_y & -B_x & 0 \end{pmatrix}$$

(Note the occurrence of $-\mathcal{B}$ in the spatial part of F.) N.B. Field invariants.

Covariant Maxwell's Equations

The first Maxwell equation can be written

$$\frac{\partial E_x}{\partial x}+\frac{\partial E_y}{\partial y}+\frac{\partial E_z}{\partial z}=\frac{\rho}{\epsilon_0} \quad \rightarrow \quad \frac{\partial F^{10}}{\partial x^1}+\frac{\partial F^{20}}{\partial x^2}+\frac{\partial F^{30}}{\partial x^3}=\frac{j^0}{\epsilon_0 c^2}$$

while the x-component of the fourth Maxwell equation is

$$\frac{\partial B_z}{\partial y}-\frac{\partial B_y}{\partial z}-\frac{1}{c^2}\frac{\partial E_x}{\partial t}=\mu_0 J_x \quad \rightarrow \quad \partial_2 F^{21}+\partial_3 F^{31}+\partial_0 F^{01}=\mu_0 j^1$$

so we can write both equations as

$$\boxed{\partial_\mu F^{\mu\nu}=\mu_0 j^\nu}$$

Obviously we can write this with other index choices, for example

$$\partial^\mu F_\mu{}^\nu=\mu_0 j^\nu \qquad \partial_\mu F^\mu{}_\nu=\mu_0 j_\nu.$$

Dual Tensor

The other two Maxwell equations can be written in terms of the *dual tensor:*

We define the altenating tensor in the same way as in three dimensions:

$$\epsilon^{\mu\nu\lambda\rho}=\begin{cases}1 & \text{if } \mu\nu\rho\lambda \text{ is a even permutation of 0123}\\ -1 & \text{if } \mu\nu\rho\lambda \text{ is an odd permutation of 0123}\\ 0 & \text{if any index is repeated}\end{cases}$$

This is a pseudotensor: that is a tensor under proper Lorentz transformations with $\det\Lambda=1$, and with the wrong sign under parity or time reversal.

We can use this to define a tensor dual to any antisymmetric tensor:

$$G^{\mu\nu}=\frac{1}{2}\epsilon^{\mu\nu\lambda\rho}F_{\lambda\rho}.$$

which gives in the case of the field tensor

$$G^{\mu\nu}=\begin{pmatrix}0 & -B_x & -B_y & -B_z\\ B_x & 0 & E_z/c & -E_y/c\\ B_y & -E_z/c & 0 & E_x/c\\ B_z & E_y/c & -E_x/c & 0\end{pmatrix}$$

which is just $F^{\mu\nu}$ with $\mathbf{E}/c$ replaced by $\mathbf{B}$, $\mathbf{B}$ by $-\mathbf{E}/c$ — the duality transformation.

Covariant Maxwell's Equations (cont'd)

Using G, and our understanding of the duality transformation, we can write the other two Maxwell equations as

$$\boxed{\partial_\mu G^{\mu\nu} = 0.}$$

We can also get an explicit form in terms of $F^{\mu\nu}$ by contracting with $\epsilon^{\mu\nu\lambda\rho}$:

$$\epsilon^{\mu\nu\lambda\rho}\partial^\chi G_{\chi\rho} = \epsilon^{\mu\nu\lambda\rho}\partial^\chi \epsilon_{\chi\rho\alpha\beta}F^{\alpha\beta}$$

which boils down to (see Landau and Lifshitz, p. 17 footnote):

$$\boxed{\partial^\mu F^{\nu\lambda} + \partial^\nu F^{\lambda\mu} + \partial^\lambda F^{\mu\nu} = 0.}$$

(This has 3 free indices, and so represents 64 equations, but only 4 of these are independent.)

Covariant Lorentz Force

The Lorentz force on a charged particle is a four-vector, and depends on the particle velocity, so the only obvious candidate is

$$qF^{\mu\nu}u_\nu \qquad \text{or equivalently} \qquad qF^\mu{}_\nu u^\nu$$

and this is indeed correct. The spatial part is $\gamma q(\mathbf{E} + \mathbf{v} \wedge \mathbf{B})$ as expected, and the zeroth component is $\gamma \mathbf{E}\cdot\mathbf{v}$

The equation of motion of a charged particle is thus

$$m\frac{du^\mu}{d\tau} = qF^\mu{}_\nu u^\nu.$$

The Lorentz force density is the rate of transfer of four-momentum per unit time per unit volume, and is thus also a four vector:

$$\boxed{f^\mu = F^\mu{}_\nu j^\nu.}$$

Motion of Charged Particle in Uniform Fields

One way to solve the equations of motion is to uncouple them. We can take the z-axis along $\mathbf{B}$, and $\mathbf{E}$ in the xz-plane. The elements of $F^{\mu}{}_{\nu}$ are then

$$\mathsf{F} = F^{\mu}{}_{\nu} = \begin{pmatrix} 0 & E_x/c & 0 & E_z/c \\ E_x/c & 0 & B & 0 \\ 0 & -B & 0 & 0 \\ E_z/c & 0 & 0 & 0 \end{pmatrix}$$

This is *not* symmetric (unless $B = 0$) but still has left eigenvectors $\mathbf{c}^T\mathsf{F} = \lambda\mathbf{c}^T$ and right eigenvectors $\mathsf{F}\mathbf{d} = \lambda\mathbf{d}$. The usual argument for orthogonality of eigenvectors applies between left and right eigenvectors: $\mathbf{c}_m^T\mathbf{d}_n = \delta_{mn}$.
If we dot the equation of motion with $\mathbf{c}_n$ we obtain

$$\frac{d\mathbf{c}_n^T\mathbf{u}}{d\tau} = \omega_n\mathbf{c}_n^T\mathbf{u} \qquad \text{where} \qquad \omega_n = \frac{q\lambda_n}{m}.$$

Hence $\mathbf{c}_n^T\mathbf{u} = U_n\exp(\omega_n\tau)$, and the initial condition is $U_n = \mathbf{c}_n^T\mathbf{u}(0)$ so

$$\mathbf{u}(\tau) = \sum_n \left(\mathbf{c}_n^T\mathbf{u}(0)\right)\exp(\omega_n\tau)\mathbf{d}_n.$$

Conservation of Four-momentum

Knowing the rate of transfer of four-momentum to matter we can re-work the conservation law. We first construct the correct right-hand side:

$$\frac{1}{\mu_0}F^{\lambda}{}_{\nu}\partial_\mu F^{\mu\nu} = F^{\lambda}{}_{\nu}j^{\nu} \qquad \left[= \frac{1}{\mu_0}F^{\lambda\nu}\partial^\mu F_{\mu\nu} \quad \text{on LHS.}\right] \tag{1}$$

We can complete the derivative from the other field equation. First contract it with $F_{\mu\nu}$:

$$F_{\mu\nu}\left(\partial^\mu F^{\lambda\nu} + \partial^\nu F^{\mu\lambda} + \partial^\lambda F^{\nu\mu}\right) = 0$$

The second term is equal to the first (first swap dummy $\mu \Leftrightarrow \nu$, then transpose both F):

$$F_{\mu\nu}\partial^\nu F^{\mu\lambda} = F_{\nu\mu}\partial^\mu F^{\nu\lambda} = F_{\mu\nu}\partial^\mu F^{\lambda\nu}$$

whereas we can re-write the final term using $D = F^{\mu\nu}F_{\mu\nu}$:

$$2F_{\mu\nu}\partial^\mu F^{\lambda\nu} + \frac{1}{2}\partial^\lambda(F_{\mu\nu}F^{\nu\mu}) = 2F_{\mu\nu}\partial^\mu F^{\lambda\nu} - \frac{1}{2}\partial^\lambda D = 0. \tag{2}$$

Adding equation (1) to $1/2\mu_0 \times$ (2), collecting terms on right and lowering the μ index we finally get

$$\boxed{\frac{1}{\mu_0}\partial_\mu\left(g^{\mu\rho}F_{\rho\nu}F^{\nu\lambda} + \frac{1}{4}g^{\mu\lambda}D\right) + F^{\lambda}{}_{\nu}j^{\nu} = 0}$$

Global Phase Invariance in Quantum Mechanics

It is well known that the *overall phase* of the wavefunction is arbitrary:

(i) If Ψ satisfies the Schrödinger equation then so does $\Psi' = e^{i\alpha}\Psi$ because of the linearity of the equation;

(ii) Ψ' is still normalised;

(iii) All physical interpretation is unchanged — for example, the probability density $\Psi\Psi^\star$ and the probability current

$$\mathbf{j} = \frac{-i\hbar}{2m}\left(\Psi^\star\nabla\Psi - \Psi\nabla\Psi^\star\right)$$

are unchanged;

(iv) Similarly if we make a measurement of an observable A, with eigenvalues A_n and eigenfunctions ϕ_n, then the expansion coefficients

$$\Psi = \sum c_n\phi_n \quad \rightarrow \quad c_n = \int \phi_n^\star \Psi \, d^3\mathbf{x}$$

acquire a phase $e^{i\alpha}$, but the probability of obtaining eigenvalue A_n is still $|c_n|^2$.

This is known as *Global Phase Invariance.*

Newton's laws of motion depend on the *forces* $\mathbf{F}$, and we are used to thinking of these as in some way primary, whereas the *potential* V from it which it is derived ($\mathbf{F} = -\nabla V$) is not.

The fact that we can alter the potential by a constant $V' = V + V_0$ without altering $\mathbf{F}$ 'confirms' this viewpoint.

Similar considerations apply to gauge transformations of the EM potentials, which have often been described as mathematical constructions as against the physical reality of the fields.

Ultimately quantum mechanics does not allow us to take this view.

It is the potential V that appears in Schrodinger's equation:

$$\frac{1}{2m}\mathbf{p}^2\Psi + V\Psi = i\hbar\frac{\partial\Psi}{\partial t}$$

(where $\mathbf{p} = -i\hbar\nabla$), or, for a charged particle q subject to a scalar potential Φ

$$\frac{1}{2m}\mathbf{p}^2\Psi + q\Phi\Psi = i\hbar\frac{\partial\Psi}{\partial t}.$$

Simple Gauge Transformation

Suppose Φ is time-independent; then we can solve by separation of variables:

$$\Psi = \sum_n c_n u_n \exp(-iE_n t/\hbar) \qquad \text{where} \qquad \frac{1}{2m}\mathbf{p}^2 u_n + q\Phi u_n = E_n u_n.$$

If we now change Φ by a constant, $\Phi' = \Phi + V_0$ then the time-independent equation changes:

$$\frac{1}{2m}\mathbf{p}^2 u'_n + q\Phi' u'_n = E'_n u'_n \quad \rightarrow \quad \frac{1}{2m}\mathbf{p}^2 u'_n + q\Phi u'_n = (E'_n - qV_0)u'_n.$$

Thus the eigenfunction has not changed: $u'_n = u_n$, but $E'_n = E_n + qV_0$ and hence

$$\Psi' = \sum_n c_n u_n \exp(-iE'_n t/\hbar) = \exp(-iqV_0 t/\hbar)\Psi.$$

The wavefunction has changed by a *time-dependent* global phase.
The interpretation is entirely classical: the energy has changed by the added energy qV_0, but the dynamics is unaltered. The gauge transformation has had no physical effect.

Vector Potential in Quantum Mechanics

The Schrödinger equation for a charged particle q subject to EM potentials $\mathbf{A}$ and Φ is

$$\frac{1}{2m}(\mathbf{p} - q\mathbf{A})^2\Psi + q\Phi\Psi = i\hbar\frac{\partial\Psi}{\partial t} \qquad \text{or} \qquad H\Psi = i\hbar\frac{\partial\Psi}{\partial t}$$

where the Hamiltonian is $H = \frac{1}{2m}\boldsymbol{\pi}^2 + q\Phi$.
$\boldsymbol{\pi} = \mathbf{p} - q\mathbf{A}(\mathbf{r})$ is known as the *kinetic momentum* whereas $\mathbf{p}$ is the *canonical momentum.*

There is a classical argument that makes this Hamiltonian obvious, but we shall look at the implied dynamics from a quantum perspective in order to make it appear plausible.

Lemma: Consider the rate of change of an expectation value $\langle Q\rangle$

$$\langle Q\rangle = \int \Psi^\star \hat{Q}\Psi\, d^3\mathbf{x} \quad \rightarrow \quad \frac{d\langle Q\rangle}{dt} = \int \frac{\partial\Psi^\star}{\partial t}\hat{Q}\Psi\, d^3\mathbf{x} + \int \Psi^\star\hat{Q}\frac{\partial\Psi}{\partial t}\, d^3\mathbf{x} + \int \Psi^\star\frac{\partial\hat{Q}}{\partial t}\Psi\, d^3\mathbf{x}.$$

We can substitute from the Schrodinger equation for the Ψ-derivatives —
$\frac{\partial\Psi}{\partial t} = \frac{1}{i\hbar}H\Psi$, $\frac{\partial\Psi^\star}{\partial t} = \frac{-1}{i\hbar}(H\Psi)^\star$ — and then use the Hermitian property of H:

$$\boxed{\frac{d\langle Q\rangle}{dt} = \frac{1}{i\hbar}\int \Psi^\star[\hat{Q}, H]\Psi\, d^3\mathbf{x} + \langle\frac{\partial\hat{Q}}{\partial t}\rangle.}$$

We use this result repeatedly to find $d\langle\mathbf{x}\rangle/dt$ and then $d^2\langle\mathbf{x}\rangle/dt^2$.
A few useful commutators:

$$[Q,P^2]=[Q,P]P+P[Q,P] \qquad [\boldsymbol{\pi},f(\mathbf{x})]=[\mathbf{p},f(\mathbf{x})]$$
$$[x_j,\pi_k]=[x_j,p_k]=i\hbar\delta_{jk} \qquad [\mathbf{p},f(\mathbf{x})]=-i\hbar\nabla f$$
$$[\pi_i,\pi_j]=-q[p_i,A_j(\mathbf{r})]-q[A_i(\mathbf{r}),p_j]=i\hbar q\left(\frac{\partial A_j}{\partial x_i}-\frac{\partial A_i}{\partial x_j}\right)=i\hbar q\epsilon_{ijk}B_k$$

We first find that $\dfrac{d\langle\mathbf{x}\rangle}{dt}=\dfrac{1}{m}\langle\boldsymbol{\pi}\rangle$. This gives the interpretation for the kinetic momentum: it is just $m\mathbf{v}$.

We can now apply the same procedure to $\dfrac{d\langle\boldsymbol{\pi}\rangle}{dt}$:

$$m\frac{d^2\langle\mathbf{x}\rangle}{dt^2}=\frac{d\langle\boldsymbol{\pi}\rangle}{dt}=\frac{1}{i\hbar}\left(\frac{1}{2m}\langle[\pi_i,\pi_j]\pi_j\rangle+\frac{1}{2m}\langle\pi_j[\pi_i,\pi_j]\rangle+q\langle[\mathbf{p},\Phi]\rangle\right)-q\langle\frac{\partial\mathbf{A}}{\partial t}\rangle$$
$$=q\left(\langle\mathbf{E}\rangle+\frac{1}{2m}\langle\boldsymbol{\pi}\wedge\mathbf{B}-\mathbf{B}\wedge\boldsymbol{\pi}\rangle\right)$$

The last term is a Hermitian operator product: π_iB_j is not Hermitian, $(\pi_iB_j+B_j\pi_i)/2$ is.

How could this be covariant under a gauge transformation of $\mathbf{A}$ and Φ:

$$\mathbf{A}'=\mathbf{A}+\nabla\chi \qquad \Phi'=\Phi-\frac{\partial\chi}{\partial t}$$

The clue is given by the simple case above: $\chi(t)=-V_0t \quad\rightarrow\quad \Psi'=\Psi\exp(-iqV_0t/\hbar)$.
This works because $\left(i\hbar\dfrac{\partial\Psi}{\partial t}-q\Phi\,\Psi\right)$ is covariant under this gauge transformation: both terms acquire extra pieces, which then cancel, so the whole expression just acquires $e^{-iqV_0t/\hbar}$.

Considering instead an arbitrary gauge transformation we try

$$\boxed{\Psi'=\Psi\exp(iq\chi/\hbar)}$$

and we find that

$$\boldsymbol{\pi}\Psi=(\mathbf{p}-q\mathbf{A})\Psi=-i\hbar\nabla\Psi-q\mathbf{A}\,\Psi$$

is *also* covariant, each term acquiring an extra piece that cancels, and hence

$$\boxed{\frac{1}{2m}\Big(\mathbf{p}-q\mathbf{A}\Big)^2\Psi=\Big(i\hbar\frac{\partial\Psi}{\partial t}-q\Phi\,\Psi\Big)}$$

is covariant under gauge transformations of $\mathbf{A}$, Φ and Ψ.

Simple Consequences

The possibility of position-dependent gauge transformations breaks the link between the spatial gradient of the wavefunction and the kinetic momentum $m\mathbf{v}$ of the particle — as we have seen canonical momentum $\mathbf{p} = -i\hbar\nabla$ no longer has this interpretation.

However the probability density is still $P = \Psi^\star\Psi$.

Putting these together we find that the expression for the probability current has to change:

$$\mathbf{j} \neq \frac{-i\hbar}{2m}\Big(\Psi^\star\nabla\Psi - \Psi\nabla\Psi^\star\Big) = \frac{1}{2m}\Big(\Psi^\star\mathbf{p}\Psi - \Psi\,\mathbf{p}\Psi^\star\Big) = \frac{1}{2m}\Big(\Psi^\star\mathbf{p}\Psi + \Psi(\mathbf{p}\Psi)^\star\Big)$$

The obvious thing to try, since the correct $\mathbf{j}$ has to reduce to this when $\mathbf{A} = 0$, is to replace $\mathbf{p}$ with $\boldsymbol{\pi}$:

$$\mathbf{j} = \frac{1}{2m}\Big(\Psi^\star\boldsymbol{\pi}\Psi + \Psi(\boldsymbol{\pi}\Psi)^\star\Big) = \frac{1}{2m}\Big(\Psi^\star\mathbf{p}\Psi - \Psi\mathbf{p}\Psi^\star\Big) - \frac{q\mathbf{A}}{m}\Psi^\star\Psi.$$

The first form shows that this is a gauge-invariant current, so we just have to show that it satisfies the right conservation equation:

$$\nabla\cdot\mathbf{j} + \frac{\partial P}{\partial t} = 0$$

Conservation of Probability

$$\frac{\partial P}{\partial t} = \frac{1}{i\hbar}\Big(\Psi^\star\,H\Psi - \Psi(H\Psi^\star)\Big) = \frac{1}{i2m\hbar}\Big(\Psi^\star\,\boldsymbol{\pi}^2\Psi - \Psi(\boldsymbol{\pi}^2\Psi^\star)\Big)$$

whereas the other term is

$$\nabla\cdot\mathbf{j} = \frac{-1}{i2m\hbar}\mathbf{p}\cdot\Big(\Psi^\star\boldsymbol{\pi}\Psi + \Psi(\boldsymbol{\pi}\Psi)^\star\Big) = \frac{-1}{i2m\hbar}\Big(\Psi^\star\mathbf{p}\cdot\boldsymbol{\pi}\Psi - (\mathbf{p}\Psi)^\star\cdot\boldsymbol{\pi}\Psi + (\mathbf{p}\Psi)\cdot(\boldsymbol{\pi}\Psi)^\star - \Psi(\mathbf{p}\cdot\boldsymbol{\pi}\Psi)^\star\Big).$$

The $\nabla\Psi\cdot\nabla\Psi^\star$ terms cancel to leave

$$\nabla\cdot\mathbf{j} = \frac{-1}{i2m\hbar}\Big(\Psi^\star\mathbf{p}\cdot\boldsymbol{\pi}\Psi + \Psi(q\mathbf{A}\cdot\mathbf{p}\Psi)^\star - \Psi^\star(q\mathbf{A}\cdot\mathbf{p}\Psi) - \Psi(\mathbf{p}\cdot\boldsymbol{\pi}\Psi)^\star\Big).$$

Interchange second and third terms, and add and subtract $\Psi^\star q^2\mathbf{A}^2\Psi$:

$$\nabla\cdot\mathbf{j} = \frac{-1}{i2m\hbar}\Big(\Psi^\star(\mathbf{p}\cdot\boldsymbol{\pi}\Psi - q\mathbf{A}\cdot\mathbf{p}\Psi + q^2\mathbf{A}^2\Psi) - \Psi(\mathbf{p}\cdot\boldsymbol{\pi}\Psi - q\mathbf{A}\cdot\mathbf{p}\Psi + q^2\mathbf{A}^2\Psi)^\star\Big)$$

which cancels the $\partial P/\partial t$ term.

Alternative form for j

If we write the wavefunction Ψ in modulus-argument form, $\Psi = R\exp(iS/\hbar)$ where R and S are real, then we get an alternative form for the current. Note that

$$\mathbf{p}\Psi = -i\hbar\,(\nabla R)\exp(iS/\hbar) - i\hbar\,R(i\nabla S/\hbar)\exp(iS/\hbar) = \left(-i\hbar\frac{\nabla R}{R} + \nabla S\right)\Psi.$$

If we put this into the original (non-invariant) form for $\mathbf{j}$ we find

$$\mathbf{j} = \frac{1}{2m}\Big(\Psi^\star\mathbf{p}\Psi + \Psi(\mathbf{p}\Psi)^\star\Big) = P\frac{\nabla S}{m}$$

which gives the usual interpretation of the wavefunction: the modulus tells us where the particle is ($P = R^2$) and the gradient of the phase tells us where it is going ($m\mathbf{v} = \nabla S$).

The same argument with the gauge-invariant current gives

$$\mathbf{j} = \frac{1}{2m}\Big(\Psi^\star\mathbf{p}\Psi - \Psi\mathbf{p}\Psi^\star\Big) - \frac{q\mathbf{A}}{m}\Psi^\star\Psi = P\left(\frac{\nabla S - q\mathbf{A}}{m}\right)$$

which is obviously gauge invariant since the gauge transformation is just

$$S' = S + q\chi \qquad\qquad \mathbf{A}' = \mathbf{A} + \nabla\chi.$$

and emphasizes that $\mathbf{A}$ changes the link between phase gradient and motion.

Reversing the Argument

- There is a *global* phase invariance $\Psi' = e^{i\alpha}\Psi$
- We postulate a *local* phase invariance $\Psi' = \exp(iq\chi(\mathbf{r},t)/\hbar)\Psi$.
- This requires the existence of potentials $\mathbf{A}(\mathbf{r},t)$ which transform as $\mathbf{A}' = \mathbf{A} + \nabla\chi$ and $\Phi' = \Phi - \partial\chi/\partial t$ which couple to the 'charge' q
- The potentials must appear with derivatives of Ψ in gauge-invariant combinations $(\mathbf{p} - q\mathbf{A})\Psi$ and $(\hat{E} - q\Phi)\Psi$.
- This turns out to imply that the gauge-invariant current (multiplied by the charge q) is the source for gauge-invariant fields derived from the potentials.

The starting point, turning a global symmetry into a local one, is *not* obvious.

Nonetheless this turns out to be an extraordinarily fruitful line of thought: everything is determined by the symmetry group, in this case the one-dimensional unitary group U(1).

Not only does this work once, but *three times!!* — the Standard Model of particle physics consists of three lots of gauge fields with three different symmetry groups, U(1), SU(2), and SU(3) operating on different internal symmetries of the wavefunction.

Atomic Hamiltonian for Interaction with Radiation

The simple Hamiltonian for hydrogen (infinite nuclear mass, no magnetic terms) is of the above form:

$$H = \frac{\mathbf{p}^2}{2m} - \frac{e^2}{4\pi\epsilon_0 r} \quad \text{or} \quad q = -e \quad \mathbf{A} = 0 \quad \Phi_{\text{nuc}} = \frac{e}{4\pi\epsilon_0 r}.$$

Now consider the effect of a plane (and plane-polarised) EM wave:

$$\mathbf{A} = \mathcal{A}\cos(\mathbf{k}\cdot\mathbf{r} - \omega t) \quad \rightarrow \quad \mathbf{E} = -\omega\mathcal{A}\sin(\mathbf{k}\cdot\mathbf{r} - \omega t) \quad \mathbf{B} = -\mathbf{k}\wedge\mathcal{A}\sin(\mathbf{k}\cdot\mathbf{r} - \omega t)$$

where $\mathbf{k}\cdot\mathcal{A} = 0$, and we have chosen a gauge where $\Phi = 0$.

But the atom is tiny compared with the wavelength: $\mathbf{k}\cdot\mathbf{r} \ll 1$ (Electric Dipole approximation). Thus we can approximate the potential as

$$\mathbf{A} = \mathcal{A}\cos(\omega t) \quad \rightarrow \quad \mathbf{E} = \omega\mathcal{A}\sin(\omega t) \quad \mathbf{B} = 0$$

and then the EM wave produces extra terms in the Hamiltonian

$$H_{\text{rad}} = \frac{e}{m}\mathbf{A}\cdot\mathbf{p} + \frac{e^2}{2m}\mathbf{A}^2 \quad = \frac{e}{m}\mathcal{A}\cdot\mathbf{p}\cos(\omega t) + \frac{e^2}{2m}\mathcal{A}^2\cos^2(\omega t)$$

But we could represent this wave with a different set of potentials.
If we do a gauge transformation with $\chi = -\mathcal{A}\cdot\mathbf{r}\cos(\omega t)$ we obtain the *same* approximate fields using the potentials

$$\mathbf{A}' = \mathbf{A} + \nabla\chi = 0 \qquad \Phi' = -\frac{\partial\chi}{\partial t} = -\omega\mathcal{A}\cdot\mathbf{r}\,\sin(\omega t) = -\mathbf{E}\cdot\mathbf{r}$$

With these potentials the interaction Hamiltonian takes the form

$$H_{\text{rad}} = q\Phi' = e\mathbf{E}\cdot\mathbf{r} = -\mathbf{E}\cdot\mathbf{D}$$

where $\mathbf{D}$ is the electric dipole moment operator $-e\mathbf{r}$.
These alternative forms of H_{rad} (known as the *velocity form* and the *length form*) must lead to identical consequences, since they are related by a gauge transformation.

This approach can be taken order by order in $\mathbf{k}\cdot\mathbf{r}$. For example the next term

$$\mathbf{A} = \mathcal{A}\,\mathbf{k}\cdot\mathbf{r}\,\sin(\omega t) \quad \rightarrow \quad \mathbf{B} = \mathbf{k}\wedge\mathcal{A}\,\sin(\omega t) \quad \mathbf{E} = -\omega\mathcal{A}\,\mathbf{k}\cdot\mathbf{r}\,\cos(\omega t)$$

can be transformed with $\chi = -(1/2)\mathcal{A}\cdot\mathbf{r}\,\mathbf{k}\cdot\mathbf{r}\,\sin(\omega t)$ to

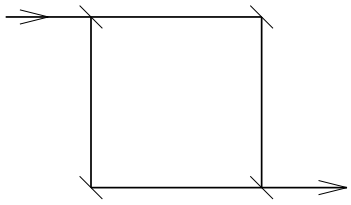
$$\mathbf{A}' = \frac{1}{2}\mathbf{B}\wedge\mathbf{r} \qquad \Phi' = \frac{1}{6}(3r_i r_j)\left(\frac{\partial E_i}{\partial r_j}\right)$$

which splits the term into Magnetic Dipole and Electric Quadrupole Hamiltonians.

It appears that the occurrence of potentials in the Schrödinger equation is consistent with a view of them as mathematical constructs: for example, a gauge transformation of potentials and wavefunction phase leaves the physical interpretation unaltered. It seems that only the fields produce observable effects.

But the phase of Ψ, like the phase of light beams, *can* have an effect when interference can occur between two beams.

This leads us to consider two-beam interference between particle waves. (This experiment has been done with many kinds of 'particles', including quite large molecules.)



At the exit point of the interferometer we have a superposition of two beams:

$$\Psi_1 = A_1 \exp[(\mathbf{P}_1 \cdot \mathbf{x} - Et)/\hbar] \qquad \Psi_2 = A_2 \exp[(\mathbf{P}_2 \cdot \mathbf{x} - Et)/\hbar]$$

($\mathbf{P}_i$ are eigenvalues of $\mathbf{p}$, A_i are complex amplitudes) which will produce a pattern of constructive or destructive interference depending on the relative phase of the two beams (through the A_i) and their co-alignment (through the $\mathbf{P}_i$).

First Question: Can we shift the pattern with a gauge transformation?

This would completely destroy our view of gauge transformations, that they induce a change between *different equivalent* descriptions of the *same* physical situation.

Make the particles charged, and do a gauge transformation with $\chi(\mathbf{r}, t)$.
Each beam coming into the final beam combiner acquires an arbitrary phase shift $q\chi(\mathbf{r}, t)/\hbar$, but when they are superposed at the same point they are both phase-shifted by the *same* amount — hence *no change* in the interference.

We conclude that there are no mysteries hidden in the gauge transformation process.

Second Question: Can we shift the pattern with a vector potential that produces no fields where the beams are?

This would challenge our view of the potentials as just constructs, and it tuns out the answer is *YES!!.*

Infinite solenoid

This is a useful model system with the right properties: an infinite solenoid with radius a and circulating current per unit length j, along the z-axis. The vector potential is (ρ is the cylindrical polar radial co-ordinate, $\rho^2 = x^2 + y^2$)

$$\mathbf{A}(\mathbf{r}) = \begin{cases} \dfrac{\mu_0 j}{2}\,\mathbf{k}\wedge\mathbf{r} & \text{for } \rho < a \\ \dfrac{\mu_0 j a^2}{2\rho^2}\,\mathbf{k}\wedge\mathbf{r} & \text{for } \rho > a. \end{cases}$$

This potential generates magnetic field inside the solenoid:

$$\mathbf{B} = \nabla\wedge\mathbf{A} = \mu_0 j\mathbf{k}$$

But in the exterior region $\mathbf{B} = 0$. However the potential cannot be zero in this region since

$$\oint \mathbf{A}\cdot d\mathbf{l} = \int (\nabla\wedge\mathbf{A})\cdot d\mathbf{S} = (\text{Flux of } \mathbf{B})$$

and this is equal to $\pi a^2\mu_0 j \neq 0$.

An alternative form for the exterior potential is $\mathbf{A} = (\mu_0 j a^2/2\rho)\hat{\boldsymbol{\phi}}$.
This is *locally,* but not globally, removable with a gauge transformation with $\chi = -(\mu_0 j a^2/2)\phi$. It is topologically impossible to remove $\mathbf{A}$ everywhere in this way.
Another way of writing this is $\chi(\mathbf{r}) = -\int_{\mathbf{r}_0}^{\mathbf{r}} \mathbf{A}\cdot d\mathbf{l}$ which is well-defined in a simply-connected region which does not enclose the solenoid. This χ sets $\mathbf{A}$ to zero within this region, but there must be compensating changes in χ elsewhere, since χ is continuous and differentiable.

We can now complete the thought experiment by setting an infinite solenoid in the middle of the interferometer. We can set the vector potential to zero along either path, but not both, with a gauge transformation. Hence the the wavefunctions in the two gauges are related by

$$\Psi_k^{(0)} = \Psi_k \exp(iq\chi_k/\hbar)$$

where χ_k is the above χ taken along a path appropriate for the k'th beam, and $\Psi_k^{(0)}$ denotes the output beam when $\mathbf{A} = 0$.

Finally we see that when the solenoid carries current, the $\mathbf{A}$ field introduces a non-trivial but gauge-independent phase shift between the two beams

$$\frac{q(\chi_2 - \chi_1)}{\hbar} = \frac{q}{\hbar}\oint \mathbf{A}\cdot d\mathbf{l} = \frac{q}{\hbar}\,\text{Flux of } \mathbf{B}.$$