

**Classical Mechanics
Electrodynamics & Relativity**

Dr. J. J. Binney
Michaelmas Term 1993

1 Relativity & Electromagnetism

Observers who move relative to one another do not always agree about the values of quantities, such as speed, mass, energy etc, associated with the same physical system. The special theory of relativity tells us how we may predict the values measured by any observer once we know the values assigned by one particular observer, for example ourselves.

Special relativity teaches us to think of experience as being made up of ‘events’, each with a definite location in the four-dimensional continuum of spacetime. Any given observer assigns to each event a unique 4-tuple of numbers (t, x, y, z) . Of course he can do this in many, many ways. But special relativity claims that there are certain specially favoured systems for assigning coordinates to events, the so-called inertial coordinate systems. O chooses one inertial system and another observer, O' , sets up a different one. But according to special relativity the coordinates (t', x', y', z') O' assigns to any event can be related to O 's coordinates (t, x, y, z) of the same event by

$$\begin{pmatrix} ct' \\ x' \\ y' \\ z' \end{pmatrix} = \begin{pmatrix} ct_0 \\ x_0 \\ y_0 \\ z_0 \end{pmatrix} + \mathbf{L} \cdot \begin{pmatrix} x \\ y \\ z \end{pmatrix}, \quad (1.1)$$

where c is the speed of light and (t_0, x_0, y_0, z_0) is a set of numbers characteristic of the two observers, as is the 4×4 matrix \mathbf{L} .

Clearly, (t_0, x_0, y_0, z_0) are the coordinates O' assigns to the event that marks the origin of O 's coordinates. For simplicity we shall assume that $(t_0, x_0, y_0, z_0) = \mathbf{0}$. In general \mathbf{L} can be represented as the product of matrices generating a rotation, a boost parallel to a coordinate direction and a second rotation: $\mathbf{L} = \mathbf{R}' \cdot \mathbf{L}_0 \cdot \mathbf{R}$, where \mathbf{R} rotates the coordinate axes so as to align the boost direction with a coordinate direction, \mathbf{L}_0 effects the boost along the given axis and \mathbf{R}' rotates the coordinates to any desired final orientation. If \mathbf{R} is chosen such that the x -axis becomes the boost direction, \mathbf{L}_0 has the form

$$\mathbf{L}_0 = \begin{pmatrix} \gamma & -\beta\gamma & 0 & 0 \\ -\beta\gamma & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad \text{where} \quad \begin{aligned} \beta &\equiv v/c \\ \gamma &\equiv 1/\sqrt{1-\beta^2}. \end{aligned} \quad (1.2)$$

For simplicity we confine ourselves to observers whose spatial coordinate systems are aligned, and whose relative motion lies along their (mutually parallel) x -axes. Then in (1.1) $\mathbf{L} = \mathbf{L}_0$ and we get the familiar equations of a Lorentz transformation:

$$\begin{aligned} t' &= \gamma t - \gamma v x / c^2 \\ x' &= \gamma x - \gamma v t \\ y' &= y \\ z' &= z \end{aligned} \quad (1.3)$$

4-vectors Lorentz transformations mix up space and time, so it is useful to define new coordinates which all have dimensions of length. We write $x^0 \equiv ct$, $x^1 \equiv x$, $x^2 \equiv y$, $x^3 \equiv z$, and refer to a general component of the 4-vector (x^0, x^1, x^2, x^3) as x^μ . (The reason for labelling the components with superscripts rather than subscripts will emerge shortly.) Then we write a Lorentz transformation as

$$x^\mu = \Lambda^\mu{}_\nu x^\nu, \quad (1.4a)$$

where

$$\mathbf{\Lambda} \equiv \begin{pmatrix} \gamma & -\beta\gamma & 0 & 0 \\ -\beta\gamma & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \quad (1.4b)$$

In (1.4a) the **Einstein summation convention** is being used in that the summation sign $\sum_{\nu=0}^1$ has been omitted for brevity. You know it's really there because ν appears twice on the right-hand side of the equation, once up and once down.

Why do we write the row index of $\mathbf{\Lambda}$ as a superscript and the column index as a subscript?

A key property of a Lorentz transformation is that $-(ct')^2 + x'^2 + y'^2 + z'^2 = -(ct)^2 + x^2 + y^2 + z^2$. This is analogous to the fact that if two vectors \mathbf{a} and \mathbf{a}' are related by a rotation matrix, then $a_x'^2 + a_y'^2 + a_z'^2 = a_x^2 + a_y^2 + a_z^2$. So a Lorentz transformation is a sort of modified, four-dimensional rotation. When we rotate a vector \mathbf{a} we like to say that the length $|\mathbf{a}|$ is invariant (i.e., stays constant). Analogously we define the length of the 4-vector \mathbf{x} to be

$$|\mathbf{x}| \equiv -(x^0)^2 + (x^1)^2 + (x^2)^2 + (x^3)^2. \quad (1.5)$$

Notes:

- (i) We don't extract a square root because we have no guarantee that $|\mathbf{x}| \geq 0$.
- (ii) 4-vectors that have negative lengths are called **time-like**, while those with positive lengths are **space-like**. Vectors with zero length are said to be **null**.
- (iii) Every book on relativity uses a different convention. The sign of the lengths of space-like vectors is called the "signature of the metric".

The lengths of 4-vectors are sufficiently important for it to be useful to have a way of writing them that does not involve writing out all the components explicitly. To achieve this we introduce this matrix, called the **Minkowski metric**:

$$\boldsymbol{\eta} \equiv \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \quad (1.6)$$

Then we have

$$|\mathbf{x}| = \mathbf{x} \cdot \boldsymbol{\eta} \cdot \mathbf{x}, \quad (1.7a)$$

or in component form

$$|\mathbf{x}| = x^\mu \eta_{\mu\nu} x^\nu. \quad (1.7b)$$

The Einstein convention is here being used to drop two summation signs. We write both of $\boldsymbol{\eta}$'s indices as subscripts so that each sum is over one up and one down index.

Covariant and contravariant vectors We write the result of matrix multiplication of \mathbf{x} by $\boldsymbol{\eta}$ as

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

We have $x_0 = -x^0 = -ct$, $x_1 = x^1$, $x_2 = x^2$ and $x_3 = x^3$. Thus the length of \mathbf{x} is

$$x^\mu x_\mu = -c^2 t^2 + x^2 + y^2 + z^2.$$

Notice that here as everywhere else, we are summing over one up and one down index. In order to stick rigidly to this rule, we define

$$\eta^{\mu\nu} \equiv \eta_{\mu\nu} \equiv \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \quad (1.8)$$

Note:

We have $\eta^{\mu\gamma} \eta_{\gamma\nu} = \delta_\nu^\mu$, or in matrix form $\boldsymbol{\eta} \cdot \boldsymbol{\eta} = \mathbf{I}$, where \mathbf{I} and δ_ν^μ are two ways of writing the 4×4 identity matrix. Also $\eta^{\mu\nu} = \eta^{\mu\gamma} \delta_\gamma^\nu$, so in a sense $\boldsymbol{\eta}$ is merely the up-up and down-down forms of the identity matrix.

From x_μ we can recover x^μ ;

$$x^\mu = \eta^{\mu\nu} x_\nu. \quad (1.9)$$

x_μ is a 4-vector, but of a slightly different type than x^μ , because under a Lorentz transformation we have

$$\begin{aligned} x'_\mu &= \eta_{\mu\nu} x'^\nu = \eta_{\mu\nu} \Lambda^\nu{}_\kappa x^\kappa = \eta_{\mu\nu} \Lambda^\nu{}_\kappa \eta^{\kappa\lambda} x_\lambda \\ &= \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \gamma & -\beta\gamma & 0 & 0 \\ -\beta\gamma & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_0 \\ x_1 \\ x_2 \\ x_3 \end{pmatrix} \\ &= \begin{pmatrix} \gamma & \beta\gamma & 0 & 0 \\ \beta\gamma & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_0 \\ x_1 \\ x_2 \\ x_3 \end{pmatrix} \equiv \Lambda_\mu{}^\nu x_\nu, \end{aligned} \quad (1.10)$$

where we have defined a new matrix

$$\Lambda_\mu{}^\lambda \equiv \eta_{\mu\nu} \Lambda^\nu{}_\kappa \eta^{\kappa\lambda}. \quad (1.11)$$

Notice that the transpose of $\Lambda_\mu{}^\nu$ is the inverse of $\Lambda^\mu{}_\nu$:

$$\Lambda^\mu{}_\kappa \Lambda_\mu{}^\nu = \delta_\kappa^\nu, \quad (1.12)$$

where we have again written the 4×4 identity matrix as δ_κ^ν .

Exercise (1):

Obtain (1.12) from the requirement that for any two vectors \mathbf{x} , \mathbf{y} , we have $x'_\mu y'^\mu = x_\mu y^\mu$.

Vectors with their indices below are called **covariant** (x_μ). Vectors with indices above are called **contravariant** (x^μ). I shall call them down and up vectors. The operation of setting two indices equal and summing from 0 to 3 is called **contraction**. In a contraction one index must be up and one down. Quantities like $\sum_\mu x_\mu x_\mu$ have nothing to do with physics. An important motivation for writing x^μ rather than \mathbf{x} is to distinguish the up from the down form of \mathbf{x} . Often an expression is equally valid for up or down vectors provided the basic rules are obeyed, and then it is neater to use conventional vector notation than to stick in indices. For example, if \mathbf{a} and \mathbf{b} are vectors and \mathbf{M} is a matrix, we can interpret $\mathbf{a} = \mathbf{M} \cdot \mathbf{b}$ as $a^\mu = M^{\mu\nu} b_\nu$, as $a_\mu = M_{\mu\nu} b^\nu$, or in yet other ways. But if you ever express a 4-vector in component form, you *must* come clean and say whether you're giving the up or the down vector, as in $x^\mu = (ct, x, y, z)$.

According to special relativity, all quantities of physical interest can be grouped into n -tuples.

1.1 1-tuples (4-scalars)

On some things all observers agree, for example the charge and total spin of the an electron. These quantities are called **4-scalars** or relativistic invariants. The length of a 4-vector is a 4-scalar.

1.2 4-tuples (4-vectors)

If O measures the wave-vector and frequency of a photon to be \mathbf{k} and ω , then an observer O' who moves at speed v along O's x -axis measures wave-vector \mathbf{k}' and frequency ω' given by

$$\begin{pmatrix} \omega'/c \\ k'_x \\ k'_y \\ k'_z \end{pmatrix} = \begin{pmatrix} \gamma & -\beta\gamma & 0 & 0 \\ -\beta\gamma & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \omega/c \\ k_x \\ k_y \\ k_z \end{pmatrix}. \quad (1.13a)$$

The matrix form of this equation is

$$\begin{pmatrix} \omega'/c \\ \mathbf{k}' \end{pmatrix} = \mathbf{\Lambda} \cdot \begin{pmatrix} \omega/c \\ \mathbf{k} \end{pmatrix} \quad \text{where} \quad \mathbf{\Lambda} \equiv \begin{pmatrix} \gamma & -\beta\gamma & 0 & 0 \\ -\beta\gamma & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \quad (1.13b)$$

Notes:

- (i) The Lorentz transformation matrix $\mathbf{\Lambda}$ is dimensionless, so ω has to be divided by c to give the same dimensions as \mathbf{k} before being put into the last place of a 4-vector with \mathbf{k} .
- (ii) Vectors written in italic boldface (\mathbf{k}) are 3-vectors, while those written in Roman boldface (\mathbf{k}) are 4-vectors.

If we define $k^0 \equiv \omega/c$, then

$$\mathbf{k}' = \mathbf{\Lambda} \cdot \mathbf{k} \quad \text{i.e.,} \quad k'^{\mu} = \Lambda^{\mu}_{\nu} k^{\nu}. \quad (1.14)$$

Exercise (2):

Determine whether the photon is blue or red shifted between its emission by O and its detection by O'. Relate this to the question of whether O' is approaching or receding from O.

The length of a photon's 4-vector is the scalar

$$|\mathbf{k}| \equiv -(k^0)^2 + (k^1)^2 + (k^2)^2 + (k^3)^2 = -\frac{\omega^2}{c^2} + |\mathbf{k}|^2 = 0.$$

One can prove that this really is a scalar by brute force:

$$\begin{aligned} |\mathbf{k}'| &= -(k'^0)^2 + (k'^1)^2 + (k'^2)^2 + (k'^3)^2 \\ &= -\left(\gamma\frac{\omega}{c} - \beta\gamma k^1\right)^2 + \left(-\beta\gamma\frac{\omega}{c} + \gamma k^1\right)^2 + (k^2)^2 + (k^3)^2 \\ &= -\gamma^2\left(1 - \beta^2\right)\frac{\omega^2}{c^2} + \gamma^2\left(1 - \beta^2\right)(k^1)^2 + (k^2)^2 + (k^3)^2 \\ &= -(k^0)^2 + (k^1)^2 + (k^2)^2 + (k^3)^2. \end{aligned}$$

Another familiar 4-tuple: if observer O measures energy E and momentum \mathbf{p} for some particle, then O' will measure E' and \mathbf{p}' given by

$$\begin{pmatrix} E'/c \\ \mathbf{p}' \end{pmatrix} = \mathbf{\Lambda} \cdot \begin{pmatrix} E/c \\ \mathbf{p} \end{pmatrix}, \quad (1.15)$$

or setting $p^0 \equiv E/c$, we have $p'^{\mu} = \Lambda^{\mu}_{\nu} p^{\nu}$.

The length of the momentum-energy 4-vector of a particle of rest mass $m_0 \neq 0$ is just $-c^2$ times the square of its rest mass m_0 . We show this by arguing that it doesn't matter in whose frame we evaluate a scalar. We choose the particle's rest frame. Then $\mathbf{p} = 0$ and $E = cp^0 = m_0 c^2$, so

$$(p^0)^2 + (p^1)^2 + (p^2)^2 + (p^3)^2 = -m_0^2 c^2.$$

1.3 6-tuples (antisymmetric 2nd rank tensors)

If the electric and magnetic fields measured by O are arranged into the antisymmetric matrix \mathbf{F} ,

$$F^{\mu\nu} \equiv \begin{pmatrix} 0 & E_x/c & E_y/c & E_z/c \\ -E_x/c & 0 & B_z & -B_y \\ -E_y/c & -B_z & 0 & B_x \\ -E_z/c & B_y & -B_x & 0 \end{pmatrix} \quad (\text{SI units}), \quad (1.16)$$

then O' will measure E' and B' as

$$\begin{pmatrix} 0 & E'_x/c & E'_y/c & E'_z/c \\ -E'_x/c & 0 & B'_z & -B'_y \\ -E'_y/c & -B'_z & 0 & B'_x \\ -E'_z/c & B'_y & -B'_x & 0 \end{pmatrix} \equiv F'^{\mu\nu} = \Lambda^\mu{}_\kappa \Lambda^\nu{}_\lambda F^{\kappa\lambda}. \quad (1.17)$$

Note that $F^{\mu\nu}$ transforms *as if* it were the product $p^\mu p^\nu$ of two down-vectors (which it isn't). Objects that transform in this way are called second-rank tensors.

\mathbf{F} is called the **Maxwell field tensor**.

Exercise (3):

Transform $F^{\kappa\lambda}$ with the matrix $\Lambda^\mu{}_\nu$ defined by (1.13b) to show that an observer who moves at speed v down the x -axis of an observer who sees fields $\mathbf{E} = (E_x, E_y, 0)$ and $\mathbf{B} = 0$, perceives fields $\mathbf{E}' = (E_x, \gamma E_y, 0)$ and $\mathbf{B}' = (0, 0, \gamma v E_y/c)$. [Hint: since $\mathbf{\Lambda}$ is symmetric, we can write $\mathbf{F}' = \mathbf{\Lambda} \cdot \mathbf{F} \cdot \mathbf{\Lambda}$.] Hence deduce the general rules $\mathbf{E}'_{\parallel} = \mathbf{E}_{\parallel}$, $\mathbf{E}'_{\perp} = \gamma(\mathbf{E}_{\perp} + \mathbf{v} \times \mathbf{B})$, $\mathbf{B}'_{\parallel} = \mathbf{B}_{\parallel}$, $\mathbf{B}'_{\perp} = \gamma(\mathbf{B}_{\perp} - \mathbf{v} \times \mathbf{E}/c^2)$. Verify that $(B^2 - E^2/c^2) = (B'^2 - E'^2/c^2)$.

Some 6-tuples correspond to elements of area. This correspondence works as follows. With any two displacements, say \mathbf{u} and \mathbf{v} , we associate the parallelogram bounded by \mathbf{u} and \mathbf{v} . Information about the size and orientation of this parallelogram is conveyed by the antisymmetric tensor $S^{\alpha\beta} \equiv u^\alpha v^\beta - u^\beta v^\alpha$; in particular, if $\mathbf{u} = \mathbf{v}$, then $\mathbf{S} = 0$. \mathbf{S} has six degrees of freedom rather than the eight numbers involved in \mathbf{u} and \mathbf{v} because we can add to \mathbf{u} any multiple of \mathbf{v} without affecting \mathbf{S} , and vice versa for \mathbf{v} and \mathbf{u} .

Exercise (4):

Prove the last statement and give a geometrical interpretation of this result.

In three-space the size and orientation of a parallelogram may be specified by giving the magnitude and direction of the normal. Hence in three-space full information about an antisymmetric 2nd rank tensor can be packed into the three components of the 3-vector which we call the cross-product of the parallelogram's sides. In four-dimensional spacetime each parallelogram has a magnitude and two mutually perpendicular normals, requiring six numbers for its full specification. Consequently there is no direct analogue of the cross product and we must represent areas directly with antisymmetric tensors.

Exercise (5):

Relate the above statements to the number of independent components of an antisymmetric $n \times n$ matrix for $n = 2, 3, 4$.

A physically interesting 6-tuple that describes an area is the tensor $(x^\mu p^\nu - x^\nu p^\mu)$ formed from the space-time coordinate vector $x^\mu = (ct, x, y, z)$ and the 4-momentum of a particle. If the angular momentum about the origin is \mathbf{L} , we have

$$H^{\mu\nu} \equiv (x^\mu p^\nu - x^\nu p^\mu) = \begin{pmatrix} 0 & \cdots & \cdots & \cdots \\ c(xE/c^2 - tp_x) & 0 & \cdots & \cdots \\ c(yE/c^2 - tp_y) & -L_z & 0 & \cdots \\ c(zE/c^2 - tp_z) & L_y & -L_x & 0 \end{pmatrix}, \quad (1.18)$$

where the diagonal dots stand for minus the quantities in the lower left triangle of the matrix. The numbers in the first column of this matrix give mc times the particle's initial position vector.

With every 6-tuple we get two free scalars. If the 6-tuple is of the form $(u^\alpha v^\beta - u^\beta v^\alpha)$, then one of these is twice the squared magnitude of the corresponding parallelogram:

$$\begin{aligned} S^{\mu\nu}(\eta_{\mu\kappa}\eta_{\nu\lambda}S^{\kappa\lambda}) &\equiv S^{\mu\nu}S_{\mu\nu} = -\text{Tr } \mathbf{S} \cdot \mathbf{S} \\ &= (u^\mu v^\nu - u^\nu v^\mu)(u_\mu v_\nu - u_\nu v_\mu) = 2[|\mathbf{u}||\mathbf{v}| - (\mathbf{u} \cdot \mathbf{v})^2]. \end{aligned}$$

Evaluation in the particle's rest frame shows that the scalar $\frac{1}{2}H_{\mu\nu}H^{\mu\nu} = [|\mathbf{x}||\mathbf{p}| - (\mathbf{x} \cdot \mathbf{p})^2] = -(m_0 c r_0)^2$, where r_0 is the distance (in the rest frame) between the particle and the origin at $t = 0$.

It is interesting to evaluate this same scalar for the Maxwell field tensor. Straightforward matrix multiplication shows that the down-down shadow of $F^{\mu\nu}$ is¹

$$F_{\mu\nu} \equiv \begin{pmatrix} 0 & -E_x/c & -E_y/c & -E_z/c \\ E_x/c & 0 & B_z & -B_y \\ E_y/c & -B_z & 0 & B_x \\ E_z/c & B_y & -B_x & 0 \end{pmatrix} \quad (\text{SI units}), \quad (1.19)$$

Multiplying each element of $F_{\mu\nu}$ by the corresponding element of $F^{\mu\nu}$ we find

$$\begin{aligned} m &\equiv \frac{1}{2}F_{\mu\nu}F^{\mu\nu} = -\frac{1}{2}\text{Tr } \mathbf{F} \cdot \mathbf{F} \\ &= \frac{1}{2}(\text{each element of } F_{\mu\nu}) \times (\text{corresponding element of } F^{\mu\nu}) \\ &= (B^2 - E^2/c^2). \end{aligned} \quad (1.20)$$

To extract another scalar from a 6-tuple we need to introduce the **Levi-Civita symbol**:

$$\epsilon^{\alpha\beta\gamma\delta} = \begin{cases} +1 & \text{if } \alpha\beta\gamma\delta \text{ is an even permutation of } 0123 \\ -1 & \text{if } \alpha\beta\gamma\delta \text{ is an odd permutation of } 0123 \\ 0 & \text{otherwise.} \end{cases} \quad (1.21)$$

Note:

Whereas when n is odd, the cyclic interchange $i_1 \rightarrow i_2 \rightarrow \dots \rightarrow i_{n-1} \rightarrow i_n \rightarrow i_1$ is an even permutation of the i_k , when n is even, this permutation is odd. (To prove this exchange i_1 and i_n and then make $n-2$ exchanges to work i_1 back to the second place.) So whereas for 3-dimensional tensors $\epsilon_{jki} = \epsilon_{ijk}$, we now have $\epsilon^{\beta\gamma\delta\alpha} = -\epsilon^{\alpha\beta\gamma\delta}$.

$\epsilon^{\alpha\beta\gamma\delta}$ allows us to form the **dual** $\overline{\mathbf{F}}$ of \mathbf{F} :

$$\begin{aligned} \overline{\mathbf{F}}^{\alpha\beta} &\equiv \frac{1}{2}\epsilon^{\alpha\beta\gamma\delta}F_{\gamma\delta} \\ &= \begin{pmatrix} 0 & B_x & B_y & B_z \\ -B_x & 0 & -E_z/c & E_y/c \\ -B_y & E_z/c & 0 & -E_x/c \\ -B_z & -E_y/c & E_x/c & 0 \end{pmatrix}. \end{aligned} \quad (1.22)$$

$\overline{\mathbf{F}}$ can be obtained from \mathbf{F} by the transformation $\mathbf{E} \rightarrow \mathbf{B}$, $\mathbf{B} \rightarrow -\mathbf{E}$. The other scalar is the trace of the product of \mathbf{F} with its dual:

$$\begin{aligned} f &\equiv \text{Tr } \mathbf{F} \cdot \overline{\mathbf{F}} \\ &= -(\text{each element of } F_{\alpha\beta}) \times (\text{corresponding element of } \overline{\mathbf{F}}^{\alpha\beta}) \\ &= \frac{4}{c}\mathbf{E} \cdot \mathbf{B}. \end{aligned} \quad (1.23)$$

¹ It is worth remembering that in special relativity the lowering operation only *changes the sign of the mixed space-time components*.

1.4 10-tuples (symmetric 2nd rank tensors)

Imagine that we move some charges around. Then the rate at which we do work *on* the e.m. field is

$$\begin{aligned}\dot{\mathcal{E}} &= - \int \mathbf{E} \cdot \mathbf{j} \, d^3\mathbf{x} \\ &= - \frac{1}{\mu_0} \int \mathbf{E} \cdot \left(\nabla \times \mathbf{B} - \frac{1}{c^2} \frac{\partial \mathbf{E}}{\partial t} \right) d^3\mathbf{x}\end{aligned}\quad (1.24)$$

But $\nabla \cdot (\mathbf{E} \times \mathbf{B}) = \mathbf{B} \cdot (\nabla \times \mathbf{E}) - \mathbf{E} \cdot (\nabla \times \mathbf{B})$, so (1.24) can be rewritten

$$\begin{aligned}\dot{\mathcal{E}} &= \frac{1}{\mu_0} \int \nabla \cdot (\mathbf{E} \times \mathbf{B}) \, d^3\mathbf{x} + \frac{1}{\mu_0} \int \left(-\mathbf{B} \cdot (\nabla \times \mathbf{E}) + \frac{1}{c^2} \mathbf{E} \cdot \frac{\partial \mathbf{E}}{\partial t} \right) d^3\mathbf{x} \\ &= \frac{1}{\mu_0} \oint (\mathbf{E} \times \mathbf{B}) \cdot d^2\mathbf{S} + \frac{1}{2\mu_0} \int \frac{\partial}{\partial t} (B^2 + E^2/c^2) \, d^3\mathbf{x}.\end{aligned}\quad (1.25)$$

If energy is to be conserved, the energy we deploy moving the charges has to go somewhere. According to (1.25) energy will be conserved if we interpret the **Poynting vector**

$$\mathbf{N} \equiv \frac{1}{\mu_0} \mathbf{E} \times \mathbf{B} \quad (1.26)$$

as the flux of e.m. energy, and

$$\frac{1}{2\mu_0} (B^2 + E^2/c^2) \quad (1.27)$$

as the density of e.m. energy.

How do the Poynting vector and the e.m. energy-density fit into the scheme of n -tuples? From \mathbf{F} we can construct the following important tensor:

$$\begin{aligned}T^{\mu\nu} &= \frac{1}{\mu_0} \left[-\frac{1}{4} (F_{\delta\gamma} F^{\delta\gamma}) \eta^{\mu\nu} - F^\mu{}_\gamma F^{\gamma\nu} \right]; \\ \mathbf{T} &= \frac{1}{\mu_0} \left[\frac{1}{4} \text{Tr}(\mathbf{F} \cdot \mathbf{F}) \boldsymbol{\eta} - \mathbf{F} \cdot \mathbf{F} \right],\end{aligned}\quad (1.28)$$

where \mathbf{F} is, as usual, the Maxwell field tensor (1.16). It's easy to see that $\text{Tr} \mathbf{T} = 0$. A little slog shows that in terms of \mathbf{E} and \mathbf{B} the tensor \mathbf{T} is

$$T^{\mu\nu} = \begin{pmatrix} \frac{1}{2\mu_0} (B^2 + E^2/c^2) & N_x/c & N_y/c & N_z/c \\ N_x/c & & & \\ N_y/c & & P_{ij} & \\ N_z/c & & & \end{pmatrix}, \quad (1.29)$$

where

$$P_{ij} \equiv \frac{1}{\mu_0} \left[\frac{1}{2} \delta_{ij} \left(B^2 + \frac{E^2}{c^2} \right) - \left(B_i B_j + \frac{E_i E_j}{c^2} \right) \right] \quad (i, j = 1, 2, 3). \quad (1.30)$$

Thus the energy density in the e.m. field is the 00 component of \mathbf{T} and the Poynting vector occupies the mixed space-time components of \mathbf{T} . It turns out that the 3×3 matrix P_{ij} describes the flux of the three kinds of momentum: P_{ix} = flux of x -momentum etc.

Exercise (6):

Show that a uniform magnetic field parallel to the z -axis is associated with tension (negative pressure) along the axis, and pressure in the perpendicular directions.

As an example of \mathbf{T} consider a plane e.m. wave running along $\hat{\mathbf{i}}$ polarized parallel to $\hat{\mathbf{j}}$. Then

$$\begin{aligned}\mathbf{E} &= (0, E, 0) \cos(\omega t - kx) \\ \mathbf{B} &= (0, 0, B) \cos(\omega t - kx).\end{aligned}$$

E and B are related by $-\partial \mathbf{B} / \partial t = \nabla \times \mathbf{E} \Rightarrow B = kE / \omega = E / c$. Hence

$$\mathbf{N} = (E^2 / \mu_0 c, 0, 0) \cos^2(\omega t - kx).$$

The first term in our expression (1.30) is non-zero only on the diagonal. The second term is non-zero only in the yy and zz slots and there cancels the first term. So \mathbf{P} is

$$P_{ij} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \frac{E^2}{\mu_0 c^2} \cos^2(\omega t - kx),$$

and finally

$$T^{\mu\nu} = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \frac{E^2}{\mu_0 c^2} \cos^2(\omega t - kx). \quad (1.31)$$

The stress tensor \mathbf{P} has only an entry in the xx slot because our wave is engaged in the business of carrying x -type momentum in the x -direction; the wave would push back a mirror placed in a plane $x = \text{constant}$. Clearly the Poynting vector is also directed along the x axis, which accounts for the off-diagonal units in \mathbf{T} . In proper relativistic units the wave employs unit energy density (“capital employed”) to carry unit fluxes of energy and momentum (“turnover”). Notice that the wave’s phase is the scalar $-\mathbf{k} \cdot \mathbf{x}$.

1.5 Derivatives of tensors

Derivatives with respect to any system of coordinates can be expressed in terms of derivatives w.r.t. any other system by use of the chain rule:

$$\frac{\partial}{\partial x'^{\mu}} = \frac{\partial x^{\nu}}{\partial x'^{\mu}} \frac{\partial}{\partial x^{\nu}}. \quad (1.32)$$

If the primed and unprimed systems are linked by a Lorentz transformation,

$$x'^{\nu} = \Lambda^{\nu}_{\mu} x^{\mu}, \quad (1.33)$$

we have on multiplying by Λ_{ν}^{κ} and summing over ν ,

$$\Lambda_{\nu}^{\kappa} x'^{\nu} = \Lambda_{\nu}^{\kappa} \Lambda^{\nu}_{\mu} x^{\mu} = x^{\kappa},$$

where the last step follows by (1.12). Differentiating we get

$$\frac{\partial x^{\kappa}}{\partial x'^{\nu}} = \Lambda_{\nu}^{\kappa}. \quad (1.34)$$

Thus

$$\frac{\partial}{\partial x'^{\mu}} = \Lambda_{\mu}^{\nu} \frac{\partial}{\partial x^{\nu}}, \quad (1.35)$$

and we see that

$$\square_{\mu} \equiv \partial / \partial x^{\mu} = \left(\frac{1}{c} \frac{\partial}{\partial t}, \frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z} \right) \quad (1.36)$$

transforms like a down vector.

Notes:

(i)

$\frac{\partial}{\partial x^\mu}$ operates on scalars to produce vectors: $G_\mu \equiv \frac{\partial \phi}{\partial x^\mu} \equiv \square_\mu \phi \equiv \phi_{,\mu}$

$\frac{\partial}{\partial x^\mu}$ operates on vectors to produce 2nd rank tensors:

$$G_{\mu\nu} \equiv \frac{\partial A_\nu}{\partial x^\mu} \equiv \square_\mu A_\nu \equiv A_{\nu,\mu}$$

$\frac{\partial}{\partial x^\mu}$ operates on tensors to produce higher-rank tensors:

$$G_{\mu\lambda\nu} \equiv \frac{\partial B_{\lambda\nu}}{\partial x^\mu} \equiv \square_\mu B_{\lambda\nu} \equiv B_{\lambda\nu,\mu}$$

The operand's indices can be either up or down: $G_\mu{}^\nu = \square_\mu A^\nu$.

- (ii) If we contract the tensor produced by operating on a vector, we get a scalar, the 4-divergence $\psi = \square_\mu A^\mu$.
- (iii) We can reduce the number of indices on a higher-rank tensor by contraction: $A^\nu = \square_\mu G^{\mu\nu}$.
- (iv) The 4-analogue of taking the curl of a vector is to antisymmetrize the tensor formed by operating on a vector: $F_{\mu\nu} = (\square_\mu A_\nu - \square_\nu A_\mu)$. If $A_\nu = \square_\nu \phi$, then $F_{\mu\nu} = 0$ because partial derivatives commute.

Example:

In e.m. the usual vector potential \mathbf{A} and the electrostatic potential ϕ form the four components of an up vector

$$A^\mu = (\phi/c, A_x, A_y, A_z) \quad [\Rightarrow \quad A_\mu = (-\phi/c, A_x, A_y, A_z)]. \quad (1.37)$$

Our old friend the Maxwell field tensor \mathbf{F} is then

$$F_{\mu\nu} = \square_\mu A_\nu - \square_\nu A_\mu. \quad (1.38)$$

Thus $F_{12} = \frac{\partial A_y}{\partial x} - \frac{\partial A_x}{\partial y} = B_z$ and $F_{01} = \frac{\dot{A}_x}{c} + \frac{1}{c} \frac{\partial \phi}{\partial x} = -E_x/c$.

Derivatives with respect to proper time The history of a particle defines a curve in space-time. Let λ be a parameter which labels points on the curve in a continuous way. Then the coordinates x^μ of points on the curve are continuous functions $x^\mu(\lambda)$. For $\delta\lambda \ll 1$ the small vector

$$\delta\mathbf{x} \equiv \frac{d\mathbf{x}}{d\lambda} \delta\lambda$$

almost joins two points on the curve. Hence it is time-like and $|\delta\mathbf{x}| < 0$. For any two points A and B on the curve, we define

$$\tau \equiv \frac{1}{c} \int_A^B \sqrt{-\left|\frac{d\mathbf{x}}{d\lambda}\right|^2} d\lambda \quad (1.39)$$

to be the **proper time** difference between A and B along the curve. If the curve is a straight line, we may transform to the coordinate system in which $x^\mu = (ct, 0, 0, 0)$ at all points on the curve, and then

$$\tau = \frac{1}{c} \int_A^B \sqrt{-\frac{dct}{d\lambda} \frac{d(-ct)}{d\lambda}} d\lambda = [t_B - t_A]. \quad (1.40)$$

Hence the name. We regard the coordinates x^μ of events along the trajectory as functions $x^\mu(\tau)$ of the proper time. Differentiating w.r.t. τ and multiplying through by the rest mass m_0 we obtain a 4-vector, the momentum

$$\mathbf{p} \equiv m_0 \frac{d\mathbf{x}}{d\tau}. \quad (1.41)$$

From the zeroth component of the up version of this equation we have $dt = \gamma d\tau$; the hearts of passengers on a fast train (they mark off units of τ) appear to beat slowly to a medic on the station platform (whose watch keeps t).

1.6 Laws of e.m. and Mechanics in Tensor Form

The relativistic generalization of Newton's second law is

$$m_0 \frac{d^2 \mathbf{x}}{d\tau^2} = \frac{d}{d\tau} \left(m_0 \frac{d\mathbf{x}}{d\tau} \right) = \frac{d\mathbf{p}}{d\tau} = \mathbf{f}, \quad (1.42)$$

where \mathbf{f} is the **4-force**. The last three components of f^μ are just the Newtonian force components f_i . With $\mu = 0$ equation (1.42) states that the zeroth component of f^μ is to $1/c$ times the rate of change of the particle's energy cp^0 ; hence physically f^0 is $1/c$ times the rate of working of the force w . In summary

$$f^\mu = (w/c, f_x, f_y, f_z). \quad (1.43)$$

The divergence of (1.16) consists of these four equations:

$$F^{\mu\nu},_{\nu} = \begin{pmatrix} \frac{1}{c} \frac{\partial E_x}{\partial x} + \frac{1}{c} \frac{\partial E_y}{\partial y} + \frac{1}{c} \frac{\partial E_z}{\partial z} \\ \partial B_z / \partial y - \partial B_y / \partial z - \frac{1}{c^2} \partial E_x / \partial t \\ -\partial B_z / \partial x + \partial B_x / \partial z - \frac{1}{c^2} \partial E_y / \partial t \\ \partial B_y / \partial x - \partial B_x / \partial y - \frac{1}{c^2} \partial E_z / \partial t \end{pmatrix} = \begin{pmatrix} \frac{1}{c} \nabla \cdot \mathbf{E} \\ \nabla \times \mathbf{B} - \frac{1}{c^2} \frac{\partial \mathbf{E}}{\partial t} \end{pmatrix}. \quad (1.44)$$

The zeroth component is by Poisson's equation equal to $\rho/(c\epsilon_0) = c\mu_0\rho$, where ρ is the charge density. By Ampere's law, the last three of these equations are equal to $\mu_0\mathbf{j}$, where \mathbf{j} is the current density. Hence if we form a 4-vector

$$j^\mu = (c\rho, j_x, j_y, j_z), \quad (1.45)$$

we may write four of Maxwell's equations as

$$F^{\mu\nu},_{\nu} = \mu_0 j^\mu. \quad (1.46)$$

It is straightforward to verify that Maxwell's other four equations can be written

$$F_{\mu\nu,\lambda} + F_{\lambda\mu,\nu} + F_{\nu\lambda,\mu} = 0 \quad (\mu \neq \nu \neq \lambda). \quad (1.47)$$

Exercises (7):

- (i) Show that when λ, μ and ν equal 1, 2 and 3 respectively, (1.47) becomes $\nabla \cdot \mathbf{B} = 0$.
- (ii) Show that with equation (1.22) equation (47) may also be written $\overline{F}^{\mu\nu},_{\nu} = 0$.

Charge conservation is expressed as

$$\mu_0 \square \cdot \mathbf{j} = \mu_0 j^\mu,_{\mu} = F^{\mu\nu},_{\nu\mu} = 0, \quad (1.48)$$

where the last step follows by the antisymmetry of \mathbf{F} .

The natural definition of the 4-current associated with a particle of charge q is

$$\mathbf{J} = q \frac{d\mathbf{x}}{d\tau}. \quad (1.49)$$

Since the force exerted on a charged particle by an e.m. field has to be linear in q , the fields represented by \mathbf{F} , and the particle's velocity vector, a suitable 4-vector to try as the force is

$$\mathbf{f} = \mathbf{F} \cdot \mathbf{J}. \quad (1.50)$$

Tentatively inserting this into (1.42) and multiplying through by $d\tau/dt = 1/\gamma$ to obtain the acceleration as measured in the laboratory frame, we get

$$\frac{d\mathbf{p}}{dt} = q\mathbf{F} \cdot \frac{d\mathbf{x}}{dt}. \quad (1.51)$$

It is straightforward to check that the last three components of the up form of this vector are

$$\frac{d}{dt} \left(m_0 \gamma \frac{d\mathbf{x}}{dt} \right) = q(\mathbf{v} \times \mathbf{B} + \mathbf{E}),$$

while the zeroth component is

$$\frac{d(m_0 c \gamma)}{dt} = \frac{q}{c} \mathbf{E} \cdot \mathbf{v},$$

or, in words, “the rate of change of the particle’s energy mc^2 is equal to the rate of working of the Lorentz force.”

Gauge invariance At a classical (i.e. non-quantum level) only \mathbf{E} and \mathbf{B} are physically meaningful— \mathbf{A} is just an abstraction from which \mathbf{E} and \mathbf{B} can be calculated via $F_{\mu\nu} = (\square_\mu A_\nu - \square_\nu A_\mu)$. So nothing physical changes if we replace \mathbf{A} by

$$\mathbf{A}' \equiv \mathbf{A} + \square\Lambda, \quad (1.52)$$

where $\Lambda(\mathbf{x})$ is any scalar-valued function of space-time coordinates. The change (1.52) in \mathbf{A} is called a **gauge transformation**.

Gauge transformations can be used to ensure that \mathbf{A} satisfies an additional equation. In particular, given \mathbf{A} we can choose Λ s.t. \mathbf{A}' satisfies one of these gauge conditions:

(i) **Lorentz gauge**:

$$\square \cdot \mathbf{A}' = 0 \quad \Rightarrow \quad \square^2 \Lambda = \square \cdot \mathbf{A} \quad (1.53)$$

The Lorentz condition (1.53) does not uniquely specify \mathbf{A}' since many non-trivial functions satisfy $\square^2 \phi = 0$ and so given one Λ satisfying the 2nd of eqs (1.53), we can construct many others $\Lambda' = \Lambda + \phi$.

(ii) **Coulomb** or **radiation** or transverse gauge

$$\nabla \cdot \mathbf{A}' = 0 \quad \Rightarrow \quad \nabla^2 \Lambda = \nabla \cdot \mathbf{A} \quad (1.54)$$

In this gauge the 0th eqn of the set $\square^\nu F_{\mu\nu} = \mu_0 j_\mu$ reads

$$\begin{aligned} \frac{\rho}{c\epsilon_0} &= -\mu_0 j_0 = -\square^\nu (\square_0 A_\nu - \square_\nu A_0) \\ &= -\square_0 \square^\nu A_\nu + \square^\nu \square_\nu A_0 \\ &= -\square_0 \square^0 A_0 + \square^\nu \square_\nu A_0 \\ &= \square^i \square_i A_0 \\ &= -\nabla^2 \phi / c \end{aligned} \quad (1.55)$$

i.e., in this gauge the electrostatic potential satisfies Poisson’s eqn, which explains the gauge’s name.

1.7 Summary

The special theory of relativity requires that any physical quantity must fit into an n -tuple of numbers, where $n = 1, 4, 6, 10, \dots$. Physical laws must be expressed as equations connecting the n -tuples associated with different physical quantities. These equations must be constructed in accordance with the rules of tensor calculus, which permit only:

- (i) the multiplication of n -tuples to form either higher-rank n -tuples (as in $H_{\mu\nu} = x_\mu p_\nu - x_\nu p_\mu$) or lower-rank n -tuples (as in $f_\mu = F_\mu{}^\nu J_\nu$), or
- (ii) the addition of n -tuples *of the same rank*.

In particular, both sides of every acceptable equation always form valid n -tuples of the same kind.

Rest-mass, electric charge and total spin are scalars (1-tuples). The most important 4-vectors (4-tuples) include any particle's energy-momentum \mathbf{p} , e.m. current \mathbf{J} or acceleration $d\mathbf{p}/d\tau$, and the potential \mathbf{A} of the e.m. field. Important 6-tuples include any particle's angular momentum \mathbf{H} and the Maxwell field tensor \mathbf{F} . An important 10-tuple is the density \mathbf{T} of the energy-momentum due to the e.m. field.

In 4-vector notation the key equation of mechanics and e.m. are

$$\begin{aligned} \mathbf{v} &= \frac{d\mathbf{x}}{d\tau} \quad ; \quad \mathbf{p} = m_0 \mathbf{v} \quad ; \quad \mathbf{J} = q\mathbf{v} \\ \mathbf{f} &= \mathbf{F} \cdot \mathbf{J} \quad ; \quad \frac{d\mathbf{p}}{d\tau} = \mathbf{f} \\ F_{\mu\nu} &= \square_\mu A_\nu - \square_\nu A_\mu \quad ; \quad F^{\mu\nu}{}_{,\nu} = \mu_0 j_\mu \quad ; \quad \overline{F}^{\mu\nu}{}_{,\nu} = 0, \end{aligned}$$

where $F^{\mu\nu} \equiv \eta^{\mu\gamma} \eta^{\nu\delta} F_{\gamma\delta}$ and $\overline{F}^{\mu\nu} \equiv \frac{1}{2} \epsilon^{\mu\nu\gamma\delta} F_{\gamma\delta}$. The energy-momentum tensor of the e.m. field is

$$T^{\mu\nu} = \frac{1}{\mu_0} \left[\frac{1}{4} \text{Tr}(\mathbf{F} \cdot \mathbf{F}) \eta^{\mu\nu} - F^\mu{}_\gamma F^{\gamma\nu} \right].$$

1.8 Physics from invariance

In favourable cases the requirement that every physically interesting number fit into some n -tuple enables us to guess how things will transform under boosts from a knowledge of how they behave at zero velocity.

For example, the e.m. field of a moving electric dipole is bound to contain a magnetic component. It is plausible that this component is generated by a magnetic dipole and thus that a moving electric dipole is also a magnetic dipole.

Let the electric and magnetic dipoles be $\boldsymbol{\mu}$ and $\boldsymbol{\epsilon}$, respectively. Then the energies due to the two types of dipole are $-\boldsymbol{\epsilon} \cdot \mathbf{E}$ and $-\boldsymbol{\mu} \cdot \mathbf{B}$. These energies are obviously the same in a lab where everything is a mirror image of our lab. So $\boldsymbol{\epsilon}$ must be a vector like \mathbf{E} and $\boldsymbol{\mu}$ a pseudo-vector like \mathbf{B} . Since $\boldsymbol{\epsilon}$ and $\boldsymbol{\mu}$ are somehow coupled by boosts, it is more than probable that they are embedded in a common 6-tuple just as \mathbf{E} and \mathbf{B} are. In other words the following must be a 4-tensor:

$$\begin{pmatrix} 0 & c\epsilon_x & c\epsilon_y & c\epsilon_z \\ -c\epsilon_x & 0 & \mu_z & -\mu_y \\ -c\epsilon_y & -\mu_z & 0 & \mu_x \\ -c\epsilon_z & \mu_y & -\mu_x & 0 \end{pmatrix}. \quad (1.56)$$

It follows that $\boldsymbol{\epsilon} \cdot \mathbf{E} - \boldsymbol{\mu} \cdot \mathbf{B}$ is a 4-scalar.

2 Radiation of Electromagnetic Waves

In the Lorentz gauge (1.53), $\square_\nu F^{\mu\nu} = \mu_0 j^\mu$ becomes

$$\mu_0 j^\mu = \square^\mu \square_\nu A^\nu - \square^2 A^\mu = -\square^2 A^\mu, \quad (2.1)$$

so each component of \mathbf{A} obeys the wave equation with the corresponding component of $-\mu_0 \mathbf{j}$ as a source. So let's find particular integrals of

$$\square^2 \phi = s(\mathbf{x}), \quad (2.2)$$

for prescribed $s(\mathbf{x})$. The key is to find an appropriate **Green's function** $G(\mathbf{x}, \mathbf{x}')$, that is, a solution for $s = \delta(\mathbf{x} - \mathbf{x}')$:

$$\square_{\mathbf{x}}^2 G(\mathbf{x}, \mathbf{x}') = \delta(\mathbf{x} - \mathbf{x}'). \quad (2.3)$$

[(Here $\delta(\mathbf{x}) \equiv \delta(x_0)\delta(x_1)\delta(x_2)\delta(x_3)$.] Clearly $G(\mathbf{x}, \mathbf{x}') = G(\mathbf{x} - \mathbf{x}')$, so we first find G in the case $\mathbf{x}' = 0$ and then obtain G for general \mathbf{x}' by substituting $\mathbf{x} \rightarrow \mathbf{x} - \mathbf{x}'$ in our solution. Thus we now solve

$$\square^2 G(\mathbf{x}) = \delta(\mathbf{x}). \quad (2.4)$$

We look for a spherically symmetric $G(r, t)$:

$$\frac{1}{r^2} \frac{\partial}{\partial r} \left(r^2 \frac{\partial G}{\partial r} \right) - \frac{1}{c^2} \frac{\partial^2 G}{\partial t^2} = \delta(ct) \delta(\mathbf{x}). \quad (2.5)$$

This can be reduced to the ordinary wave equation for the new variable $S \equiv rG$:

$$\frac{\partial^2 S}{\partial r^2} - \frac{1}{c^2} \frac{\partial^2 S}{\partial t^2} = r \delta(ct) \delta(\mathbf{x}). \quad (2.6)$$

Away from $\mathbf{x} = 0$ this is the homogeneous wave eqn. So there $S = f(r - ct) + g(r + ct)$ and we see that G is a superposition of an outward-going wave (f) and an inward-going one (g). On physical grounds we take the amplitude of the latter to be zero: $g = 0$.

Physically we expect that a blip in s at $(r = 0, t = 0)$ will produce a disturbance confined to the surface $r = ct$, so we try $f(r - ct) = A\delta(r - ct)$. To check that this works and get the proportionality constant A , we integrate (2.4) with $G = A\delta(r - ct)/r$ from t_1 to t_2 :

$$\int_{t_1}^{t_2} dt \square^2 (A\delta(r - ct)/r) = \begin{cases} \nabla^2 \left(\frac{A}{r} \right) - \frac{A}{cr} \left[\frac{\partial}{\partial t} \delta(r - ct) \right]_{t_1}^{t_2} & \text{if } ct_1 < r < ct_2 \\ 0 & \text{otherwise.} \end{cases} \quad (2.7)$$

The second term on the upper right of (2.7) is zero because the δ -function and all its derivatives vanishes when its argument is non-zero, as here. On the other hand, we know from electrostatics that $\nabla^2 r^{-1} = -4\pi\delta(\mathbf{x})$. Hence the first term on the upper right of (2.7) is non-zero only if $r = 0 \Rightarrow t_1 < 0 < t_2$, and then is appropriately infinite. Bearing in mind the lower line on the right of (2.7), we see that $\square^2 (A\delta(r - ct)/r)$ is non-zero only in the neighbourhood of $\mathbf{x} = 0$, as we had hoped. Integrating both sides of (2.7) through a very small sphere centred on the origin, we see that

$$\int_{t_1 < t < t_2}^{r < a} d^4 \mathbf{x} \square^2 (A\delta(r - ct)/r) = -4\pi A. \quad (2.8)$$

Since we require $\int d^4 \mathbf{x} G = 1$, we set $A = -1/4\pi$ and have, finally

$$G(\mathbf{x}) = -\frac{\delta(|\mathbf{x}| - ct)}{4\pi|\mathbf{x}|} \quad \Rightarrow \quad G(\mathbf{x} - \mathbf{x}') = -\frac{\delta[|\mathbf{x} - \mathbf{x}'| - c(t - t')]}{4\pi|\mathbf{x} - \mathbf{x}'|}. \quad (2.9)$$

Box 1: Two Lemmas on δ -functions

(I) For any function $f(x)$ which never touches the x axis we have

$$\delta[f(x)] = \sum_{x_i \text{ s.t. } f(x_i)=0} \frac{\delta(x - x_i)}{|df/dx|}. \quad (\text{B1.1})$$

Proof: For arbitrary g form the integral $\int \delta[f(x)]g(x) dx$. Obviously this gets contributions only from the neighbourhoods of zeros x_i of f . In each such neighbourhood, f is a monotone function of x , so we can use f as the independent variable. Then $dx = df/(df/dx)$ and each interval contributes

$$\delta I = \frac{g}{df/dx} \Big|_{x_i} \int_{f(x_i-)}^{f(x_i+)} \delta(f) df, \quad (\text{B1.2})$$

If $df/dx < 0$, then $f(x_i-) > f(x_i+)$ and the integral in (B1.2) is -1 . On combining the contributions from different intervals, we therefore have

$$\int \delta[f(x)]g(x) dx = \sum_{x_i \text{ s.t. } f=0} \frac{g(x_i)}{|df/dx|_{x_i}}.$$

But this is just the integral of $g(x)$ times the r.h.s. of (B1.1).

(II) For any two functions a, b which never touch the x -axis we have

$$\delta(ab) = \frac{\delta(a)}{|b|} + \frac{\delta(b)}{|a|}. \quad (\text{B1.3})$$

Proof: The zeros of ab split into those where $a = 0$ and those where $b = 0$. So with (B1.1) we have

$$\delta(ab) = \sum_{x_i \text{ s.t. } a=0} \frac{\delta(x - x_i)}{|b(da/dx)|} + \sum_{x_i \text{ s.t. } b=0} \frac{\delta(x - x_i)}{|a(db/dx)|}. \quad (\text{B1.4})$$

But this is precisely what we get on applying (B1.1) to the r.h.s. of (B1.3).

Eqn (2.9) is not manifestly Lorentz invariant, as one might expect G to be since it solves (2.3), which is. However, from Lemma II in the box we have

$$\begin{aligned} \delta(|\mathbf{x} - \mathbf{x}'|) &= \delta[-c^2(t - t')^2 + |\mathbf{x} - \mathbf{x}'|^2] \\ &= \delta[(-c(t - t') + |\mathbf{x} - \mathbf{x}'|)(c(t - t') + |\mathbf{x} - \mathbf{x}'|)] \\ &= \frac{\delta[-c(t - t') + |\mathbf{x} - \mathbf{x}'|]}{2|\mathbf{x} - \mathbf{x}'|} + \frac{\delta[c(t - t') + |\mathbf{x} - \mathbf{x}'|]}{2|\mathbf{x} - \mathbf{x}'|}, \end{aligned} \quad (\text{2.10})$$

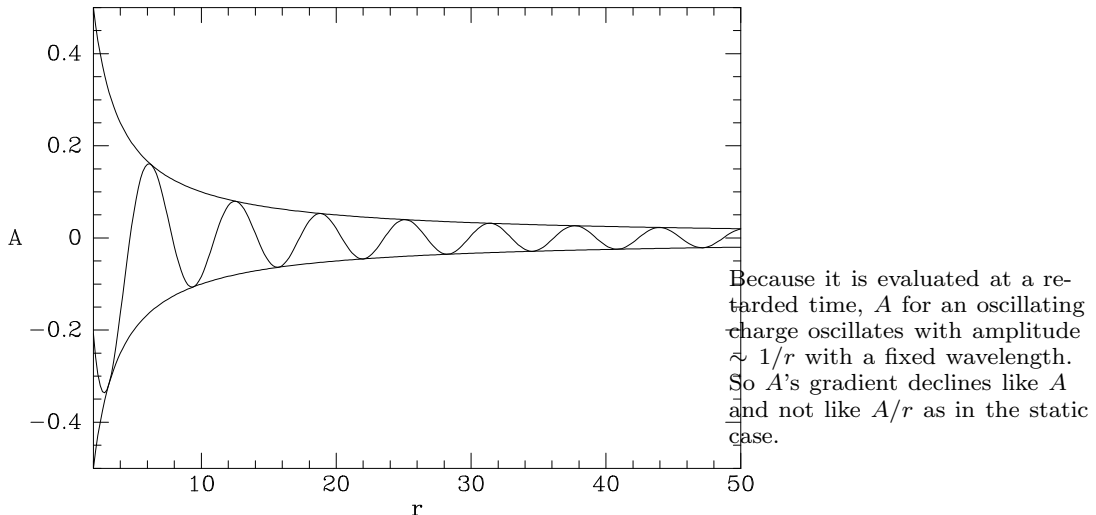
(2.10) allows us to rewrite (2.9) in the manifestly invariant form

$$G(\mathbf{x} - \mathbf{x}') = -\frac{1}{2\pi} \theta(t - t') \delta(|\mathbf{x} - \mathbf{x}'|) \quad \text{where} \quad \theta(t) = \begin{cases} 0 & t < 0 \\ 1 & t > 0 \end{cases}. \quad (\text{2.11})$$

We can now write down the solution to (2.1) in two forms:

$$\begin{aligned} \mathbf{A}(\mathbf{x}) &= -\mu_0 \int G(\mathbf{x}, \mathbf{x}') \mathbf{j}(\mathbf{x}') d^4 \mathbf{x}' \\ &= \frac{\mu_0}{2\pi} \int \theta(t - t') \delta(|\mathbf{x} - \mathbf{x}'|) \mathbf{j}(\mathbf{x}') d^4 \mathbf{x}' \end{aligned} \quad (\text{2.12})$$

$$\begin{aligned} \mathbf{A}(t, \mathbf{x}) &= \mu_0 \int d^3 \mathbf{x}' \int dt' \mathbf{j}(t', \mathbf{x}') \frac{\delta[|\mathbf{x} - \mathbf{x}'| - c(t - t')]}{4\pi |\mathbf{x} - \mathbf{x}'|} \\ &= \frac{\mu_0}{4\pi} \int \frac{\mathbf{j}(t - |\mathbf{x} - \mathbf{x}'|/c, \mathbf{x}')}{|\mathbf{x} - \mathbf{x}'|} d^3 \mathbf{x}'. \end{aligned} \quad (\text{2.13})$$



The quantity \mathbf{A} given by these eqns is called the **retarded potential**. From (2.12) or (2.13) we can find the potential \mathbf{A} generated by any system of charges.

2.1 Dipole radiation

What potential does (2.13) predict for a point very far away from a fluctuating charge distribution? We put the coordinate origin in the middle of the charge distribution. Then in $|\mathbf{x} - \mathbf{x}'|$ we can neglect \mathbf{x}' by comparison with \mathbf{x} , and have

$$\mathbf{A}(t, \mathbf{x}) \simeq \frac{\mu_0}{4\pi|\mathbf{x}|} \int \mathbf{j}(t - |\mathbf{x}|/c, \mathbf{x}') d^3\mathbf{x}'. \quad (2.14)$$

The equation of charge conservation, $\partial\rho/\partial t = -\nabla \cdot \mathbf{j}$, enables us to relate the spatial components of this to the dipole moment of the charges grouped around the origin. The latter is

$$\mathbf{p} \equiv \int \rho(\mathbf{x}) \mathbf{x} d^3\mathbf{x} \quad (2.15)$$

so

$$\begin{aligned} \frac{d\mathbf{p}}{dt} &= \int \frac{\partial\rho}{\partial t} \mathbf{x} d^3\mathbf{x} \\ &= - \int (\nabla \cdot \mathbf{j}) \mathbf{x} d^3\mathbf{x} \\ &= \int (\mathbf{j} \cdot \nabla) \mathbf{x} d^3\mathbf{x} \\ &= \int \mathbf{j} d^3\mathbf{x}. \end{aligned} \quad (2.16)$$

Thus

$$\mathbf{A}(t, \mathbf{x}) \simeq \frac{\mu_0}{4\pi|\mathbf{x}|} \left. \frac{d\mathbf{p}}{dt} \right|_{(t-|\mathbf{x}|/c)}. \quad (2.17)$$

Notice that the amplitude of the potential falls off as $1/r$ as for the electrostatic field of a point charge. But if the charge distribution fluctuates, the potential acquires extra spatial variation through $\dot{\mathbf{p}}$, and this enables the fields \mathbf{E} , \mathbf{B} that one derives from \mathbf{A} to be fall off like $1/r$ also. This is the effect that enables electrons in the Sun to perceptibly wiggle your electrons as you lie on a beach 93,000,000 miles away.

Exercise (8):

Show that (2.14) gives $A^0 = 0$ in the far field. [Hint: exploit charge conservation.]

At what rate does a fluctuating charge distribution radiate energy? Ideally we'd calculate this by integrating the energy-momentum tensor \mathbf{T} around a sphere. But we want to avoid calculating curls in spherical polars because that would be tedious, so we cheat a bit and argue that at large \mathbf{x} the wave-fronts can be treated as planar. For a planar wave propagating in the $\hat{\mathbf{k}}$ direction

$$\mathbf{A}(t, \mathbf{x}) = \tilde{\mathbf{A}}f(t - \hat{\mathbf{k}} \cdot \mathbf{x}/c), \quad (2.18)$$

where $\tilde{\mathbf{A}}$ is a constant vector. This gives

$$\mathbf{B} = \nabla \times \mathbf{A} = -\frac{1}{c}\hat{\mathbf{k}} \times \tilde{\mathbf{A}}f'(t - \hat{\mathbf{k}} \cdot \mathbf{x}/c).$$

Furthermore, the set $(\mathbf{E}, \mathbf{B}, \mathbf{k})$ is known to form a right-handed set of mutually orthogonal vectors. Thus

$$\mathbf{E} = c\mathbf{B} \times \hat{\mathbf{k}}.$$

The Poynting vector is therefore

$$\begin{aligned} \mathbf{N} &= \frac{1}{\mu_0}\mathbf{E} \times \mathbf{B} = \frac{cB^2}{\mu_0}\hat{\mathbf{k}} \\ &= \frac{1}{\mu_0 c}[\tilde{A}^2 - (\hat{\mathbf{k}} \cdot \tilde{\mathbf{A}})^2]f'^2\hat{\mathbf{k}}. \end{aligned} \quad (2.19)$$

Comparing (2.17) with (2.18) we have that

$$\hat{\mathbf{k}} = \hat{\mathbf{x}} \quad \text{and} \quad \tilde{\mathbf{A}}f = \frac{\mu_0}{4\pi|\mathbf{x}|}\dot{\mathbf{p}}. \quad (2.20)$$

Using this in (2.19) we find

$$\mathbf{N} = \frac{1}{\mu_0 c} \left(\frac{\mu_0}{4\pi|\mathbf{x}|} \right)^2 [\ddot{\mathbf{p}}^2 - (\hat{\mathbf{x}} \cdot \ddot{\mathbf{p}})^2] \hat{\mathbf{x}}. \quad (2.21)$$

The angular distribution of the flux $|\mathbf{N}|$ is that of a quadrupole field: it vanishes for $\mathbf{x} \parallel \ddot{\mathbf{p}}$ and peaks in the orthogonal directions. Integrating over a sphere of radius $|\mathbf{x}|$ we find that the radiated power is

$$P = \frac{\mu_0|\ddot{\mathbf{p}}|^2}{6\pi c}. \quad (2.22)$$

This is known as **Larmor's formula** for the power radiated by a system of charges. If the charge distribution consists of a single charge with position vector \mathbf{r} , then $\mathbf{p} = q\mathbf{r}$ and from (2.22) the power radiated is

$$P = \frac{q^2}{4\pi\epsilon_0} \frac{2|\ddot{\mathbf{r}}|^2}{3c^3}. \quad (2.23)$$

(Notice that power has units of *force* \times *speed*, so $q^2/4\pi\epsilon_0$ needs to be multiplied by something with units V/L^2 made out of $\ddot{\mathbf{r}}$ and c .)

2.2 Single accelerated charge*

One important special case is that of an accelerated particle of charge q : if the charge's world-line is $\mathbf{X}(\tau)$, one has

$$\begin{aligned} \mathbf{j}(\mathbf{x}) &= qc \int \dot{\mathbf{X}}\delta(\mathbf{x} - \mathbf{X}) d\tau \\ &= qc \int \delta(\mathbf{x} - \mathbf{X}) d\mathbf{X}. \end{aligned} \quad (2.24)$$

* Lies beyond the syllabus

Exercise (9):

Check the validity of (2.24) by (i) showing that it is dimensionally correct, (ii) showing that $\int \mathbf{j} d^3\mathbf{x} = q(d\mathbf{X}/dt)$, i.e., the total current is just q times the Newtonian velocity, and (iii) showing similarly that the total charge in any spatial slice is always q .

Plugging (2.24) into (2.12) yields

$$\begin{aligned} \mathbf{A}(\mathbf{x}) &= \frac{\mu_0 q c}{2\pi} \int d^4\mathbf{x}' \theta(t-t') \delta(|\mathbf{x}-\mathbf{x}'|) \int \dot{\mathbf{X}} \delta(\mathbf{x}'-\mathbf{X}) d\tau \\ &= \frac{\mu_0 q c}{2\pi} \int \theta(t-T) \delta(|\mathbf{x}-\mathbf{X}|) \dot{\mathbf{X}} d\tau \\ &= -\frac{q}{4\pi\epsilon_0 c} \frac{\dot{\mathbf{X}}}{(\mathbf{x}-\mathbf{X}) \cdot \dot{\mathbf{X}}} \Big|_{|\mathbf{x}-\mathbf{X}|=0}. \end{aligned} \quad (2.25)$$

where use has again been made of (B1.1). This is called the **Liénard-Weichert** potential. Notice that for $\dot{\mathbf{X}} = (c, \mathbf{o})$ it gives the Coulomb potential as it should.

3 Lagrangian Mechanics

3.1 Paths, functionals & the calculus of variations

Before a 'plane takes off from New York for London, its computer chooses an optimal path $\mathbf{x}(t)$; i.e., it finds that sequence of longitudes, latitudes and altitudes at each moment t of the flight which, given prevailing winds, will get it to London at the prescribed time with least expenditure of fuel. The quantity of fuel required to get to London in a given time is a single number F that depends on the whole path $\mathbf{x}(t)$; one says that F is a **functional** $F[\mathbf{x}]$ of the path $\mathbf{x}(t)$.

The simplest functionals are integrals along the path of functions of $\mathbf{x}(t)$ and its derivatives with respect to t :

$$\begin{aligned} F_1[\mathbf{x}] &\equiv \int_0^{t_0} |\mathbf{x}(t)|^2 dt \\ F_2[\mathbf{x}] &\equiv \int_0^{t_0} |\dot{\mathbf{x}}(t)|^2 dt \\ F_3[\mathbf{x}] &\equiv \int_0^{t_0} \mathbf{x} \cdot \dot{\mathbf{x}}(t) dt \\ &\dots\dots \end{aligned}$$

How do we find the path that minimizes a functional

$$F[\mathbf{x}] \equiv \int_0^{t_0} f(\mathbf{x}, \dot{\mathbf{x}}) dt?$$

Let $\bar{\mathbf{x}}(t)$ be the minimizing path and let $\boldsymbol{\eta}(t)$ be a small variation, so that $\mathbf{x}(t) \equiv \bar{\mathbf{x}}(t) + \boldsymbol{\eta}(t) \approx \bar{\mathbf{x}}(t)$. We insist on $\boldsymbol{\eta}$ vanishing at $t = 0, t_0$ so that $\bar{\mathbf{x}}(t)$ and the modified path both start and finish at the same places at the same times. Then

$$\begin{aligned} F[\bar{\mathbf{x}}] \leq F[\mathbf{x}] &= \int_0^{t_0} f(\bar{\mathbf{x}} + \boldsymbol{\eta}, \dot{\bar{\mathbf{x}}} + \dot{\boldsymbol{\eta}}) dt \\ &= \int_0^{t_0} \left(f(\bar{\mathbf{x}}, \dot{\bar{\mathbf{x}}}) + \frac{\partial f}{\partial \mathbf{x}} \cdot \boldsymbol{\eta} + \frac{\partial f}{\partial \dot{\mathbf{x}}} \cdot \dot{\boldsymbol{\eta}} + \dots \right) dt \\ &= F[\bar{\mathbf{x}}] + \int_0^{t_0} \left(\frac{\partial f}{\partial \mathbf{x}} \cdot \boldsymbol{\eta} + \frac{\partial f}{\partial \dot{\mathbf{x}}} \cdot \dot{\boldsymbol{\eta}} + \dots \right) dt. \end{aligned} \quad (3.1)$$

We now integrate by parts the second term in the integral of the last line:

$$\int_0^{t_0} \frac{\partial f}{\partial \dot{\mathbf{x}}} \cdot \dot{\boldsymbol{\eta}} dt = \left[\frac{\partial f}{\partial \dot{\mathbf{x}}} \cdot \boldsymbol{\eta} \right]_0^{t_0} - \int_0^{t_0} \frac{d}{dt} \left(\frac{\partial f}{\partial \dot{\mathbf{x}}} \right) \cdot \boldsymbol{\eta} dt. \quad (3.2)$$

Since $\boldsymbol{\eta}(0) = \boldsymbol{\eta}(t_0) = 0$, the $[\cdot]$ vanishes. Putting this into (3.1) we have

$$0 \geq F[\mathbf{x}] - F[\bar{\mathbf{x}}] = \int_0^{t_0} \left[\left(\frac{\partial f}{\partial \mathbf{x}} - \frac{d}{dt} \frac{\partial f}{\partial \dot{\mathbf{x}}} \right) \cdot \boldsymbol{\eta} + \dots \right] dt. \quad (3.3)$$

This relation must hold for any $\boldsymbol{\eta}$, no matter how small. So the higher terms indicated by $+\dots$ can be neglected. The remaining integrand is proportional to $\boldsymbol{\eta}$, so if it were non-zero for some particular function $\boldsymbol{\eta}(t)$, it would have the opposite sign for $\boldsymbol{\eta}' \equiv -\boldsymbol{\eta}$. The inequality on the extreme left would then be violated for one of $\boldsymbol{\eta}$ or $\boldsymbol{\eta}'$. Hence the integral must vanish for all $\boldsymbol{\eta}$. This is possible only if the coefficient of $\boldsymbol{\eta}$ vanishes for all $0 < t < t_0$: if it did not vanish for some t , say t_1 , the integral would fail to vanish for the particular choice $\boldsymbol{\eta} = \delta(t - t_1)$. So $\bar{\mathbf{x}}(t)$ minimizes F if and only if

$$\frac{d}{dt} \frac{\partial f}{\partial \dot{\mathbf{x}}} - \frac{\partial f}{\partial \mathbf{x}} = 0. \quad (3.4)$$

This is called the **Euler-Lagrange equation** ('EL eqn'), and the theory that underlies it is called **the calculus of variations**. It is one of the few results we have in the theory of functionals—one everywhere in physics encounters problems that cry out for a fully fledged calculus of functions that shows how to integrate, Taylor expand, exponentiate etc functionals the way we do functions.

Legend has it that the calculus of variations was invented by Newton after dinner one evening to solve this challenge problem (set in 1695 by Johann Bernoulli):

Exercise (10):

A bead slides on a smooth wire that passes through two rings, one at the origin, the other at $(x', y', z') = (x_0, 0, -z_0)$ with $z_0 > 0$. To what curve (the 'brachistochrone') must the wire be bent in order to minimize the time required for the bead to slide from rest at the upper ring to the lower ring?

Solution: The optimal curve obviously lies in the plane $y' = 0$. It is convenient to work in coordinates (x, y, z) such that z increases downwards. Then the time of flight is

$$\tau = \int_0^{z_0} \frac{dz}{\dot{z}}.$$

But $\frac{1}{2}(\dot{x}^2 + \dot{z}^2) = gz$, so $\dot{z} = \sqrt{2gz/[(dx/dz)^2 + 1]}$ and

$$\tau = \int_0^{z_0} \frac{dz}{\sqrt{2gz}} \sqrt{\left(\frac{dx}{dz}\right)^2 + 1}.$$

We need to minimize this with respect to the path $x(z)$. The EL eqn tells us that the optimal path satisfies

$$0 = \frac{d}{dz} \left(\frac{dx/dz}{\sqrt{z} \sqrt{(dx/dz)^2 + 1}} \right),$$

which implies

$$x(z) = \int_0^z \sqrt{\frac{Az}{1 - Az}} dz,$$

where A is a constant of integration. (A may be determined by first solving $x_0/z_0 = (\theta_0 - \frac{1}{2} \sin 2\theta_0)/\sin^2 \theta_0$ for θ_0 and then using this value in $A = \sin^2 \theta_0/z_0$.) In terms of variable $\sin^2 \theta \equiv Az$ the answer is

$$x = \frac{1}{A} \left(\theta - \frac{1}{2} \sin 2\theta \right). \quad (3.5)$$

If we write $\phi \equiv 2\theta$ this may be written $z = (1 - \cos \phi)/2A$, $x = (\phi - \sin \phi)/2A$, which is a cycloid with the origin at its cusp.

3.2 Lagrangian for relativistic motion

Quantum mechanics ensures that every particle moves from one event \mathbf{x}_1 to another \mathbf{x}_2 in such a way that a functional S of its world-line between these events, $\mathbf{x}(\tau)$, is extremized. We'll see how q.m. does this in §5. But granted that it *can* be done, it is easy to see what the functional $S[\mathbf{x}]$ must be: we write

$$S[\mathbf{x}] = \int s(\mathbf{x}) d\tau$$

and ask what form the function s may take. S is called the **action** for single-particle motion. The action must be a scalar since there is no obvious higher n -tuple into which we can fit it. τ is a scalar, so s must be a scalar too, and we have only to ask what scalars we can construct from the world-line $\mathbf{x}(\tau)$ and quantities such as \mathbf{A} , \mathbf{F} associated with the e.m. field.

First we note that S shouldn't depend on our choice of origin, so only derivatives $\dot{\mathbf{x}}$, $\ddot{\mathbf{x}}$ etc should occur in s , not \mathbf{x} itself. Furthermore, the EL eqn involves differentiation with respect to the variable that parameterizes position along the extremal path, in this case τ . So we will get as 2nd-order eqn of motion, if s depends on $\dot{\mathbf{x}}$, but not on higher derivatives of $\mathbf{x}(\tau)$. Similarly, the EL eqn involves differentiation w.r.t. the general position vector \mathbf{x} , so if the eqn of motion is to depend on \mathbf{F} and not its derivatives, s should depend on \mathbf{A} but not \mathbf{F} . So the invariants to consider are (i) $|\dot{\mathbf{x}}|^2 = -c^2$ and (ii) $\dot{\mathbf{x}} \cdot \mathbf{A}$ —we exclude $|\mathbf{A}|^2$ from consideration since its contribution to S proves to be both gauge- and path-dependent. So the simplest thing to try is

$$S = \int (-m_0 c^2 + q \dot{\mathbf{x}} \cdot \mathbf{A}) d\tau, \quad (3.6)$$

where we've included the rest mass m_0 for future convenience and q is some constant.

Exercise (11):

Show that the gauge-dependent part of the action (3.6) is path-independent.

Unfortunately we cannot apply the EL eqn to (3.6) as it stands because we want to hold constant the events of arrival and departure, \mathbf{x}_1 and \mathbf{x}_2 , rather than the proper-time elapse between these events. So we have first to eliminate τ from (3.6) in favour of, say, t . $dt/d\tau = p^0/m_0 c = \gamma$, so

$$S = \int \left(-m_0 c^2 \sqrt{1 - v^2/c^2} + q \frac{d\mathbf{x}}{dt} \cdot \mathbf{A} \right) dt. \quad (3.7)$$

If this guess is to be right it must predict rectilinear motion when $\mathbf{A} = 0$. Applying (3.4) with $\mathbf{A} = 0$ gives

$$\frac{d}{dt} \frac{m_0 \mathbf{v}}{\sqrt{1 - v^2/c^2}} = 0 \quad \Rightarrow \quad \frac{d\gamma \mathbf{v}}{dt} = 0. \quad (3.8)$$

The correctness of this result is manifested by multiplying through by γ to produce

$$m_0 \frac{d}{d\tau} \frac{d\mathbf{x}}{d\tau} = 0.$$

So far so good. Now for the case $\mathbf{A} \neq 0$. Since $A^\mu = (\phi/c, \mathbf{A})$ and $dx_\mu/dt = (-c, \mathbf{v})$, the action is

$$S = \int \left[-m_0 c^2 \sqrt{1 - v^2/c^2} + q(-\phi + \mathbf{v} \cdot \mathbf{A}) \right] dt, \quad (3.9)$$

so the EL eqn is

$$\frac{d}{dt} \left(\frac{m_0 \mathbf{v}}{\sqrt{1 - v^2/c^2}} + q \mathbf{A} \right) + q \nabla(\phi - \mathbf{v} \cdot \mathbf{A}) = 0. \quad (3.10)$$

Here the derivative w.r.t. t is along the path, so

$$\frac{d\mathbf{A}}{dt} = \frac{\partial \mathbf{A}}{\partial t} + (\mathbf{v} \cdot \nabla) \mathbf{A}. \quad (3.11)$$

The partial derivative here can be combined with the $\nabla\phi$ term in (3.10) to produce $\mathbf{E} = -\nabla\phi - \partial\mathbf{A}/\partial t$. Putting all these things back into the EL eqn (3.10) yields

$$m_0 \frac{d\gamma\mathbf{v}}{dt} = q[\mathbf{E} + \nabla(\mathbf{v} \cdot \mathbf{A}) - (\mathbf{v} \cdot \nabla)\mathbf{A}]. \quad (3.12)$$

It's now straightforward to show that the last two terms on the right of (3.12) equal $\mathbf{v} \times \mathbf{B}$ as one would hope: bearing in mind that $\nabla\mathbf{v} = 0$ we have

$$\begin{aligned} \mathbf{v} \times \mathbf{B} &= \mathbf{v} \times (\nabla \times \mathbf{A}) \\ &= \nabla(\mathbf{v} \cdot \mathbf{A}) - (\mathbf{v} \cdot \nabla)\mathbf{A} \end{aligned}$$

Thus the EL eqn applied to the action (3.6) gives

$$m_0 \frac{d\gamma\mathbf{v}}{dt} = q(\mathbf{E} + \mathbf{v} \times \mathbf{B}) \quad (3.13)$$

and we are able to identify q with the particle's charge.

3.3 Lagrangian for non-relativistic motion

From (3.9) the action for non-relativistic motion in an electrostatic field is

$$S = \int [-m_0c^2 + L(\mathbf{x}, \dot{\mathbf{x}}, t)] dt, \quad \text{where} \quad L(\mathbf{x}, \dot{\mathbf{x}}) \equiv \frac{1}{2}m_0\dot{\mathbf{x}}^2 - q\phi(\mathbf{x}, t). \quad (3.14)$$

Since $\int m_0c^2 dt$ is the same for all paths that start and finish at the given events, it plays no role in picking out the true path. So it can be dropped, and we obtain the **principle of least action**:

$$\delta S = 0 \quad \text{where} \quad S \equiv \int L(\mathbf{x}, \dot{\mathbf{x}}, t) dt. \quad (3.15)$$

The function L is called the **Lagrangian**. By (3.14) it is the difference between the particle's kinetic and potential energies.

The equations of motion of every conservative dynamical system can be derived from a Lagrangian. At heart every system is conservative; friction arises from neglect of the microscopic degrees of freedom to which energy tends to be transferred from the macroscopic ones to which the eqns of motion apply. So Lagrangians are a very general tool.

The Lagrangian of a dynamical system depends on the system's n coordinates, conventionally written \mathbf{q} and velocities $\dot{\mathbf{q}}$. The $2n$ -dimensional space of initial conditions $(\mathbf{q}, \dot{\mathbf{q}})$ is called **phase space**. (Notice that we are here treating $\dot{\mathbf{q}}$ as completely independent of \mathbf{q} ; from the point of view of phase space \mathbf{q} isn't part of a path, just a configuration of the system.) The eqns of motion allow one to determine uniquely the system's future and past from its present position in phase space. Geometrically, through every point of phase space there runs a curve along which the system evolves. These curves never intersect one another.

3.4 Equations of motion from Lagrangians

Lagrangians provide a neat way of calculating eqns of motion in odd coordinate systems because it is easier to transform a single function to new-fangled coordinates than a set of eqns of motion. Consider, for example, motion in a rotating frame.

Suppose both primed and unprimed coordinates share the same origin, but the primed coordinates rotate with angular velocity $\boldsymbol{\omega}$ with respect to the unprimed coordinates, which are inertial. Then

$$\mathbf{v}_{\text{inertial}} = \dot{\mathbf{r}}' + \boldsymbol{\omega} \times \mathbf{r}'.$$

So written in terms of the primed coordinates the k.e. is

$$\begin{aligned} T &= \frac{1}{2}mv^2 = \frac{1}{2}m|\dot{\mathbf{r}}' + \boldsymbol{\omega} \times \mathbf{r}'|^2 \\ &= \frac{1}{2}m|\dot{\mathbf{r}}'|^2 + m\dot{\mathbf{r}}' \cdot (\boldsymbol{\omega} \times \mathbf{r}') + \frac{1}{2}m|\boldsymbol{\omega} \times \mathbf{r}'|^2 \end{aligned} \quad (3.16)$$

The p.e. is just $V(\mathbf{r}', t)$, and we have $\dot{\mathbf{r}}' \cdot (\boldsymbol{\omega} \times \mathbf{r}') = \mathbf{r}' \cdot (\dot{\mathbf{r}}' \times \boldsymbol{\omega})$, so

$$L = \frac{1}{2}m|\dot{\mathbf{r}}'|^2 + m\dot{\mathbf{r}}' \cdot (\boldsymbol{\omega} \times \mathbf{r}') + \frac{1}{2}m|\boldsymbol{\omega} \times \mathbf{r}'|^2 - V \quad (3.17)$$

and the EL eqns are

$$\begin{aligned} 0 &= \frac{d}{dt} \frac{\partial L}{\partial \dot{\mathbf{r}}'} - \frac{\partial L}{\partial \mathbf{r}'} \\ &= \frac{d}{dt} (m\dot{\mathbf{r}}' + m\boldsymbol{\omega} \times \mathbf{r}') - \left[m\dot{\mathbf{r}}' \times \boldsymbol{\omega} + \frac{\partial}{\partial \mathbf{r}'} \left(\frac{1}{2}m|\boldsymbol{\omega} \times \mathbf{r}'|^2 - V \right) \right]. \end{aligned} \quad (3.18)$$

Collecting everything together we have finally

$$m\ddot{\mathbf{r}}' = 2m\dot{\mathbf{r}}' \times \boldsymbol{\omega} - \frac{\partial V_{\text{eff}}}{\partial \mathbf{r}'} \quad \text{where} \quad V_{\text{eff}} \equiv V - \frac{1}{2}m|\boldsymbol{\omega} \times \mathbf{r}'|^2. \quad (3.19)$$

In a rotating frame there is a contribution to the ‘‘acceleration’’ $\ddot{\mathbf{r}}'$ from the **Coriolis force** $2m\boldsymbol{\omega} \times \dot{\mathbf{r}}'$, and the potential needs to be augmented by a term that gives rise to the **centrifugal force** $m\boldsymbol{\omega} \times (\boldsymbol{\omega} \times \mathbf{r}') = -m(\boldsymbol{\omega} \cdot \mathbf{r}')\boldsymbol{\omega}$. Forces such as these, which appear because one’s frame is non-inertial, are called **pseudo-forces**.

A second example illustrates that Lagrangians work even for coordinates that depend explicitly on time. In cosmology it is handy to use ‘comoving’ coordinates such that the spatial coordinates of particles that move apart as the Universe expands are constant. Let the primed system be inertial and the unprimed system comoving. Then $\mathbf{r}' = a(t)\mathbf{r}$, where $a(t)$ is the cosmic scale factor. So

$$T = \frac{1}{2}m\dot{\mathbf{r}}'^2 = \frac{1}{2}m(a\dot{\mathbf{r}} + \dot{a}\mathbf{r})^2. \quad (3.20)$$

Writing the potential energy as $V = m\Phi$ the EL eqns are

$$0 = \frac{d}{dt} [m(a\dot{\mathbf{r}} + \dot{a}\mathbf{r})] - m(a\dot{\mathbf{r}} + \dot{a}\mathbf{r})\dot{a} + m\frac{\partial \Phi}{\partial \mathbf{r}}.$$

Cleaning up we get

$$\ddot{\mathbf{r}} + 2\frac{\dot{a}}{a}\dot{\mathbf{r}} + \frac{\ddot{a}}{a}\mathbf{r} = -\frac{1}{a^2}\frac{\partial \Phi}{\partial \mathbf{r}}. \quad (3.21)$$

A final example illustrates how to get T in a weird curvilinear coordinate system. Oblate spheroidal coordinates (u, v, ϕ) are related to regular cylindrical polars (R, z, ϕ) by

$$R = \Delta \cosh u \cos v \quad ; \quad z = \Delta \sinh u \sin v. \quad (3.22)$$

Slightly changing u , v and ϕ in turn while leaving the other coordinates alone, generates small displacements

$$\begin{aligned} \mathbf{e}_u &= \Delta \delta u (\sinh u \cos v \hat{\mathbf{R}} + \cosh u \sin v \hat{\mathbf{z}}) \\ \mathbf{e}_v &= \Delta \delta v (-\cosh u \sin v \hat{\mathbf{R}} + \sinh u \cos v \hat{\mathbf{z}}) \\ \mathbf{e}_\phi &= R \delta \phi. \end{aligned}$$

The R , z and ϕ unit vectors are orthogonal, so the distance one goes on changing all of (u, v, ϕ) simultaneously is

$$\begin{aligned} ds^2 &= \Delta^2 [(\delta u)^2 (\sinh^2 u \cos^2 v + \cosh^2 u \sin^2 v) \\ &\quad + (\delta v)^2 (\cosh^2 u \sin^2 v + \sinh^2 u \cos^2 v) + (\delta \phi)^2 \cosh^2 u \cos^2 v] \\ &= \Delta^2 \{ (\cosh^2 u - \cos^2 v) [(\delta u)^2 + (\delta v)^2] + \cosh^2 u \cos^2 v (\delta \phi)^2 \}. \end{aligned} \quad (3.23)$$

Dividing through by dt^2 we get the kinetic energy in terms of $(\dot{u}, \dot{v}, \dot{\phi})$:

$$T = \frac{1}{2}m\Delta^2\{(\cosh^2 u - \cos^2 v)[\dot{u}^2 + \dot{v}^2] + \cosh^2 u \cos^2 v \dot{\phi}^2\}. \quad (3.24)$$

The eqns of motion are therefore

$$\begin{aligned} m\Delta^2 \left\{ \frac{d}{dt} [((\cosh^2 u - \cos^2 v)\dot{u}) - \frac{1}{2} \sinh 2u(\dot{u}^2 + \dot{v}^2 + \cos^2 v \dot{\phi}^2)] \right\} + \frac{\partial V}{\partial u} &= 0 \\ m\Delta^2 \left\{ \frac{d}{dt} [((\cosh^2 u - \cos^2 v)\dot{v}) - \frac{1}{2} \sin 2v(\dot{u}^2 + \dot{v}^2 - \cosh^2 u \dot{\phi}^2)] \right\} + \frac{\partial V}{\partial v} &= 0 \\ m\Delta^2 \left[\frac{d}{dt} (\cosh^2 u \cos^2 v \dot{\phi}) \right] + \frac{\partial V}{\partial \phi} &= 0. \end{aligned}$$

3.5 Normal modes from Lagrangians

Obviously, when a system is in equilibrium all its time derivatives vanish. From the EL eqns we infer that equilibrium configurations correspond to $\partial V/\partial q_i = 0$, where q_i is any coordinate. (q_i is often called a **generalized coordinate** to emphasize that it needn't be Cartesian or otherwise special.) By expanding $V(\mathbf{q})$ around the stationary point \mathbf{q}_s corresponding to an equilibrium configuration and plugging the expansion into the EL eqns, one sees that the equilibrium is stable if \mathbf{q}_s is a local minimum of V , and unstable otherwise.

When slightly disturbed from an stable equilibrium, the system will oscillate in a motion that can be represented as a superposition of normal modes. Lagrangians provide a relatively painless route to the frequencies and forms of these normal modes. The trick is to expand $L(\mathbf{q}, \dot{\mathbf{q}})$ in a Taylor series around the equilibrium configuration $\mathbf{q} = \mathbf{q}_s$, $\dot{\mathbf{q}} = 0$, discarding terms of higher than second order in $\mathbf{q} - \mathbf{q}_s$ and its derivatives. Thus we write

$$L \simeq \frac{1}{2}M_{ij}\dot{q}_i\dot{q}_j + C_{ij}\dot{q}_i q_j + \frac{1}{2}F_{ij}q_i q_j + A_i\dot{q}_i + B_i q_i + L_0, \quad (3.25)$$

where \mathbf{M} , \mathbf{C} , \mathbf{F} , \mathbf{A} and \mathbf{B} are constant matrices or vectors and the summation convention is in force. Since the EL eqns involve only derivatives of L , we can discard the constant L_0 . It is also easy to check that if \mathbf{C} is symmetric, the terms proportional to \mathbf{C} and \mathbf{A} make no net contribution to the equations of motion. Further, if we assume that $\mathbf{q}_s = 0$, then $\mathbf{B} = 0$. So in this case the EL eqns are

$$M_{ij}\ddot{q}_j = F_{ij}q_j. \quad (3.26)$$

This is easily solved by writing $\mathbf{q}(t) = \mathbf{Q}e^{i\omega t}$, whence the eigenfrequencies ω are the roots of

$$\det(\mathbf{F} + \omega^2\mathbf{M}) = 0. \quad (3.27)$$

Exercise (12):

A cylinder of mass m and radius a rolls on a rough horizontal table. A second cylinder, mass m and radius $\frac{1}{2}a$ rolls inside the first. Find the normal frequencies for small disturbances from equilibrium.

Solution: Let θ be the angle through which the first cylinder has turned from equilibrium, and ϕ be the angle turned by the second cylinder. Then the line between the two centres makes an angle

$$\psi = \frac{1}{2}\phi - \theta \quad (3.28)$$

with the vertical. The kinetic energy of the first cylinder (translational plus rotational) is

$$T_1 = \frac{1}{2}m(a\dot{\theta})^2 + \frac{1}{2}ma^2\dot{\theta}^2. \quad (3.29)$$

The motion of the centre of the second cylinder is a compound of the horizontal motion $a\dot{\theta}$ of the centre of the first cylinder, plus $\frac{1}{2}a\dot{\psi}$ tangent to the line joining the centres. The total kinetic energy is therefore

$$T = m(a\dot{\theta})^2 + \frac{1}{2}m[(a\dot{\theta} + \frac{1}{2}a\dot{\psi} \cos \psi)^2 + (\frac{1}{2}a\dot{\psi} \sin \psi)^2] + \frac{1}{2}m(a/2)^2\dot{\phi}^2. \quad (3.30)$$

The potential energy is simply

$$V = -mg\frac{1}{2}a \cos \psi. \quad (3.31)$$

In T , which is quadratic in the velocities, we set $\psi = 0$ and we expand V to second order in ψ , to find

$$\begin{aligned} T &= \frac{1}{2}ma^2(\frac{9}{4}\dot{\theta}^2 + \frac{1}{4}\dot{\theta}\dot{\phi} + \frac{5}{16}\dot{\phi}^2), \\ V &= \text{constant} + \frac{1}{4}mga(\frac{1}{2}\phi - \theta)^2. \end{aligned} \quad (3.32)$$

Defining $\omega_0 \equiv \sqrt{g/a}$ the equations of motion become

$$\begin{aligned} \frac{9}{2}\ddot{\theta} + \frac{1}{4}\ddot{\phi} - \omega_0^2(\frac{1}{2}\phi - \theta) &= 0, \\ \frac{1}{4}\ddot{\theta} + \frac{5}{8}\ddot{\phi} + \frac{1}{2}\omega_0^2(\frac{1}{2}\phi - \theta) &= 0. \end{aligned} \quad (3.33)$$

The eigenfrequencies are now easily found to be $\omega = 0$ and $\omega = \sqrt{8/11}\omega_0$.

3.6 Noether's theorem

A **constant of motion** is any function $C(\mathbf{q}, \dot{\mathbf{q}})$ that satisfies $dC/dt = 0$, where $\mathbf{q}(t)$ is a solution of the eqns of motion. For example, in a 'conservative' system, energy is conserved, so $E(\mathbf{q}, \dot{\mathbf{q}})$ is a constant of motion. Finding a constant of motion is a big step towards obtaining a general solution of the equations of motion.

In general, a system with n degrees of freedom q_1, \dots, q_n admits $2n - 1$ independent constants of motion. We show this by arguing that given the phase-space position $(\mathbf{q}, \dot{\mathbf{q}})$ at any time t , the equations of motion allow us to give the phase-space position $(\mathbf{q}^{(0)}, \dot{\mathbf{q}}^{(0)})$ at any reference time t_0 . Thus $q_i^{(0)}$ or $\dot{q}_i^{(0)}$ is a function $f_\alpha(\mathbf{q}, \dot{\mathbf{q}}, t)$ with $\alpha = 1, \dots, 2n$. On eliminating t between these $2n$ functions, we have $2n - 1$ constants of motion.

It seldom happens that we can find $2n - 1$ constants of motion—a rare exception is the case of motion in a Kepler potential $V \propto 1/r$. In fact it turns out that essentially complete information about solutions of the equations of motion can be extracted from n constants of motion. A system for which n constants of motion can be found is said to be **integrable**.

A theorem proved by Emmy Noether (1882–1935) provides a powerful way of extracting constants of motion from Lagrangians. Noether's theorem involves identifying a flow in phase space that leaves L invariant. A 'flow' is an infinitesimal transformation $(\mathbf{q} \rightarrow \mathbf{q} + (d\mathbf{q}/d\lambda)\delta\lambda, \dot{\mathbf{q}} \rightarrow \dot{\mathbf{q}} + (d\dot{\mathbf{q}}/d\lambda)\delta\lambda)$. For example, the transformation $\mathbf{x} \rightarrow \mathbf{x} + \hat{\mathbf{i}}\delta\lambda$ with all the velocities remaining the same, is a flow. Invariance of L just means that L takes the same value at all points that are joined by the flow. Noether's theorem states that if L is invariant under $(\mathbf{q} \rightarrow \mathbf{q} + (d\mathbf{q}/d\lambda)\delta\lambda, \dot{\mathbf{q}} \rightarrow \dot{\mathbf{q}} + (d\dot{\mathbf{q}}/d\lambda)\delta\lambda)$, then $(d\mathbf{q}/d\lambda) \cdot (\partial L/\partial \dot{\mathbf{q}})$ is a constant of motion. Thus from the invariance of L under translation $\mathbf{x} \rightarrow \mathbf{x} + \hat{\mathbf{i}}\delta\lambda$ along the x -axis, Noether's theorem deduces the constancy of the x -momentum $\partial L/\partial \dot{x}$, which is equal to $m\dot{x}$ for a particle moving in a velocity-independent potential.

The proof of Noether's theorem is simple. We have

$$0 = \delta L = \frac{\partial L}{\partial \mathbf{q}} \cdot \delta \mathbf{q} + \frac{\partial L}{\partial \dot{\mathbf{q}}} \cdot \delta \dot{\mathbf{q}}.$$

Using the EL eqns to eliminate $\partial L/\partial \dot{\mathbf{q}}$ this becomes

$$\begin{aligned} 0 &= \frac{d}{dt} \left(\frac{\partial L}{\partial \dot{\mathbf{q}}} \right) \cdot \delta \mathbf{q} + \frac{\partial L}{\partial \dot{\mathbf{q}}} \cdot \delta \dot{\mathbf{q}} \\ &= \frac{d}{dt} \left(\frac{\partial L}{\partial \dot{\mathbf{q}}} \cdot \delta \mathbf{q} \right), \end{aligned}$$

and the result follows on writing $\delta\mathbf{q} = (d\mathbf{q}/d\lambda)\delta\lambda$.

Consider the proof of conservation of angular momentum by Noether's theorem. A rotation by $\delta\theta$ about the unit vector $\hat{\mathbf{n}}$ changes \mathbf{x} by $\delta\theta\hat{\mathbf{n}} \times \mathbf{x}$. So if L is invariant under this rotation, the following is a constant of motion:

$$\begin{aligned} J &\equiv \hat{\mathbf{n}} \times \mathbf{x} \cdot \frac{\partial L}{\partial \dot{\mathbf{x}}} \\ &= \hat{\mathbf{n}} \cdot \mathbf{x} \times \frac{\partial L}{\partial \dot{\mathbf{x}}}. \end{aligned} \quad (3.34)$$

For a particle moving in a velocity-independent potential this is just the component of $m\mathbf{x} \times \dot{\mathbf{x}}$ parallel to $\hat{\mathbf{n}}$.

Here's an application to motion in a uniform magnetic field $\mathbf{B} = B\hat{\mathbf{k}}$: By (3.9) the non-relativistic Lagrangian is

$$L = \frac{1}{2}m_0v^2 + q\mathbf{v} \cdot \mathbf{A}. \quad (3.35)$$

Let's choose $\mathbf{A} = (-By, 0, 0)$. Then $L = \frac{1}{2}m_0v^2 - qBv_xy$ is invariant under two flows: (i) $\mathbf{x} \rightarrow \mathbf{x} + \lambda\hat{\mathbf{i}}$ and (ii) $\mathbf{x} \rightarrow \mathbf{x} + \lambda\hat{\mathbf{k}}$. Hence we have two invariants

$$p_x \equiv \frac{\partial L}{\partial \dot{x}} = m_0v_x - qBy \quad ; \quad p_z \equiv \frac{\partial L}{\partial \dot{z}} = m_0v_z. \quad (3.36a)$$

Choosing $\mathbf{A} = (0, Bx, 0)$ we find a third invariant for the same physical problem:

$$p_y \equiv \frac{\partial L}{\partial \dot{y}} = m_0v_y + qBx. \quad (3.36b)$$

The physical meaning of p_z is obvious, but what do p_x and p_y mean physically? Add them up:

$$\begin{aligned} P &\equiv p_x + ip_y = m_0(v_x + iv_y) + qB(ix - y) \\ &= m_0\dot{\xi} + iqB\xi \end{aligned} \quad \text{where } \xi \equiv x + iy. \quad (3.37)$$

Solving this first-order d.e. for ξ we find

$$\xi(t) = \xi(0)e^{-i\omega t} + \frac{iP}{m_0\omega}, \quad \text{where } \omega \equiv \frac{qB}{m_0} \quad (3.38)$$

is the Larmor frequency. It is now easy to see that the real and imaginary parts of P encode the y and x coordinates of the guiding centre around which the particle gyrates.

4 Hamiltonian Dynamics

It turns out that $(\dot{\mathbf{q}}, \mathbf{q})$ are not the ideal coordinates for phase space. The natural coordinates are (\mathbf{p}, \mathbf{q}) , where

$$\mathbf{p} \equiv \frac{\partial L}{\partial \dot{\mathbf{q}}} \quad (4.1)$$

is the **momentum** 'conjugate to \mathbf{q} '. Changing coordinates from $\dot{\mathbf{q}}$ to \mathbf{p} is analogous in thermodynamics to replacing V by P since $P = -(\partial U/\partial V)_S$ just as $\mathbf{p} = (\partial L/\partial \dot{\mathbf{q}})_{\mathbf{q}}$. We are replacing a variable by the gradient of some function of that variable. Transformations of this type are called **Legendre transforms**. When in thermodynamics we eliminate V in favour of P it is expedient to introduce a new function $H(S, P) \equiv U + PV$. So here we introduce the **Hamiltonian**

$$H(\mathbf{p}, \mathbf{q}) \equiv \mathbf{p} \cdot \dot{\mathbf{q}} - L, \quad (4.2)$$

where it is understood that $\dot{\mathbf{q}}$ is to be eliminated in favour of \mathbf{q} , \mathbf{p} , and t using equation (4.1).

The total derivative of the Hamiltonian is

$$\begin{aligned} dH &= \mathbf{p} \cdot d\dot{\mathbf{q}} + \dot{\mathbf{q}} \cdot d\mathbf{p} - \left(\frac{\partial L}{\partial \mathbf{q}} \right)_{\dot{\mathbf{q}}, t} \cdot d\mathbf{q} - \left(\frac{\partial L}{\partial \dot{\mathbf{q}}} \right)_{\mathbf{q}, t} \cdot d\dot{\mathbf{q}} - \left(\frac{\partial L}{\partial t} \right)_{\mathbf{q}, \dot{\mathbf{q}}} dt \\ &= \dot{\mathbf{q}} \cdot d\mathbf{p} - \left(\frac{\partial L}{\partial \mathbf{q}} \right)_{\dot{\mathbf{q}}, t} \cdot d\mathbf{q} - \left(\frac{\partial L}{\partial t} \right)_{\mathbf{q}, \dot{\mathbf{q}}} dt, \end{aligned} \quad (4.3)$$

where the first and fourth terms cancel by (4.1). But we may also write

$$dH = \left(\frac{\partial H}{\partial \mathbf{p}} \right)_{\mathbf{q}, t} \cdot d\mathbf{p} + \left(\frac{\partial H}{\partial \mathbf{q}} \right)_{\mathbf{p}, t} \cdot d\mathbf{q} + \left(\frac{\partial H}{\partial t} \right)_{\mathbf{q}, \mathbf{p}} dt. \quad (4.4)$$

Since equations (4.3) and (4.4) must be the same, we have

$$\dot{\mathbf{q}} = \left(\frac{\partial H}{\partial \mathbf{p}} \right)_{\mathbf{q}, t}; \quad \left(\frac{\partial H}{\partial \mathbf{q}} \right)_{\mathbf{p}, t} = - \left(\frac{\partial L}{\partial \mathbf{q}} \right)_{\dot{\mathbf{q}}, t}; \quad \left(\frac{\partial H}{\partial t} \right)_{\mathbf{q}, \mathbf{p}} = - \left(\frac{\partial L}{\partial t} \right)_{\mathbf{q}, \dot{\mathbf{q}}}. \quad (4.5)$$

Using the EL eqns and simplifying the notation, the first two of these equations lead us to **Hamilton's equations**

$$\dot{\mathbf{q}} = \frac{\partial H}{\partial \mathbf{p}}; \quad \dot{\mathbf{p}} = - \frac{\partial H}{\partial \mathbf{q}}. \quad (4.6)$$

Along a trajectory $\{\mathbf{q}(t), \mathbf{p}(t)\}$, the Hamiltonian $H[\mathbf{q}(t), \mathbf{p}(t), t]$ changes at a rate

$$\frac{dH}{dt} = \frac{\partial H}{\partial \mathbf{q}} \cdot \dot{\mathbf{q}} + \frac{\partial H}{\partial \mathbf{p}} \cdot \dot{\mathbf{p}} + \frac{\partial H}{\partial t} = \frac{\partial H}{\partial t}. \quad (4.7)$$

Hence, if $\partial L/\partial t = 0$, it follows from equation (4.5) that the Hamiltonian is conserved along all dynamical trajectories. We can think of this as an extension of Noether's theorem: the integral H arises from the time-translation invariance of L .

Thus, for example, consider motion in the time-independent potential $V(\mathbf{x})$. If we work in Cartesian coordinates, the Lagrangian $L = \frac{1}{2}m\dot{\mathbf{x}}^2 - V(\mathbf{x})$ depends only on \mathbf{x} and $\dot{\mathbf{x}}$, so $\partial L/\partial t = 0$. Hence the Hamiltonian H is conserved. The physical quantity to which H corresponds is easily found. We have $\mathbf{p} = \partial L/\partial \dot{\mathbf{x}} = m\dot{\mathbf{x}}$ and

$$\begin{aligned} H(\mathbf{x}, \mathbf{p}) &= \mathbf{p} \cdot \dot{\mathbf{x}} - L \\ &= \frac{p^2}{2m} + V(\mathbf{x}), \end{aligned} \quad (4.8)$$

which is simply the total energy $E = \text{k.e.} + \text{p.e.}$. Thus for motion in a fixed potential the Hamiltonian is equal to the total energy.

What are \mathbf{p} and H in a rotating frame? From (4.1) and (3.17) we have

$$\mathbf{p} = m(\dot{\mathbf{r}} + \boldsymbol{\omega} \times \mathbf{r}) \quad (4.9)$$

which shows that \mathbf{p} isn't always the same as $\dot{\mathbf{q}}$. The Hamiltonian for a rotating frame is

$$\begin{aligned} H &= m(\dot{\mathbf{r}} + \boldsymbol{\omega} \times \mathbf{r}) \cdot \dot{\mathbf{r}} - \left[\frac{1}{2}m|\dot{\mathbf{r}}|^2 + m\dot{\mathbf{r}} \cdot (\boldsymbol{\omega} \times \mathbf{r}) + \frac{1}{2}m|\boldsymbol{\omega} \times \mathbf{r}|^2 - V \right] \\ &= \frac{1}{2}m|\dot{\mathbf{r}}|^2 - \frac{1}{2}m|\boldsymbol{\omega} \times \mathbf{r}|^2 + V. \end{aligned} \quad (4.10)$$

This is not the same as the energy in an inertial frame: the kinetic contribution is different and there is an additional centrifugal potential energy $-\frac{1}{2}m|\boldsymbol{\omega} \times \mathbf{r}|^2$. Indeed, unless V is axisymmetric [$V = V(|\boldsymbol{\omega} \times \mathbf{r}|)$], the energy in an inertial frame changes as V does work on the potential, but H is nonetheless constant.

The Lagrangian for non-relativistic motion in an e.m. field is (see (3.9))

$$L = \frac{1}{2}m|\dot{\mathbf{x}}|^2 + q(\dot{\mathbf{x}} \cdot \mathbf{A} - \phi), \quad (4.11)$$

so in this case

$$\begin{aligned} \mathbf{p} &= m\dot{\mathbf{x}} + q\mathbf{A} \\ H &= (m\dot{\mathbf{x}} + q\mathbf{A}) \cdot \dot{\mathbf{x}} - \left[\frac{1}{2}m|\dot{\mathbf{x}}|^2 + q(\dot{\mathbf{x}} \cdot \mathbf{A} - \phi) \right] \\ &= \frac{1}{2}m|\dot{\mathbf{x}}|^2 + q\phi \\ &= \frac{1}{2m}|\mathbf{p} - q\mathbf{A}|^2 + q\phi. \end{aligned} \quad (4.12)$$

Thus in an e.m. field \mathbf{p} is not just $m\dot{\mathbf{x}}$. Although H is just what one would naïvely think of as the energy, when expressed in terms of \mathbf{p} it looks odd. In quantum mechanics the distinction between \mathbf{p} and $\dot{\mathbf{x}}$ is of the utmost importance because it turns out that when one quantizes, it is \mathbf{p} rather than $m\dot{\mathbf{x}}$ that should be replaced by $-i\hbar\nabla$

4.1 Liouville's theorem

If we imagine releasing a bunch of dynamically identical systems from neighbouring initial conditions, then the 'phase points' describing these systems flow through phase space like a fluid. This flow is governed by Hamilton's equations (4.6). It is an incompressible flow: the 'velocity' of the fluid is $(\dot{\mathbf{p}}, \dot{\mathbf{q}})$ and the divergence of this velocity is

$$\begin{aligned} \text{div}(\dot{\mathbf{p}}, \dot{\mathbf{q}}) &= \left(\frac{\partial \dot{\mathbf{p}}}{\partial \mathbf{p}} + \frac{\partial \dot{\mathbf{q}}}{\partial \mathbf{q}} \right) \\ &= \left(-\frac{\partial^2 H}{\partial \mathbf{p} \partial \mathbf{q}} + \frac{\partial^2 H}{\partial \mathbf{q} \partial \mathbf{p}} \right) = 0. \end{aligned}$$

The divergence-freeness of the phase flow is known as **Liouville's theorem**.

Let f be the probability density of systems in phase-space. Then conservation of systems requires that f obey the continuity equation

$$\begin{aligned} 0 &= \frac{\partial f}{\partial t} + \text{div}((\dot{\mathbf{p}}, \dot{\mathbf{q}})f) \\ &= \frac{\partial f}{\partial t} + \frac{\partial f}{\partial \mathbf{p}} \cdot \dot{\mathbf{p}} + \frac{\partial f}{\partial \mathbf{q}} \cdot \dot{\mathbf{q}} \\ &= \frac{\partial f}{\partial t} - \frac{\partial f}{\partial \mathbf{p}} \cdot \frac{\partial H}{\partial \mathbf{q}} + \frac{\partial f}{\partial \mathbf{q}} \cdot \frac{\partial H}{\partial \mathbf{p}} \end{aligned} \quad (4.13)$$

where Liouville's theorem has been used. The continuity equation of f in either of the last two forms is known as **Liouville's equation**.

4.2 Poincaré invariants*

Since Hamilton's equations (4.6) are first-order differential equations, if we are given a particle's phase-space coordinates $(\mathbf{q}_0, \mathbf{p}_0)$ at time $t = 0$, we can solve Hamilton's equations for the coordinates $(\mathbf{q}_t, \mathbf{p}_t)$ at any later time t . Thus through each point $(\mathbf{q}_0, \mathbf{p}_0)$ in phase space there passes a unique phase-space trajectory $[\mathbf{q}(\mathbf{q}_0, \mathbf{p}_0, t), \mathbf{p}(\mathbf{q}_0, \mathbf{p}_0, t)]$, which gives the future and past phase-space coordinates of the particle that at $t = 0$ has coordinates $(\mathbf{q}_0, \mathbf{p}_0)$. We define the **time-evolution operator** \mathbf{H}_t by

$$\mathbf{H}_t(\mathbf{q}_0, \mathbf{p}_0) \equiv [\mathbf{q}(\mathbf{q}_0, \mathbf{p}_0, t), \mathbf{p}(\mathbf{q}_0, \mathbf{p}_0, t)]. \quad (4.14)$$

* Lies beyond the syllabus

We say that the operator \mathbf{H}_t is generated by the function $H(\mathbf{q}, \mathbf{p})$.

Let \mathcal{S}_0 be any two-dimensional surface in phase space, and (u, v) any pair of coordinates that may be used to specify points of \mathcal{S}_0 . The time-evolution operator \mathbf{H}_t maps each point of \mathcal{S}_0 into a new surface \mathcal{S}_t and we denote by (u, v) the point of \mathcal{S}_t into which \mathbf{H}_t maps the point (u, v) of \mathcal{S}_0 . With these definitions, all $2n$ phase-space coordinates q_i and p_i are functions of the three variables u, v , and t .

We define

$$\begin{aligned} A(t) &\equiv \iint_{\mathcal{S}_t} \mathbf{dp} \cdot \mathbf{dq} \equiv \sum_{i=1}^n \iint_{\mathcal{S}_t} dp_i dq_i \\ &= \sum_{i=1}^n \iint_{\mathcal{S}_t} \frac{\partial(p_i, q_i)}{\partial(u, v)} du dv, \end{aligned} \quad (4.15)$$

and calculate dA/dt . We set $t' \equiv t + \delta t$, where δt is small, and introduce the conventions $q(u, v) \equiv q(u, v, t)$, $q'(u, v) \equiv q(u, v, t')$ and similar notations for p and p' . To first order in the small time interval δt , Hamilton's equations (4.6) yield

$$(\mathbf{q}', \mathbf{p}') = \mathbf{H}_{\delta t}(\mathbf{q}, \mathbf{p}) = \left(\mathbf{q} + \frac{\partial H}{\partial \mathbf{p}} \delta t, \mathbf{p} - \frac{\partial H}{\partial \mathbf{q}} \delta t \right). \quad (4.16)$$

Differentiating these equations with respect to u and v , we find

$$\begin{aligned} \frac{\partial(p'_i, q'_i)}{\partial(u, v)} &= \frac{\partial(p_i, q_i)}{\partial(u, v)} - \left[\frac{\partial q_i}{\partial v} \frac{\partial^2 H}{\partial u \partial q_i} - \frac{\partial p_i}{\partial u} \frac{\partial^2 H}{\partial v \partial p_i} + \frac{\partial p_i}{\partial v} \frac{\partial^2 H}{\partial u \partial p_i} - \frac{\partial q_i}{\partial u} \frac{\partial^2 H}{\partial v \partial q_i} \right] \delta t \\ &\quad + O(\delta t)^2. \end{aligned} \quad (4.17)$$

Thus

$$\begin{aligned} \frac{dA}{dt} &= \lim_{\delta t \rightarrow 0} \left\{ \frac{1}{\delta t} \iint du dv \sum_i \left[\frac{\partial(p'_i, q'_i)}{\partial(u, v)} - \frac{\partial(p_i, q_i)}{\partial(u, v)} \right] \right\} \\ &= - \sum_i \iint du dv \left[\frac{\partial q_i}{\partial v} \frac{\partial^2 H}{\partial u \partial q_i} - \frac{\partial p_i}{\partial u} \frac{\partial^2 H}{\partial v \partial p_i} + \frac{\partial p_i}{\partial v} \frac{\partial^2 H}{\partial u \partial p_i} - \frac{\partial q_i}{\partial u} \frac{\partial^2 H}{\partial v \partial q_i} \right]. \end{aligned} \quad (4.18)$$

One may show that the sum of the square brackets in equation (18) vanishes by replacing every occurrence of $\frac{\partial}{\partial u}$ in the second derivatives by $\sum_k \left(\frac{\partial q_k}{\partial u} \frac{\partial}{\partial q_k} + \frac{\partial p_k}{\partial u} \frac{\partial}{\partial p_k} \right)$ and similarly for $\frac{\partial}{\partial v}$. Hence $\dot{A} = 0$ and we have:

Poincaré invariant theorem *If $\mathcal{S}(0)$ is any two-surface in phase space, and $\mathcal{S}(t)$ is the surface into which $\mathcal{S}(0)$ is mapped by the time-evolution operator \mathbf{H}_t , then*

$$\iint_{\mathcal{S}(0)} \mathbf{dp} \cdot \mathbf{dq} = \iint_{\mathcal{S}(t)} \mathbf{dp} \cdot \mathbf{dq}. \quad (4.19)$$

Corollary *If $\gamma(0)$ is any closed path through phase space, and $\gamma(t)$ is the path to which $\gamma(0)$ is mapped by the time-evolution operator, then*

$$\oint_{\gamma(0)} \mathbf{p} \cdot \mathbf{dq} = \oint_{\gamma(t)} \mathbf{p} \cdot \mathbf{dq}. \quad (4.20)$$

Proof. By Green's theorem,

$$\oint_{\gamma(t)} \mathbf{p} \cdot \mathbf{dq} = \sum_i \oint_{\gamma(t)} p_i dq_i = \sum_i \iint_{\mathcal{S}(t)} dp_i dq_i, \quad (4.21)$$

where $\mathcal{S}(t)$ is any surface that has $\gamma(t)$ as its boundary. The result now follows from the Poincaré invariant theorem. \triangleleft

Any mapping of phase space onto itself which, like \mathbf{H}_t , conserves line integrals of the form $\oint \mathbf{p} \cdot \mathbf{dq}$, is called a **canonical map**.

4.3 Poisson brackets and canonical coordinates

Let $A(\mathbf{q}, \mathbf{p})$ and $B(\mathbf{q}, \mathbf{p})$ be any two functions of the phase-space coordinates. Then the **Poisson bracket** $[A, B]$ is defined by

$$[A, B] \equiv \frac{\partial A}{\partial \mathbf{q}} \cdot \frac{\partial B}{\partial \mathbf{p}} - \frac{\partial A}{\partial \mathbf{p}} \cdot \frac{\partial B}{\partial \mathbf{q}}. \quad (4.22)$$

It is straightforward to verify the following properties of Poisson brackets:

- (i) $[A, B] = -[B, A]$ and $[A + B, C] = [A, C] + [B, C]$,
- (ii) $[[A, B], C] + [[B, C], A] + [[C, A], B] = 0$ (**Jacobi identity**),
- (iii) The coordinates (\mathbf{q}, \mathbf{p}) satisfy the **canonical commutation relations**

$$[p_i, p_j] = [q_i, q_j] = 0 \quad \text{and} \quad [q_i, p_j] = \delta_{ij}. \quad (4.23)$$

(iv) Hamilton's equations may be written

$$\dot{q}_i = [q_i, H] \quad ; \quad \dot{p}_i = [p_i, H]. \quad (4.24)$$

If we write $(w_i \equiv q_i, w_{n+i} \equiv p_i \quad i = 1, \dots, n)$, and define the **symplectic matrix** \mathbf{c} by

$$c_{\alpha\beta} \equiv [w_\alpha, w_\beta] = \begin{cases} \pm 1 & \text{for } \beta = \alpha \pm n, 1 \leq \alpha, \beta \leq 2n; \\ 0 & \text{otherwise,} \end{cases} \quad (25a)$$

we have

$$[A, B] = \sum_{\alpha, \beta=1}^{2n} c_{\alpha\beta} \frac{\partial A}{\partial w_\alpha} \frac{\partial B}{\partial w_\beta}. \quad (4.25b)$$

Any set of $2n$ phase-space coordinates $\{W_\alpha, \alpha = 1, \dots, 2n\}$ is called a set of **canonical coordinates** if $[W_\alpha, W_\beta] = c_{\alpha\beta}$. Let $\{W_\alpha\}$ be such a set; then with equation (25b) and the chain rule we have

$$\begin{aligned} [A, B] &= \sum_{\alpha, \beta=1}^{2n} c_{\alpha\beta} \frac{\partial A}{\partial w_\alpha} \frac{\partial B}{\partial w_\beta} = \sum_{\kappa\lambda} \left(\sum_{\alpha\beta} c_{\alpha\beta} \frac{\partial W_\kappa}{\partial w_\alpha} \frac{\partial W_\lambda}{\partial w_\beta} \right) \frac{\partial A}{\partial W_\kappa} \frac{\partial B}{\partial W_\lambda} \\ &= \sum_{\kappa\lambda} [W_\kappa, W_\lambda] \frac{\partial A}{\partial W_\kappa} \frac{\partial B}{\partial W_\lambda} = \sum_{\kappa\lambda} c_{\kappa\lambda} \frac{\partial A}{\partial W_\kappa} \frac{\partial B}{\partial W_\lambda}. \end{aligned} \quad (4.26)$$

Thus the derivatives involved in the definition (4.22) of the Poisson bracket can be taken with respect to any set of canonical coordinates, just as the vector formula $\nabla \cdot \mathbf{a} = \sum_i (\partial a_i / \partial x_i)$ is valid in any Cartesian coordinate system.

The rate of change of an arbitrary canonical coordinate W_α along an orbit is

$$\dot{W}_\alpha = \sum_{\beta=1}^{2n} \frac{\partial W_\alpha}{\partial w_\beta} \dot{w}_\beta, \quad (4.27)$$

where, as usual, $\mathbf{w} \equiv (\mathbf{q}, \mathbf{p})$. With Hamilton's equations (4.24) and equation (4.26) this becomes

$$\begin{aligned} \dot{W}_\alpha &= \sum_{\beta=1}^{2n} \frac{\partial W_\alpha}{\partial w_\beta} [w_\beta, H] = \sum_{\beta\gamma\delta} \frac{\partial W_\alpha}{\partial w_\beta} c_{\gamma\delta} \frac{\partial w_\beta}{\partial w_\gamma} \frac{\partial H}{\partial w_\delta} = \sum_{\gamma\delta} c_{\gamma\delta} \frac{\partial W_\alpha}{\partial w_\gamma} \frac{\partial H}{\partial w_\delta} \\ &= [W_\alpha, H]. \end{aligned} \quad (4.28)$$

<i>Parallelism of Lorentz invariance & Symplectic structure</i>	
inertial coordinates	\leftrightarrow canonical coordinates
Lorentz transformations	\leftrightarrow canonical transformations
$\eta_{\mu\nu}$	\leftrightarrow $c_{\alpha\beta}$
$ \mathbf{x} ^2$	\leftrightarrow $\iint d\mathbf{p} \cdot d\mathbf{q}$

Thus Hamilton's equations (4.24) are valid in *any* canonical coordinate system.

Finally,* we show that equation (4.15) for the Poincaré invariant of a two-surface is valid in any canonical coordinate system. We first note that if we are given any function $B(\mathbf{q}, \mathbf{p})$ we may obtain a one-parameter family of maps \mathbf{B}_a of phase space onto itself by the following procedure. From each point $(\mathbf{q}_0, \mathbf{p}_0)$ of some $(2n - 1)$ -dimensional surface in phase space we integrate the coupled ordinary differential equations

$$\frac{d\mathbf{q}}{db} = [\mathbf{q}, B] \quad , \quad \frac{d\mathbf{p}}{db} = [\mathbf{p}, B] \quad (4.29)$$

from the initial conditions $\mathbf{q}(0) = \mathbf{q}_0, \mathbf{p}(0) = \mathbf{p}_0$. If the initial $(2n - 1)$ -surface is large enough, the **integral curves** $\{\mathbf{q}(b), \mathbf{p}(b)\}$ of B reach every point of phase space. Then the map \mathbf{B}_b is defined by

$$\mathbf{B}_b(\mathbf{q}(b'), \mathbf{p}(b')) = (\mathbf{q}(b + b'), \mathbf{p}(b + b')). \quad (4.30)$$

The function $B(\mathbf{q}, \mathbf{p})$ is indistinguishable from a Hamiltonian, since it satisfies Hamilton's equations (4.29), with b playing the role of the time t . Thus the Poincaré invariant theorem shows that \mathbf{B}_b is a canonical map.

Now let \mathcal{S} be any two-surface in phase space, and U and V two functions on phase space whose integral curves lie within \mathcal{S} . Then we may use the maps \mathbf{U}_u and \mathbf{V}_v to lay out a coordinate grid on \mathcal{S} : we pick a point $(\mathbf{q}_0, \mathbf{p}_0)$ of \mathcal{S} and define $(\mathbf{q}, \mathbf{p})_{(u,v)} \equiv \mathbf{V}_v \mathbf{U}_u(\mathbf{q}_0, \mathbf{p}_0)$. The Poincaré invariant (4.15) of \mathcal{S} may now be written

$$\begin{aligned} A &= \iint_{\mathcal{S}} d\mathbf{p} \cdot d\mathbf{q} = \iint \sum_i \frac{\partial(p_i, q_i)}{\partial(u, v)} du dv \\ &= \iint \sum_i ([p_i, U][q_i, V] - [p_i, V][q_i, U]) du dv, \end{aligned} \quad (4.31)$$

where the last equality follows by expanding the Jacobian and using equation (4.29). But by equation (4.22), $[p_i, U] = -(\partial U / \partial q_i)$ and $[q_i, V] = (\partial V / \partial p_i)$, so we may rewrite equation (31) as

$$A = - \iint \sum_i \left(\frac{\partial U}{\partial q_i} \frac{\partial V}{\partial p_i} - \frac{\partial V}{\partial q_i} \frac{\partial U}{\partial p_i} \right) du dv = - \iint [U, V] du dv. \quad (4.32)$$

Now notice that the Poisson bracket $[U, V]$ is the same when evaluated in terms of derivatives with respect to any set of canonical coordinates. For the same reason, the map between (u, v) and the surface \mathcal{S} is independent of the canonical coordinates used in equation (4.29). Hence we may change to a new set of canonical coordinates (\mathbf{Q}, \mathbf{P}) , retrace each step in the derivation of equation (32), and thus show that

$$\iint_{\mathcal{S}} d\mathbf{p} \cdot d\mathbf{q} = \iint_{\mathcal{S}} d\mathbf{P} \cdot d\mathbf{Q} \quad \text{for any } \mathcal{S}. \quad (4.33)$$

* The rest of this section lies beyond the syllabus.

4.4 Canonical transformations

Suppose you have a function $S(\mathbf{P}, \mathbf{q})$ of some new variables P_i , $i = 1, n$ and the regular coordinates q_i such that the equation

$$\mathbf{p} = \frac{\partial S}{\partial \mathbf{q}} \quad (4.34a)$$

can be interpreted as defining $\mathbf{P}(\mathbf{p}, \mathbf{q})$. Then the coordinates (\mathbf{P}, \mathbf{Q}) are canonical, where

$$\mathbf{Q} \equiv \frac{\partial S}{\partial \mathbf{P}}. \quad (4.34b)$$

That is, $[Q_i, Q_j] = 0$, $[Q_i, P_j] = \delta_{ij}$, $[P_i, P_j] = 0$. The transformation $(\mathbf{p}, \mathbf{q}) \rightarrow (\mathbf{P}, \mathbf{Q})$ is called a **canonical transformation** and S the **generating function** of the transformation.

Before we prove that $[Q_i, P_j] = \delta_{ij}$ etc., we demonstrate the relationship of these canonical transformations to the canonical maps introduced earlier. We do this by considering functions S of the form

$$S = \mathbf{P} \cdot \mathbf{q} + s(\mathbf{P}, \mathbf{q})\delta u, \quad (4.35)$$

where $\delta u \ll 1$. For S of this form we have

$$\begin{aligned} \mathbf{Q} &= \mathbf{q} + \frac{\partial s}{\partial \mathbf{P}}\delta u \quad ; \quad \mathbf{p} = \mathbf{P} + \frac{\partial s}{\partial \mathbf{q}}\delta u \quad \Rightarrow \\ \mathbf{P} &= \mathbf{p} - \frac{\partial s}{\partial \mathbf{q}}\delta u. \end{aligned} \quad (4.36)$$

Thus $S = \mathbf{P} \cdot \mathbf{q}$ generates the identity transformation $\mathbf{P} = \mathbf{p}$, $\mathbf{Q} = \mathbf{q}$. Moreover,

$$\begin{aligned} \frac{\mathbf{Q} - \mathbf{q}}{\delta u} &= \frac{\partial s}{\partial \mathbf{P}} \\ \frac{\mathbf{P} - \mathbf{p}}{\delta u} &= -\frac{\partial s}{\partial \mathbf{q}} \end{aligned} \quad (4.37)$$

In the limit $\delta u \rightarrow 0$ we can identify \mathbf{P} with \mathbf{p} on the right, and these equations become

$$\frac{d\mathbf{q}}{du} = [\mathbf{q}, s] \quad ; \quad \frac{d\mathbf{p}}{du} = [\mathbf{p}, s], \quad (4.38)$$

which is identical with (4.29). Thus canonical transformations generated by functions of the form (4.35) may be thought of as infinitesimal canonical maps.

There is no fundamental difference between a map and a coordinate transformation: every map generates a coordinate transformation and every transformation a map since one can treat changed coordinates as new numbers describing an old point (a coordinate change), or as old numbers describing a new point (a mapping).

We now prove* that \mathbf{P} and \mathbf{Q} satisfy the canonical commutation relations. From the chain rule we have that

$$\begin{aligned} \left. \frac{\partial}{\partial \mathbf{p}} \right|_{\mathbf{q}} &= \left(\left. \frac{\partial \mathbf{P}}{\partial \mathbf{p}} \right|_{\mathbf{q}} \cdot \left. \frac{\partial}{\partial \mathbf{P}} \right|_{\mathbf{q}} \right) \\ \left. \frac{\partial}{\partial \mathbf{q}} \right|_{\mathbf{p}} &= \left. \frac{\partial}{\partial \mathbf{q}} \right|_{\mathbf{P}} + \left(\left. \frac{\partial \mathbf{P}}{\partial \mathbf{q}} \right|_{\mathbf{p}} \cdot \left. \frac{\partial}{\partial \mathbf{P}} \right|_{\mathbf{q}} \right). \end{aligned} \quad (4.39)$$

* The rest of this section lies beyond the syllabus.

Applying these formulae to p_i and using $\partial p_i/\partial \mathbf{P} = \partial^2 S/\partial q_i \partial \mathbf{P} = \partial \mathbf{Q}/\partial q_i$ yields

$$\begin{aligned} \delta_{ij} &= \left(\frac{\partial \mathbf{P}}{\partial p_j} \right)_{\mathbf{q}} \cdot \left(\frac{\partial \mathbf{Q}}{\partial q_i} \right)_{\mathbf{q}} \\ - \left(\frac{\partial p_i}{\partial q_j} \right)_{\mathbf{P}} &= \left(\frac{\partial \mathbf{P}}{\partial q_j} \right)_{\mathbf{p}} \cdot \left(\frac{\partial \mathbf{Q}}{\partial q_i} \right)_{\mathbf{q}} \end{aligned} \quad (4.40)$$

Multiplying these equations together and summing over j we find

$$\begin{aligned} \sum_{kl} \left(\frac{\partial Q_k}{\partial q_i} \right)_{\mathbf{q}} \left(\frac{\partial Q_l}{\partial q_{i'}} \right)_{\mathbf{q}} [P_k, P_l] &= - \left(\frac{\partial p_i}{\partial q_{i'}} \right)_{\mathbf{P}} + \left(\frac{\partial p_{i'}}{\partial q_i} \right)_{\mathbf{P}} \\ &= - \frac{\partial^2 S}{\partial q_{i'} \partial q_i} + \frac{\partial^2 S}{\partial q_i \partial q_{i'}} = 0. \end{aligned} \quad (4.41)$$

Since the matrix $\partial Q_k/\partial q_i$ has an inverse by (4.40), this shows that $[P_k, P_l] = 0$.

Working again from equations (4.39) we have

$$\begin{aligned} [Q_i, P_j] &= \left(\frac{\partial Q_i}{\partial \mathbf{q}} \right)_{\mathbf{p}} \cdot \left(\frac{\partial P_j}{\partial \mathbf{p}} \right)_{\mathbf{q}} - \left(\frac{\partial Q_i}{\partial \mathbf{p}} \right)_{\mathbf{q}} \cdot \left(\frac{\partial P_j}{\partial \mathbf{q}} \right)_{\mathbf{p}} \\ &= \left[\left(\frac{\partial Q_i}{\partial \mathbf{q}} \right)_{\mathbf{P}} + \left(\frac{\partial Q_i}{\partial \mathbf{P}} \right)_{\mathbf{q}} \cdot \left(\frac{\partial \mathbf{P}}{\partial \mathbf{q}} \right)_{\mathbf{p}} \right] \cdot \left(\frac{\partial P_j}{\partial \mathbf{p}} \right)_{\mathbf{q}} \\ &\quad - \left(\frac{\partial Q_i}{\partial \mathbf{P}} \right)_{\mathbf{q}} \cdot \left(\frac{\partial \mathbf{P}}{\partial \mathbf{p}} \right)_{\mathbf{q}} \cdot \left(\frac{\partial P_j}{\partial \mathbf{q}} \right)_{\mathbf{p}} \\ &= \left(\frac{\partial Q_i}{\partial \mathbf{q}} \right)_{\mathbf{P}} \cdot \left(\frac{\partial P_j}{\partial \mathbf{p}} \right)_{\mathbf{q}} + \left(\frac{\partial Q_i}{\partial \mathbf{P}} \right)_{\mathbf{q}} \cdot [\mathbf{P}, P_j] \\ &= \frac{\partial^2 S}{\partial P_i \partial \mathbf{q}} \cdot \left(\frac{\partial P_j}{\partial \mathbf{p}} \right)_{\mathbf{q}} \\ &= \left(\frac{\partial \mathbf{p}}{\partial P_i} \right)_{\mathbf{q}} \cdot \left(\frac{\partial P_j}{\partial \mathbf{p}} \right)_{\mathbf{q}} = \delta_{ij}. \end{aligned} \quad (4.42)$$

Similarly,

$$\begin{aligned} [Q_i, Q_j] &= \left[\left(\frac{\partial Q_i}{\partial \mathbf{q}} \right)_{\mathbf{P}} + \left(\frac{\partial Q_i}{\partial \mathbf{P}} \right)_{\mathbf{q}} \cdot \left(\frac{\partial \mathbf{P}}{\partial \mathbf{q}} \right)_{\mathbf{p}} \right] \cdot \left(\frac{\partial Q_j}{\partial \mathbf{p}} \right)_{\mathbf{q}} \\ &\quad - \left(\frac{\partial Q_i}{\partial \mathbf{P}} \right)_{\mathbf{q}} \cdot \left(\frac{\partial \mathbf{P}}{\partial \mathbf{p}} \right)_{\mathbf{q}} \cdot \left(\frac{\partial Q_j}{\partial \mathbf{q}} \right)_{\mathbf{p}} \\ &= \left(\frac{\partial Q_i}{\partial \mathbf{q}} \right)_{\mathbf{P}} \cdot \left(\frac{\partial Q_j}{\partial \mathbf{p}} \right)_{\mathbf{q}} + \left(\frac{\partial Q_i}{\partial \mathbf{P}} \right)_{\mathbf{q}} \cdot [\mathbf{P}, Q_j] \\ &= \left(\frac{\partial Q_i}{\partial \mathbf{q}} \right)_{\mathbf{P}} \cdot \left(\frac{\partial Q_j}{\partial \mathbf{p}} \right)_{\mathbf{q}} - \left(\frac{\partial Q_i}{\partial P_j} \right)_{\mathbf{q}} \\ &= \frac{\partial^2 S}{\partial P_i \partial q_k} \left(\frac{\partial Q_j}{\partial \mathbf{P}} \right)_{\mathbf{q}} \cdot \left(\frac{\partial \mathbf{P}}{\partial p_k} \right)_{\mathbf{q}} - \left(\frac{\partial Q_i}{\partial P_j} \right)_{\mathbf{q}}. \end{aligned}$$

But $\left(\frac{\partial p_k}{\partial \mathbf{P}} \right)_{\mathbf{q}} = \frac{\partial^2 S}{\partial q_k \partial \mathbf{P}}$, so

$$\begin{aligned} [Q_i, Q_j] &= \left(\frac{\partial Q_j}{\partial P_l} \right)_{\mathbf{q}} \left(\frac{\partial P_l}{\partial p_k} \right)_{\mathbf{q}} \left(\frac{\partial p_k}{\partial P_i} \right)_{\mathbf{q}} - \left(\frac{\partial Q_i}{\partial P_j} \right)_{\mathbf{q}} \\ &= \frac{\partial Q_j}{\partial P_i} - \frac{\partial Q_i}{\partial P_j} \\ &= \frac{\partial^2 S}{\partial P_i \partial P_j} - \frac{\partial^2 S}{\partial P_j \partial P_i} = 0. \end{aligned} \quad (4.43)$$

The function that generates a canonical transformation need not be of the form $S(\mathbf{P}, \mathbf{q})$; other forms are $S(\mathbf{P}, \mathbf{p})$, $S(\mathbf{Q}, \mathbf{q})$ and $S(\mathbf{Q}, \mathbf{p})$. The generating function is always a function of one old coordinate and one new one. An entertaining transformation is generated by $S = \mathbf{Q} \cdot \mathbf{q}$:

$$\mathbf{p} = \frac{\partial S}{\partial \mathbf{q}} = \mathbf{Q} \quad ; \quad \mathbf{P} = \frac{\partial S}{\partial \mathbf{Q}} = \mathbf{q}. \quad (4.44)$$

4.5 Point transformations

If $(Q_i(\mathbf{q}), i = 1, \dots, n)$ are any three independent functions of the generalized coordinates \mathbf{q} , then by equation (4.1) we obtain the new momenta $P_i = (\partial L / \partial \dot{Q}_i)$ by expressing the Lagrangian as a function $L(\mathbf{Q}, \dot{\mathbf{Q}})$ of the Q_i and their time derivatives. The coordinate change $(\mathbf{q}, \mathbf{p}) \rightarrow (\mathbf{Q}, \mathbf{P})$ is called a **point transformation**, because the new coordinates are functions only of the old. It is straightforward to show that the new coordinates are canonical, by evaluating their Poisson brackets.

The importance of these results is that it is often convenient to work in curvilinear coordinates \mathbf{Q} and derive the corresponding momenta $\mathbf{P} = (\partial L / \partial \dot{\mathbf{Q}})$. Since the coordinates (\mathbf{Q}, \mathbf{P}) are canonical, the Poisson bracket (4.22) can be equally well evaluated by taking derivatives with respect to \mathbf{Q} and \mathbf{P} as with respect to \mathbf{q} and \mathbf{p} . Hence all curvilinear coordinates have equal status in Hamiltonian mechanics.

4.6 Phase-space volumes*

Often, for example when doing statistical mechanics, one needs a credible definition of ‘phase-space volume’. If one is using Cartesian coordinates to describe a system of N particles of mass m_i , it is natural to take the volume element to be $d\tau = \prod_i^N (m_i^3 d^3 \mathbf{x}_i d^3 \mathbf{v}_i)$. But it isn’t immediately obvious what to use for $d\tau$ in a more complex case. In particular, if one decided to describe the system of particles by some curvilinear coordinates $\mathbf{q}(\mathbf{x})$ and their conjugate momenta \mathbf{p} , one would expect $d\tau$ to be of the form

$$d\tau = \prod_i \left(\frac{\partial(\mathbf{v}_i, \mathbf{x}_i)}{\partial(\mathbf{p}_i, \mathbf{q}_i)} d^3 \mathbf{p} d^3 \mathbf{q} \right). \quad (4.45)$$

One of the most beautiful and useful results in the subject is that the Jacobian here is just one. In fact, the Jacobian between *any* pair of canonical coordinates is always one. That is, the volume of an arbitrary region is

$$V = \iiint_{\mathcal{V}} d^n \mathbf{p} d^n \mathbf{q} = \iiint_{\mathcal{V}} d^n \mathbf{P} d^n \mathbf{Q}, \quad (4.46)$$

where (\mathbf{p}, \mathbf{q}) and (\mathbf{P}, \mathbf{Q}) are *any* canonical coordinates.

It is possible to prove this result by brute-force calculation of the Jacobian. But it is more elegant and convenient to stand back and look for a coordinate-free definition of phase-space volume and to show that in coordinate form this is $d\tau = d^n \mathbf{p} d^n \mathbf{q}$ no matter what the coordinate system. The exercise then is to define the volume of an arbitrary parallelepiped in phase space without reference to any system of coordinates.

We start by finding $2n$ functions on phase space $U_1(\mathbf{p}, \mathbf{q}), \dots, U_{2n}(\mathbf{p}, \mathbf{q})$ whose integral curves (i.e., the solutions of equations (4.29) with $B = U_i, b = u_i$) run along the edges of the parallelepiped. We denote by u_i the parameter b that measures displacement along the integral curve generated by the function U_i through equations (4.29). Equation (4.32) for the Poincaré invariant of the parallelogram bounded by the invariant curves of U and V suggests that we try

$$d\tau \equiv \frac{(-1)^n}{2^n n!} \sum (-1)^\nu [U_{i_1}, U_{i_2}] \times \dots \times [U_{i_{2n-1}}, U_{i_{2n}}] du_1 \times \dots \times du_{2n}, \quad (4.47)$$

* Lies beyond the syllabus

where the sum is over all permutations (i_1, \dots, i_{2n}) of the numbers $(1, \dots, 2n)$, and ν is 0 or 1 according as the permutation is even or odd. Dimensionally at least this definition is correct: reference to (4.32) shows that it has the dimensions of the product of n Poincaré invariants, i.e., a volume. To understand the origin of the numerical prefactor and the role of the factor $(-1)^\nu$, let's evaluate (4.47) for the special case in which the parallelepiped has edges parallel to the axes of the (\mathbf{p}, \mathbf{q}) system.

A function U_i whose integral curves run parallel to the q_i axis, is $U_i = p_i$, and the corresponding parameter is $u_i = q_i$; if we use $B = p_i$ in (4.32), we find $dq_j/db = \delta_{ij}$. Similarly, the function $U_{n+i} = q_i$ has integral curves parallel to the p_i axis, with parameter $u_{n+i} = -p_i$. When we use these choices for the U_i in (4.47) we find that the product of Poisson brackets fails to vanish only when each p is bracketed with its own q , and then the product equals ± 1 according as (i_1, \dots, i_{2n}) is an even or odd permutation of the numbers $(1, \dots, 2n)$. There are $n!$ ways in which the pairs of p 's and q 's can be arranged, and each arrangement gives rise to 2^n non-zero terms since for each term with $[p_i, q_i]$ there is one with $[q_i, p_i]$. So all these permutations add up to a factor that is cancelled by the prefactor $1/2^n n!$. We are then left with the product of the differentials in the parameters $du_i = dq_i$, $du_{n+i} = -dp_i$. Hence for a parallelepiped with edges parallel to the coordinate axes, $d\tau = d^n \mathbf{p} d^n \mathbf{q}$. Since every volume can be decomposed into such special parallelepipeds for any coordinate system, (4.46) now follows.

5 Relating quantum to classical mechanics

5.1 Phase-space operators

The Poisson bracket (4.22) turns every function F on phase space into an operator \hat{F} on other functions of phase space. Thus given $F(\mathbf{p}, \mathbf{q})$ we define the operator \hat{F} by its action on an arbitrary function $\psi(\mathbf{p}, \mathbf{q})$:

$$\hat{F}\psi \equiv -i\hbar[\psi, F]. \quad (5.1)$$

Here \hbar is some constant with the dimensions of a Poincaré invariant $\mathbf{p} \cdot \mathbf{q}$ —i.e., the inverse of the dimensions of the Poisson bracket. It is interesting to express the operators associated with p_x , x , $H = \frac{1}{2}p^2/m + V(\mathbf{x})$ and L_z as differential operators:

$$\begin{aligned} \hat{p}_x &= -i\hbar[\cdot, p_x] = -i\hbar \frac{\partial}{\partial x} \\ \hat{x} &= -i\hbar[\cdot, x] = i\hbar \frac{\partial}{\partial p_x} \\ \hat{H} &= -i\hbar[\cdot, H] = -i\hbar \left(\frac{\mathbf{p}}{m} \cdot \nabla - \nabla V \cdot \frac{\partial}{\partial \mathbf{p}} \right) \\ \hat{L}_z &= -i\hbar[\cdot, L_z] = -i\hbar[\cdot, xp_y - yp_x] \\ &= -i\hbar \left(x \frac{\partial}{\partial y} - y \frac{\partial}{\partial x} + p_x \frac{\partial}{\partial p_y} - p_y \frac{\partial}{\partial p_x} \right). \end{aligned} \quad (5.2)$$

Notice that $\widehat{(p^2)} \neq (\hat{p})^2$. With the obvious definition of an inner product, these operators are self-adjoint (Hermitian):

$$\begin{aligned} \int d^n \mathbf{p} d^n \mathbf{q} \phi^* \hat{A} \psi &= -i\hbar \left(\int d^n \mathbf{p} d^n \mathbf{q} \phi^* \frac{\partial \psi}{\partial \mathbf{q}} \cdot \frac{\partial A}{\partial \mathbf{p}} - \int d^n \mathbf{q} d^n \mathbf{p} \phi^* \frac{\partial \psi}{\partial \mathbf{p}} \cdot \frac{\partial A}{\partial \mathbf{q}} \right) \\ &= i\hbar \left(\int d^n \mathbf{p} d^n \mathbf{q} \frac{\partial \phi^*}{\partial \mathbf{q}} \cdot \frac{\partial A}{\partial \mathbf{p}} \psi - \int d^n \mathbf{q} d^n \mathbf{p} \frac{\partial \phi^*}{\partial \mathbf{p}} \cdot \frac{\partial A}{\partial \mathbf{q}} \psi \right) \\ &= \int d^n \mathbf{p} d^n \mathbf{q} (\hat{A} \phi)^* \psi. \end{aligned} \quad (5.3)$$

We can use the Jacobi identity to express the commutator of two of these operators in terms of the Poisson bracket of the underlying functions:

$$\begin{aligned}(\hat{f}\hat{g} - \hat{g}\hat{f})\psi &= -\hbar^2 ([[\psi, g], f] - [[\psi, f], g]) \\ &= \hbar^2 [\psi, [f, g]] \\ &= i\hbar \widehat{[f, g]}\psi.\end{aligned}\tag{5.4}$$

For example

$$\begin{aligned}\hat{x}\hat{p}_x - \hat{p}_x\hat{x} &= i\hbar\hat{1} \\ \hat{L}_x\hat{L}_y - \hat{L}_y\hat{L}_x &= i\hbar\hat{L}_z.\end{aligned}\tag{5.5}$$

Here the operator $\hat{1}$ is not the identity operator but the operator to which the function 1 gives rise—this operator annihilates all functions: $\hat{1}\psi = 0 \forall \psi$.

Let $A(\mathbf{p}, \mathbf{q})$ be some function on phase space. Then the rate of change of the value of A along a phase trajectory is

$$\begin{aligned}\frac{dA}{dt} &= \frac{\partial A}{\partial \mathbf{p}} \cdot \dot{\mathbf{p}} + \frac{\partial A}{\partial \mathbf{q}} \cdot \dot{\mathbf{q}} \\ &= [A, H].\end{aligned}\tag{5.6}$$

Consequently A is a constant of motion if it commutes with the Hamiltonian, i.e., if we have $\hat{A}\hat{H} - \hat{H}\hat{A} = 0$.

Hamilton's equations (4.24) allow Liouville's equation (4.13) to be written

$$\begin{aligned}i\hbar \frac{\partial f}{\partial t} &= -i\hbar [f, H] \\ &= \hat{H}f.\end{aligned}\tag{5.7}$$

The similarity of these entirely classical operators to the usual quantum-mechanical ones is obvious. Where they differ from the usual q.m. operators is in that they don't operate on simple functions on configuration space: even if ψ happens to be a function $\psi(\mathbf{x})$ of \mathbf{x} only, $\hat{H}\psi = -(i\hbar\mathbf{p}/m) \cdot \nabla\psi$ does depend on \mathbf{p} .

The 'canonical' procedure for quantizing a classical theory is to set up this Hamiltonian formalism in phase space, discover what commutation relations hold between the operators \hat{H} , \hat{F} , \hat{A} of the 'observables' H , F , A etc., and then to represent them as operators on functions on just half a set of canonical coordinates, that is functions $\psi(\mathbf{q})$ or $\psi(\mathbf{p})$ or $\psi(\mathbf{Q})$...

5.2 Hamilton-Jacobi Equation*

Suppose we could find n constants of motion I_1, \dots, I_n . And suppose it were possible to find a system of canonical coordinates (\mathbf{P}, \mathbf{Q}) such that $P_i = I_i$ etc. Then the equations of motion for the P 's would be trivial,

$$\begin{aligned}0 &= \dot{P}_i = [P_i, H] \\ &= -\frac{\partial H}{\partial Q_i}.\end{aligned}\tag{5.8}$$

and would demonstrate that $H(\mathbf{P})$ would be independent of the Q 's. This last observation would allow us to solve the equations of motion for the Q 's: we would have

$$\dot{Q}_i = \frac{\partial H}{\partial P_i} \equiv \omega_i, \quad \text{a constant} \quad \Rightarrow \quad Q_i(t) = Q_i(0) + \omega_i t.\tag{5.9}$$

* Lies beyond the syllabus

So everything would lie at our feet if we could find n constants of the motion and could embed these as the ‘momenta’ of a system of canonical coordinates.² The magic coordinates $\mathbf{P} \equiv \mathbf{I}$ and \mathbf{Q} are called **action-angle** coordinates, the I ’s being the actions and the Q ’s the angles.

Let $S(\mathbf{I}, \mathbf{q})$ be the generating function of the transformation between regular coordinates (\mathbf{p}, \mathbf{q}) and angle-action coordinates. Then we can use this to eliminate $\mathbf{p} = \partial S / \partial \mathbf{q}$ from H , expressing H as a function of (\mathbf{I}, \mathbf{q}) :

$$H(\mathbf{I}, \mathbf{q}) = H\left(\frac{\partial S}{\partial \mathbf{q}}, \mathbf{q}\right). \quad (5.10)$$

By moving on an orbit we can vary the q_i pretty much at will while holding constant the I_i . As we vary the q_i in this way H must remain constant at the energy E of the orbit in question. This suggests that we investigate the non-linear partial differential equation

$$H\left(\frac{\partial S}{\partial \mathbf{q}}, \mathbf{q}\right) = E, \quad (\text{Hamilton-Jacobi equation}). \quad (5.11)$$

If we can solve this equation, we identify the arbitrary constants on which the solution $S(\mathbf{q})$ depends with functions of the constants of motion I_i . For example, the H-J eqn for a free particle moving in two dimensions is

$$\frac{|\nabla S|^2}{2m} = E \quad (5.12)$$

We write $S(\mathbf{x}) = S_x(x) + S_y(y)$ and solve (5.12) by separation of variables:

$$\text{constant} \equiv I_x = \left(\frac{\partial S}{\partial x}\right)^2 = 2mE - \left(\frac{\partial S}{\partial y}\right)^2 \equiv I_y. \quad (5.13)$$

This example is very tame, but the technique works also for more complicated Hamiltonians that cannot be solved by other means.

The similarity between the H-J eqn and the time-independent Schrödinger eqn is obvious.

5.3 Path integrals*

The principle of least action (3.15) is concerned with paths $\mathbf{q}(t)$ through coordinate space. We can derive classical mechanics from another, closely related, variational principle which involves paths $(\mathbf{p}(t), \mathbf{q}(t))$ through phase space rather than coordinate space. This principle is that the path actually followed between (t_i, q_i) and (t_f, q_f) is that for which

$$\delta S = 0 \quad \text{where} \quad S \equiv \int \mathbf{p} \cdot d\mathbf{q} - H(\mathbf{p}, \mathbf{q}) dt. \quad (5.14)$$

Here the path of integration runs between (t_i, q_i) and (t_f, q_f) . Showing that this principle yields Hamilton’s equations (4.6) is easy:

$$\begin{aligned} \delta S &= \int \left(\delta \mathbf{p} \cdot \dot{\mathbf{q}} + \mathbf{p} \cdot \delta \dot{\mathbf{q}} - \frac{\partial H}{\partial \mathbf{p}} \cdot \delta \mathbf{p} - \frac{\partial H}{\partial \mathbf{q}} \cdot \delta \mathbf{q} \right) dt \\ &= \int \left[\left(\dot{\mathbf{q}} - \frac{\partial H}{\partial \mathbf{p}} \right) \cdot \delta \mathbf{p} - \left(\dot{\mathbf{p}} + \frac{\partial H}{\partial \mathbf{q}} \right) \cdot \delta \mathbf{q} \right] dt + [\mathbf{p} \cdot \delta \mathbf{q}]_{t_i}^{t_f}. \end{aligned} \quad (5.15)$$

Since $\delta \mathbf{q}$ vanishes at t_i and t_f by hypothesis, the final term in (5.15) vanishes. Then, with $\delta \mathbf{p}$ and $\delta \mathbf{q}$ subject to arbitrary variation, it is clear that $\delta S = 0$ only if the contents of the pairs

² Notice that to be able to embed the I ’s as a set of momenta, we require $[I_i, I_j] = 0$; functions satisfying this condition are said to be ‘in involution’.

* Lies beyond the syllabus

of large round brackets in (5.15) vanish. But the vanishing of brackets is precisely the content of Hamilton's equations.

Notice that a very remarkable thing is being done with the variational principle (5.14): we are treating \mathbf{p} as quite independent of the value of $\dot{\mathbf{q}}$ along the path. This makes perfectly good sense from the point of view of phase-space geometry, but it makes a mockery of our original definition (4.1) of \mathbf{p} . This definition is recovered for the true path as a consequence of the variational principle (5.14):

$$\begin{aligned}\dot{\mathbf{q}} &= \frac{\partial H}{\partial \mathbf{p}} = \frac{\partial}{\partial \mathbf{p}}(\mathbf{p} \cdot \dot{\mathbf{q}} - L) \\ &= \dot{\mathbf{q}} + \left(\mathbf{p} - \frac{\partial L}{\partial \dot{\mathbf{q}}} \right) \cdot \frac{\partial \dot{\mathbf{q}}}{\partial \mathbf{p}}.\end{aligned}$$

Recall that we introduced H as $\mathbf{p} \cdot \dot{\mathbf{q}} - L$, with $\dot{\mathbf{q}}$ eliminated in favour of \mathbf{p} . Now that we are treating \mathbf{p} as independent of $\dot{\mathbf{q}}$, $\mathbf{p} \cdot \dot{\mathbf{q}} - H$ becomes a quantity different from L ; indeed, L depends only on the projection of a phase-space path $(\mathbf{p}(t), \mathbf{q}(t))$ onto configuration space, while $\mathbf{p} \cdot \dot{\mathbf{q}} - H$ depends on $\mathbf{p}(t)$ as well as $\mathbf{q}(t)$. Thus the action principle (5.14) is entirely different from (3.15), although the extremal values of the two integrals are the same because along the extremal path $\mathbf{p} = \partial L / \partial \dot{\mathbf{q}}$.

It is interesting to derive (5.14) from quantum mechanics. The basic idea is simple: we calculate the quantum amplitude to get from (t_i, \mathbf{q}_i) to (t_f, \mathbf{q}_f) and show that it can be expressed as a sum over all possible paths between these events of amplitudes proportional to $e^{iS/\hbar}$, where S is defined by (5.14). The only paths that make a net contribution to the overall amplitude are those whose values of S lie within $\sim \hbar$ of a stationary value, since the contributions of other paths are cancelled by oppositely signed contributions from neighbouring paths. Thus the overall amplitude is dominated by contributions from paths that lie within $\sim \hbar$ of the classical, extremizing, path, and from a macroscopic point of view these paths are identical with the classical path.

We start by finding the amplitude A_{12} to get from (t_1, \mathbf{q}_1) to (t_2, \mathbf{q}_2) , where the interval $t_2 - t_1$ is small. In Dirac's notation, this amplitude is

$$A_{12} = \langle \mathbf{q}_2 | \psi, t_2 \rangle, \quad (5.16)$$

where $|\psi, t_2\rangle$ is the ket into which $|\mathbf{q}_1\rangle$ has evolved at t_2 . In other words, $|\psi, t_2\rangle$ is the solution of the time-dependent Schrödinger equation (TDSE) for initial condition $|\psi, t_1\rangle = |\mathbf{q}_1\rangle$. This is

$$|\psi, t_2\rangle = e^{-i\hat{H}(t_2-t_1)/\hbar} |\mathbf{q}_1\rangle. \quad (5.17)$$

Here the exponential is the operator with the same eigen-kets $|E_n\rangle$ as the Hamiltonian \hat{H} , and eigenvalues equal to $e^{iE_n(t_2-t_1)/\hbar}$, where the E_n are the eigen-values of \hat{H} . That is,

$$e^{i\hat{H}(t_2-t_1)/\hbar} \equiv \sum_n |E_n\rangle e^{-iE_n(t_2-t_1)/\hbar} \langle E_n|. \quad (5.18)$$

(To prove that (5.17) satisfies the TDSE, just substitute (5.18) into (5.17) and differentiate w.r.t. t_2 .) Our amplitude can now be written

$$\begin{aligned}A_{12} &= \langle \mathbf{q}_2 | e^{-i\hat{H}(t_2-t_1)/\hbar} | \mathbf{q}_1 \rangle \\ &= \int d^3\mathbf{p} \langle \mathbf{q}_2 | \mathbf{p} \rangle \langle \mathbf{p} | e^{-i\hat{H}(t_2-t_1)/\hbar} | \mathbf{q}_1 \rangle,\end{aligned} \quad (5.19)$$

where use has been made of the fact that $\int d^3\mathbf{p} |\mathbf{p}\rangle \langle \mathbf{p}|$ is just the identity operator since the states $|\mathbf{p}\rangle$ of well-defined momentum form a complete set.

\hat{H} and thus the function of it appearing in (5.19) is a function of the operators $\hat{\mathbf{p}}$ and $\hat{\mathbf{q}}$. Let's assume that every $\hat{\mathbf{p}}$ has been positioned to the left of every $\hat{\mathbf{q}}$. Then every $\hat{\mathbf{p}}$ can be considered to act to the left and be replaced by its eigen-value \mathbf{p} , while every $\hat{\mathbf{q}}$ acts similarly to the right. So the complex number $\langle \mathbf{p} | e^{-i\hat{H}(t_2-t_1)/\hbar} | \mathbf{q}_1 \rangle$ becomes simply

$$e^{-iH(t_2-t_1)/\hbar} \langle \mathbf{p} | \mathbf{q}_1 \rangle = e^{-iH(t_2-t_1)/\hbar} \frac{e^{-i\mathbf{p}\cdot\mathbf{q}_1/\hbar}}{\sqrt{2\pi\hbar}}, \quad (5.20)$$

where H is the classical Hamiltonian evaluated at the classical phase-space point (\mathbf{p}, \mathbf{q}) and we have used the fact that $\langle \mathbf{p} | \mathbf{q}_1 \rangle$ is just the complex conjugate of the wave-function of a particle of well-defined momentum \mathbf{p} . When we insert (5.20) into (5.19) and similarly replace $\langle \mathbf{q}_2 | \mathbf{p} \rangle$ by a plane wave, we find

$$A_{12} = \frac{1}{h} \int d^3\mathbf{p} \exp \left[\frac{i}{\hbar} \left(\mathbf{p} \cdot (\mathbf{q}_2 - \mathbf{q}_1) - H(t_2 - t_1) \right) \right]. \quad (5.21)$$

Equation (5.21) for the amplitude to get from one event to another is only valid for infinitesimal $t_2 - t_1$. There are two issues: (i) \hat{H} may be time-dependent; (ii) for finite τ the operator $e^{-i\hat{H}\tau} = 1 - i\hat{H}\tau + \frac{1}{2!}(\hat{H}\tau)^2 + \dots$ involves high powers of \hat{H} and so many reversals of the order of the operators $\hat{\mathbf{p}}$ and $\hat{\mathbf{q}}$ will be required to ensure that the $\hat{\mathbf{p}}$'s are to the left of all $\hat{\mathbf{q}}$'s. In view of these objections we use (5.21) only for small $t_2 - t_1$. Given two widely separated events (t_i, \mathbf{q}_i) and (t_f, \mathbf{q}_f) , we express the amplitude to pass between them by a particular path $\mathbf{q}_i \rightarrow \mathbf{q}_1 \rightarrow \dots \rightarrow \mathbf{q}_f$ as the product

$$A_{i1} A_{12} \times \dots \times A_{m-1,f} \quad (5.22)$$

of m amplitudes of the form (5.21) over small intervals (t_{j+1}, t_j) . We then obtain the amplitude to pass between (t_i, \mathbf{q}_i) and (t_f, \mathbf{q}_f) by *any* path by summing (5.22) over all values of the intermediate positions \mathbf{q}_j . The final amplitude is

$$\begin{aligned} A_{if} &= \lim_{m \rightarrow \infty} \frac{1}{h^{3m}} \int \prod_j^m (d^3\mathbf{p}_j d^3\mathbf{q}_j) \exp \left[\frac{i}{\hbar} \sum_k^m \left(\mathbf{p}_k \cdot (\mathbf{q}_{k+1} - \mathbf{q}_k) - H(t_{k+1} - t_k) \right) \right] \\ &= \text{constant} \times \int \mathcal{D}\mathbf{p} \mathcal{D}\mathbf{q} \exp \left[\frac{i}{\hbar} \int (\mathbf{p} \cdot d\mathbf{q} - H dt) \right]. \end{aligned} \quad (5.23)$$

Here the symbol $\mathcal{D}\mathbf{p} \mathcal{D}\mathbf{q}$ means one is to sum the integrand over all paths $(\mathbf{p}(t), \mathbf{q}(t))$ which pass through (t_i, \mathbf{q}_i) and (t_f, \mathbf{q}_f) .

When \hat{H} is of the form $\hat{H} = \frac{1}{2}\hat{\mathbf{p}}^2/m + V(\hat{\mathbf{q}})$, one can do the path integral over $\mathbf{p}(t)$ in (5.23) and obtain an expression for A_{if} from which (3.15) follows in the classical limit.

Exercise (13):

In (5.23) replace H with $\frac{1}{2}\mathbf{p}^2/m + V(\mathbf{q})$ and $d\mathbf{q}$ by $\dot{\mathbf{q}}dt$. Then do the integration over every \mathbf{p}_j by completing the square and using $\int_{-\infty}^{\infty} e^{-x^2} dx = \sqrt{\pi}$. Explain the relation of the resulting expression for A_{if} to (3.15).

6 Lagrangian Field Theory

We've seen how equations of motion may be derived from the least-action principle

$$0 = \delta S \equiv \int dt L(x, \dot{x}) \quad (6.1)$$

Starting with an action has many advantages:

- Since L is a scalar, transforming to new coordinates is easy;
- It's easy to ensure that the eqns of motion are Lorentz invariant (or Galilean invariant as appropriate) by imposing the desired invariance on L ;
- Given the required invariance and the basic form of the desired eqns (second-order, linear, say) only a few simple expressions are candidates for Lagrangians;
- Certain constants of motion can be readily derived from evident symmetries of L .

Can we obtain partial differential eqns such as the wave eqn or Maxwell's eqns from Lagrangians?

Specimen problem: derive the wave eqn

$$\frac{1}{c^2} \frac{\partial^2 \phi}{\partial t^2} - \frac{\partial^2 \phi}{\partial x^2} = 0. \quad (6.2)$$

Regard $\phi(t, x)$ as a set of ∞ -dimensional vectors $\phi_x(t)$, where x labels components of ϕ . The Lagrangian has to be a scalar, so ϕ 's indices have to be 'soaked up' somehow. We make a scalar out of an ordinary vector by dotting it with another vector—this soaks up the indices of both vectors by introducing a sum over that index. Analogously, we soak up indices x with generalizations of dot products; that is, one sums over x by means of an integral:

$$s = \mathbf{a} \cdot \mathbf{b} = \sum_i a_i b_i \quad \leftrightarrow \quad s = (\psi, \phi) = \int dx \psi(x) \phi(x). \quad (6.3)$$

This leads one to expect that many (but not all) actions for partial differential equations are evaluated by integrating a Lagrangian density \mathcal{L} over space before performing the usual integral over time:

$$S[\phi] = \int dt \int dx \mathcal{L}(\phi, \dot{\phi}). \quad (6.4)$$

When we were doing Lagrangian mechanics, S was a functional of the particle's history $x(t)$. Now S is a functional of the field's history $\phi(t, x)$. So ϕ has stepped into x 's place, and x has become an independent variable with a similar standing to that of t . Consequently, in (6.4) we're integrating over both space and time.

In order to make the symmetry between x and t complete we henceforth allow \mathcal{L} to involve derivatives w.r.t. x as well as w.r.t. t ; then $\mathcal{L} = \mathcal{L}(\phi, \square_\mu \phi)$ and

$$S[\phi] = \int \mathcal{L}(\phi, \square_\mu \phi) dx dt. \quad (6.5)$$

Finally, it doesn't make things significantly more complicated to allow space to be fully three-dimensional. So x becomes the 3-vector \mathbf{x} and (ct, \mathbf{x}) becomes the usual 4-vector \mathbf{x} . Since $d^4\mathbf{x} = c dt d^3\mathbf{x}$ and nothing of significance is changed when S is multiplied by a constant, we can write simply

$$S[\phi] = \frac{1}{c} \int \mathcal{L}(\phi, \square_\mu \phi) d^4\mathbf{x}. \quad (6.6)$$

At each t between t_i and t_f the field's configuration $\phi(t, \mathbf{x})$ is chosen such that the integral (6.6) through the space-time volume bounded by $t = t_i$ and $t = t_f$ is extremized:

As in Lagrangian mechanics we are specifying a solution to the 2nd order equations of motion by giving values of the 'coordinates' at two times, t_i and t_f , rather than the coordinates and velocities at a single time. In this case specifying the 'coordinates' involves giving the functional dependence of ϕ on \mathbf{x} at some fixed t .

Here's how we extremize S :

$$\begin{aligned} 0 = \delta S &= S[\phi + \psi] - S[\phi] \quad \text{where} \quad |\psi(t, \mathbf{x})| \ll |\phi(t, \mathbf{x})| \\ &\simeq \int \left(\frac{\partial \mathcal{L}}{\partial \phi} \psi + \frac{\partial \mathcal{L}}{\partial (\square_\mu \phi)} \square_\mu \psi \right) d^4 \mathbf{x} \\ &= \int \left(\frac{\partial \mathcal{L}}{\partial \phi} - \frac{\partial}{\partial x^\mu} \frac{\partial \mathcal{L}}{\partial (\square_\mu \phi)} \right) \psi d^4 \mathbf{x} + \oint \frac{\partial \mathcal{L}}{\partial (\square_\mu \phi)} \psi d^3 \mathbf{x}_\mu. \end{aligned} \quad (6.7)$$

Here the final integral \oint is the integral over the closed 3-surface that bounds the 4-dimensional region of space-time through which \mathcal{L} is integrated. The surface consists of the initial and final hypersurfaces, and the 3-surface swept out by a 2-surface at spatial ∞ as t varies from t_i to t_f . This integral vanishes because ψ is zero throughout the domain integrated over: the variation ψ vanishes on the initial and final hypersurfaces by hypothesis, and we force it to vanish at spatial ∞ also in order to ensure that the varied field $\phi + \psi$ satisfies the same bdy condition as the unvaried field ϕ . Thus

$$\delta S = \int \left(\frac{\partial \mathcal{L}}{\partial \phi} - \frac{\partial}{\partial x^\mu} \frac{\partial \mathcal{L}}{\partial (\square_\mu \phi)} \right) \psi d^4 \mathbf{x} \quad (6.8)$$

If this is to hold for any $\psi(t, \mathbf{x})$ that vanishes on the initial and final hypersurfaces, we clearly require that

$$\frac{\partial \mathcal{L}}{\partial \phi} - \frac{\partial}{\partial x^\mu} \left(\frac{\partial \mathcal{L}}{\partial (\square_\mu \phi)} \right) = 0. \quad (6.9)$$

This p.d.e. is the Euler-Lagrange equation for a field. It is the field equation that follows from the Lagrangian density \mathcal{L} .

What do we choose for \mathcal{L} to make (6.9) the wave eqn, $\square^2 \phi = 0$? Try³

$$\mathcal{L} = \frac{1}{2} |\square \phi|^2 = \frac{1}{2} \eta^{\alpha\beta} \square_\alpha \phi \square_\beta \phi. \quad (6.10)$$

Then $\partial \mathcal{L} / \partial \phi = 0$ and $\partial \mathcal{L} / \partial (\square_\mu \phi) = \frac{1}{2} (\eta^{\mu\beta} \square_\beta \phi + \eta^{\alpha\mu} \square_\alpha \phi) = \square^\mu \phi$, so (6.9) yields

$$\begin{aligned} 0 &= \frac{\partial^2 \phi}{\partial x^0{}^2} - \nabla^2 \phi \\ &= \frac{1}{c^2} \frac{\partial^2 \phi}{\partial t^2} - \nabla^2 \phi. \end{aligned}$$

Thus the wave equation emerges from the Lagrangian density which is the simplest possible function of $\square_\mu \phi$ only.

6.1 Maxwell's equations*

What about Maxwell's equations? These are 2nd order in \mathbf{A} , so we look for a Lagrangian density \mathcal{L} that depends on \mathbf{A} and its derivatives, $\square_\mu \mathbf{A}$. Moreover, Maxwell's eqns are linear in the fields, and thus in \mathbf{A} . So \mathcal{L} should be quadratic in \mathbf{A} and $\square_\mu \mathbf{A}$. Finally, \mathcal{L} should be invariant under gauge transformations $\mathbf{A} \rightarrow \mathbf{A}' + \square \Lambda$, and should involve $\square_\mu \mathbf{A}$ only in the

³ At this point the factor 1/2 is gratuitous since the field equations are independent of the normalization of S . When we relate the energy-momentum tensor to \mathcal{L} we shall find this factor 1/2 handy, however.

* Lies beyond the syllabus

combination contained in \mathbf{F} . The shortlist of functions satisfying these criteria contains (up to an unimportant normalization) only one candidate:

$$\begin{aligned}\mathcal{L}_{\text{vac}}(\mathbf{A}, \square_{\mu}\mathbf{A}) &= \frac{1}{4\mu_0} \text{Tr } \mathbf{F} \cdot \mathbf{F} \\ &= -\frac{1}{4\mu_0} F_{\mu\nu} F^{\mu\nu} \\ &= \frac{1}{2\mu_0} (E^2/c^2 - B^2),\end{aligned}\tag{6.11}$$

where the last equality is from (1.20). (Notice that if we associate E with kinetic energy ($E = -\dot{\mathbf{A}}/c + \dots$) and B with potential energy, \mathcal{L}_{vac} is of the form k.e. $-$ p.e..) The field equations associated with the Lagrangian (6.11) density are

$$\frac{\partial}{\partial x^{\beta}} \left(\frac{\partial \mathcal{L}_{\text{vac}}}{\partial (\square_{\beta} A_{\mu})} \right) = 0.$$

Now

$$\begin{aligned}\frac{\partial F_{\mu\nu}}{\partial (\square_{\beta} A_{\alpha})} &= \frac{\partial}{\partial (\square_{\beta} A_{\alpha})} (\square_{\mu} A_{\nu} - \square_{\nu} A_{\mu}) \\ &= \delta_{\mu}^{\beta} \delta_{\nu}^{\alpha} - \delta_{\nu}^{\beta} \delta_{\mu}^{\alpha},\end{aligned}\tag{6.12}$$

so

$$\begin{aligned}\frac{\partial \mathcal{L}_{\text{vac}}}{\partial (\square_{\beta} A_{\alpha})} &= -\frac{1}{4\mu_0} \frac{\partial (F_{\mu\nu} F^{\mu\nu})}{\partial (\square_{\beta} A_{\alpha})} = -\frac{1}{4\mu_0} \frac{\partial (F_{\mu\nu} \eta^{\mu\kappa} \eta^{\nu\lambda} F_{\kappa\lambda})}{\partial (\square_{\beta} A_{\alpha})} \\ &= -\frac{1}{2\mu_0} (\delta_{\mu}^{\beta} \delta_{\nu}^{\alpha} - \delta_{\nu}^{\beta} \delta_{\mu}^{\alpha}) F^{\mu\nu} \\ &= -\frac{1}{2\mu_0} (F^{\beta\alpha} - F^{\alpha\beta}) \\ &= \frac{1}{\mu_0} F^{\alpha\beta}.\end{aligned}\tag{6.13}$$

The field equations are therefore

$$\frac{\partial F^{\alpha\beta}}{\partial x^{\beta}} = 0,\tag{6.14}$$

that is, 4 of Maxwell's 8 field eqns for an e.m. field in vacuo.

To get Maxwell's eqns in the presence of charges we need to add to the action S obtained by integrating (6.11) over spacetime, the action of particles in a given e.m. field. For a single charged particle the latter is given by (3.6). What does this suggest for the action associated with a swarm of particles of charge q , mass m_0 that are moving with 4-velocity $\mathbf{v}(\mathbf{x})$ and in their rest-frame have number density $n(\mathbf{x})$? Well, the form of (3.6) suggests that the part of \mathcal{L} which depends on both the e.m. field and the particles (the 'interaction term'), is proportional to the dot product of \mathbf{A} with the current density $\mathbf{j} = qn_0\mathbf{v}$ associated with the particles. So we speculate that the interaction term is $k\mathbf{j} \cdot \mathbf{A}$, where k is a suitable constant. We check this conjecture by substituting from (2.24) for the current density due to a single particle:

$$\begin{aligned}S_{\text{interaction}} &= k \int (\mathbf{j} \cdot \mathbf{A})|_{\mathbf{x}} d^4\mathbf{x} \\ &= kqc \int \dot{\mathbf{X}} \cdot \mathbf{A}(\mathbf{x}) \delta(\mathbf{x} - \mathbf{X}) d^4\mathbf{x} d\tau \\ &= kqc \int \dot{\mathbf{X}} \cdot \mathbf{A}(\mathbf{X}) d\tau\end{aligned}\tag{6.15}$$

which agrees with (3.6) for $k = 1/c$.

So long as we are only interested in getting the field eqns, which are obtained by varying \mathbf{A} , we don't need to bother with the contribution to S from matter alone (which is independent of \mathbf{A}). So let's see whether this action begets Maxwell's eqns with sources:

$$S = \frac{1}{c} \int (\mathbf{j} \cdot \mathbf{A} + \frac{1}{4\mu_0} \text{Tr } \mathbf{F} \cdot \mathbf{F}) d^4\mathbf{x}. \quad (6.16)$$

Varying \mathbf{A} with the aid of previous results, the field eqns are found to be

$$j_\mu - \frac{1}{\mu_0} \frac{\partial F_{\mu\nu}}{\partial x^\nu} = 0 \quad (6.17)$$

in agreement with (1.46). The other four Maxwell's eqns don't come from minimizing the action but from the fact that \mathbf{F} is the 4-curl of \mathbf{A} . So they are geometrical rather than dynamical in nature.

6.2 Klein-Gordon Equation*

The wave function of a spin-zero particle of mass m_0 should satisfy $\hat{p}^2\psi = -m_0^2c^2\psi$, where

$$\hat{p}^2 = |\hat{E}/c, \hat{\mathbf{p}}|^2 = \left| \frac{i\hbar}{c} \frac{\partial}{\partial t}, -i\hbar\nabla \right|^2 = \hbar^2 \left(\frac{\partial^2}{c^2\partial t^2} - \nabla^2 \right).$$

That is, we require $\square^2\psi = \frac{m_0^2c^2}{\hbar^2}\psi$. What Lagrangian density generates this eqn? We try

$$\mathcal{L}(\psi, \square_\mu\psi) = \frac{1}{2} \left(|\square\psi|^2 + \frac{m_0^2c^2}{\hbar^2} |\psi|^2 \right). \quad (6.18)$$

By $|\square\psi|^2$ we mean

$$|\square\psi|^2 = -\frac{1}{c^2} \frac{\partial\psi^*}{\partial t} \frac{\partial\psi}{\partial t} + \nabla\psi^* \cdot \nabla\psi. \quad (6.19)$$

Differentiating w.r.t. ψ is slightly tricky because ψ^* is a function $\psi^*(\psi)$ of ψ . We handle this by writing $\psi = u + iv$ and treating the real and imaginary parts of u and v as independent real fields:

$$\begin{aligned} \frac{\partial|\psi|^2}{\partial u} &= \frac{\partial}{\partial u}(u^2 + v^2) = 2u, \\ \frac{\partial|\psi|^2}{\partial v} &= 2v. \end{aligned} \quad (6.20)$$

Further

$$|\square\psi|^2 = \square(u - iv) \cdot \square(u + iv) = |\square u|^2 + |\square v|^2.$$

So

$$\frac{\partial|\square\psi|^2}{\partial(\square_\mu u)} = 2\square^\mu u \quad ; \quad \frac{\partial|\square\psi|^2}{\partial(\square_\mu v)} = 2\square^\mu v. \quad (6.21)$$

Hence the field eqns are

$$\frac{\partial}{\partial x^\mu} \square^\mu u - \frac{m_0^2c^2}{\hbar^2} u = 0, \quad (6.22a)$$

$$\frac{\partial}{\partial x^\mu} \square^\mu v - \frac{m_0^2c^2}{\hbar^2} v = 0. \quad (6.22b)$$

The Klein-Gordon eqn is obtained by adding i times (6.22b) to (6.22a).

* Lies beyond the syllabus

The following result simplifies the variation of an action that depends on a complex field ψ . Suppose $\delta f(\psi, \psi^*) = 0$. We have

$$0 = \delta f = \frac{\partial f}{\partial \psi}(\delta u + i\delta v) + \frac{\partial f}{\partial \psi^*}(\delta u - i\delta v).$$

Since δu and δv are arbitrary, we conclude

$$\left. \begin{aligned} 0 &= \frac{\partial f}{\partial \psi} + \frac{\partial f}{\partial \psi^*} \\ 0 &= \frac{\partial f}{\partial \psi} - \frac{\partial f}{\partial \psi^*} \end{aligned} \right\} \Leftrightarrow \left\{ \begin{aligned} 0 &= \frac{\partial f}{\partial \psi} \\ 0 &= \frac{\partial f}{\partial \psi^*}. \end{aligned} \right.$$

Thus we can proceed as though $\delta\psi$ and $\delta\psi^*$ were independent, though they are not.

6.3 Einstein field equations*

General relativity buffs will be wondering what action generates Einstein's equations. Well, $\mathcal{L}d^4\mathbf{x}$ must be a scalar made up of the metric tensor $g_{\mu\nu}$ and its derivatives. There is only one serious candidate for this job, the Ricci scalar $R \equiv R^\mu{}_\mu \equiv R^{\mu\alpha}{}_{\mu\alpha}$, where $R^\mu{}_{\alpha\beta\gamma}$ is the curvature tensor. So is $\mathcal{L}d^4\mathbf{x}$ just $Rd^4\mathbf{x}$? Not quite, since we want $S = \int \mathcal{L} d^4\mathbf{x}$ to be a scalar under general changes of coordinates $\mathbf{x} \rightarrow \mathbf{x}'$, not just under Lorentz transformations. In a general coordinate change the Jacobian $\partial(\mathbf{x}')/\partial(\mathbf{x})$ is not unity, so $d^4\mathbf{x}$ is not a scalar, and \mathcal{L} cannot be a scalar either if S is to be a scalar. We have

$$d^4\mathbf{x} = \frac{\partial(\mathbf{x})}{\partial(\mathbf{x}')} d^4\mathbf{x}'. \quad (6.23)$$

But

$$g_{\mu\nu} = \frac{\partial x'^\alpha}{\partial x^\mu} \frac{\partial x'^\beta}{\partial x^\nu} g'_{\alpha\beta}, \quad (6.24)$$

so taking determinants we have

$$|g_{\mu\nu}| = \left(\frac{\partial(\mathbf{x}')}{\partial(\mathbf{x})} \right)^2 |g'_{\mu\nu}|, \quad (6.25)$$

since $\partial(\mathbf{x}')/\partial(\mathbf{x})$ is by definition the determinant of the matrix $\partial x'^\alpha/\partial x^\beta$. Multiplying (6.23) by the square root of (6.25) we see that $\sqrt{-|g_{\mu\nu}|}d^4\mathbf{x}$ is a scalar. ($|g| < 0$ no matter what signature one uses.) Hence

$$S_{\text{grav}}[\mathbf{g}] = -\frac{c^3}{16\pi G} \int R \sqrt{-|g_{\mu\nu}|} d^4\mathbf{x} \quad (6.26)$$

is a general scalar and a candidate for the action that generates Einstein's equations. For a demonstration that this action works, and justification of the numerical prefactor (which is only of consequence when one adds in the action of matter), see e.g., Weinberg *Gravitation and Cosmology* p. 364.

* Lies beyond the syllabus

6.4 Noether's theorem for internal symmetries*

Does Noether's theorem for the Lagrangians of particle motion extend to Lagrangian densities for fields? Actually it yields *two* closely related results: one for internal symmetries and one for external symmetries, such as Lorentz invariance. We deal with internal symmetries first.

Often $\mathcal{L}(\mathbf{A}, \square_\mu \mathbf{A})$ is invariant under some transformation of the field \mathbf{A} . For example, in the case of e.m. \mathcal{L} is invariant under $\mathbf{A} \rightarrow \mathbf{A} + \square \Lambda$ where $\Lambda(\mathbf{x})$ is any scalar function.⁴ Whenever there is a point-by-point invariance of this type, we can write

$$\begin{aligned} 0 = \delta \mathcal{L} &= \frac{\partial \mathcal{L}}{\partial \mathbf{A}} \cdot \delta \mathbf{A} + \frac{\partial \mathcal{L}}{\partial (\square_\mu \mathbf{A})} \cdot \delta (\square_\mu \mathbf{A}) \\ &= \frac{\partial}{\partial x^\mu} \left(\frac{\partial \mathcal{L}}{\partial (\square_\mu \mathbf{A})} \right) \cdot \delta \mathbf{A} + \frac{\partial \mathcal{L}}{\partial (\square_\mu \mathbf{A})} \cdot \square_\mu (\delta \mathbf{A}) \\ &= \frac{\partial}{\partial x^\mu} \left(\delta \mathbf{A} \cdot \frac{\partial \mathcal{L}}{\partial (\square_\mu \mathbf{A})} \right), \end{aligned} \quad (6.27)$$

where the field eqns (6.9) have been used. The final line states that the current density j^μ has vanishing divergence, where

$$j^\mu \equiv \delta \mathbf{A} \cdot \frac{\partial \mathcal{L}}{\partial \square_\mu \mathbf{A}}. \quad (6.28)$$

The vanishing of $\square \cdot \mathbf{j}$ implies that the integral $J \equiv \int j^\mu d^3 \mathbf{x}_\mu \equiv \int j^\mu \epsilon_{\mu\alpha\beta\gamma} dx^\alpha dx^\beta dx^\gamma$ is the same for any two 3-dimensional spatial hypersurfaces: Given two such surfaces we can extend these into the closed surface bounding a spacetime volume like this:

$\square \cdot \mathbf{j} = 0$ implies that the flux into this volume has to equal that out of it, so the upward flux through the bottom surface has to equal the upward flux through the top surface. Thus the internal symmetry of \mathcal{L} has generated a conserved flux J .

E.m. charge conservation How does this idea work out in e.m? Setting $\delta \mathbf{A} = \square \Lambda$, we have

$$\begin{aligned} j^\mu &= (\square_\alpha \Lambda) \frac{\partial \mathcal{L}_{\text{vac}}}{\partial (\square_\mu A_\alpha)} \\ &= \frac{1}{\mu_0} (\square_\alpha \Lambda) F^{\alpha\mu}, \end{aligned} \quad (6.29)$$

where use has been made of (6.13). Equating to zero the divergence of this we find that

$$\begin{aligned} 0 &= \frac{\partial^2 \Lambda}{\partial x^\mu \partial x^\alpha} F^{\alpha\mu} + \frac{\partial \Lambda}{\partial x^\alpha} \frac{\partial F^{\alpha\mu}}{\partial x^\mu} \\ &= \frac{\partial \Lambda}{\partial x^\alpha} \square_\mu F^{\alpha\mu}, \end{aligned}$$

where the first term on the right has been eliminated by virtue of \mathbf{F} 's antisymmetry. Since we can arrange for $\square \Lambda$ to be any vector at a given point, (6.29) implies that $\square_\mu F^{\alpha\mu} = 0$. This is just (6.14), the standard field eqn for e.m. in vacuo.

To obtain a more interesting Noether invariant one has to start from \mathcal{L} for the e.m. field plus a matter field, say ψ .

* Lies beyond the syllabus

⁴ Notice the difference with the least-action principle, which states that $0 = \delta \int d^4 \mathbf{x} \mathcal{L}$ for *any* variation $\delta \mathbf{A}$; for most variations, \mathcal{L} changes at each point, it is just its integral which is invariant.

Klein-Gordon current The Klein-Gordon \mathcal{L} (6.18) is invariant under changes in the phase of ψ , i.e., $\psi \rightarrow e^{i\theta}\psi$. When θ is small we have $\delta u + i\delta v = \delta\psi \simeq i\theta\psi$, so the changes in the real and imaginary parts of ψ are

$$\delta u = -\theta v \quad ; \quad \delta v = \theta u. \quad (6.30)$$

Since we are considering \mathcal{L} to be a function of (u, v) and their derivatives, the dot in (6.28) has to be interpreted as a sum over u and v . Using our results (6.21) we find that the conserved current is

$$\begin{aligned} j_\mu &= \delta u \square_\mu u + \delta v \square_\mu v \\ &= \theta \left(-v \frac{\partial u}{\partial x^\mu} + u \frac{\partial v}{\partial x^\mu} \right) \\ &= \frac{\theta}{2i} \left(\psi^* \frac{\partial \psi}{\partial x^\mu} - \psi \frac{\partial \psi^*}{\partial x^\mu} \right). \end{aligned} \quad (6.31)$$

It is simple to verify $\square \cdot \mathbf{j} = 0$ by taking the divergence and using the Klein-Gordon equation and its complex conjugate to eliminate \square^2 .

From simple non-relativistic q.m. we recognize (6.31) as $\theta m/\hbar$ times the particle flux: For Hamiltonian $H = p^2/2m$ we have

$$\begin{aligned} \frac{d}{dt} \int |\psi|^2 d^3\mathbf{x} &= \int \left(\frac{\partial \psi^*}{\partial t} \psi + \psi^* \frac{\partial \psi}{\partial t} \right) d^3\mathbf{x} \\ &= \int \left(\frac{H\psi^*}{-i\hbar} \psi + \psi^* \frac{H\psi}{i\hbar} \right) d^3\mathbf{x} \\ &= \frac{\hbar}{2im} \int \left((\nabla^2 \psi^*) \psi - \psi^* \nabla^2 \psi \right) d^3\mathbf{x} \\ &= \frac{\hbar}{2im} \oint \left((\nabla_i \psi^*) \psi - \psi^* \nabla_i \psi \right) d^2\mathbf{x}_i. \end{aligned} \quad (6.32)$$

In the relativistic formulation (6.31) the left side of (6.32) is to be found in the terms on the right of the equality that involve time derivatives.

6.5 Noether's theorem and Lorentz invariance

The Lagrangian density \mathcal{L} of a Lorentz-covariant theory depends on \mathbf{x} only through the field \mathbf{A} and its derivatives, i.e., it has no explicit space-time dependence. Consider an infinitesimal shift in the coordinate origin which changes the coordinates of the point \mathbf{x} to $\mathbf{x}' \equiv \mathbf{x} + \mathbf{a}$, where \mathbf{a} is very small. Then the difference in the value of \mathcal{L} at \mathbf{x} and at the point $\mathbf{x} + \mathbf{a}$ whose coordinates in the unprimed frame coincide with \mathbf{x} 's coordinates in the primed frame is

$$\begin{aligned} \delta\mathcal{L} &= \left(\frac{\partial\mathcal{L}}{\partial\mathbf{A}} \cdot \frac{\partial\mathbf{A}}{\partial x^\alpha} + \frac{\partial\mathcal{L}}{\partial(\square_\nu\mathbf{A})} \cdot \frac{\partial(\square_\nu\mathbf{A})}{\partial x^\alpha} \right) a^\alpha \\ &= \left(\frac{\partial}{\partial x^\nu} \left(\frac{\partial\mathcal{L}}{\partial(\square_\nu\mathbf{A})} \right) \cdot \frac{\partial\mathbf{A}}{\partial x^\alpha} + \frac{\partial\mathcal{L}}{\partial(\square_\nu\mathbf{A})} \cdot \frac{\partial^2\mathbf{A}}{\partial x^\alpha \partial x^\nu} \right) a^\alpha \\ &= \frac{\partial}{\partial x^\nu} \left(\frac{\partial\mathcal{L}}{\partial(\square_\nu\mathbf{A})} \cdot \frac{\partial\mathbf{A}}{\partial x^\alpha} \right) a^\alpha. \end{aligned} \quad (6.33)$$

On the other hand, if we simply regard \mathcal{L} as a function of \mathbf{x} through the fields, we have

$$\delta\mathcal{L} = a^\alpha \frac{\partial\mathcal{L}}{\partial x^\alpha} = \frac{\partial}{\partial x^\nu} (\mathcal{L} \delta_\alpha^\nu a^\alpha). \quad (6.34)$$

Equating these two expressions for $\delta\mathcal{L}$ we have

$$0 = \frac{\partial}{\partial x^\nu} \left(\frac{\partial\mathcal{L}}{\partial(\square_\nu\mathbf{A})} \cdot \frac{\partial\mathbf{A}}{\partial x^\alpha} - \mathcal{L} \delta_\alpha^\nu \right) a^\alpha. \quad (6.35)$$

Furthermore, \mathbf{a} is an arbitrary small vector so its coefficient in (6.35) must vanish. Thus from the fact that \mathcal{L} depends on \mathbf{x} only through the fields we can conclude that the tensor

$$\hat{T}^\nu{}_\mu \equiv \left(\frac{\partial \mathcal{L}}{\partial(\square_\nu \mathbf{A})} \cdot \frac{\partial \mathbf{A}}{\partial x^\mu} - \mathcal{L} \delta^\nu{}_\mu \right) \quad (6.36)$$

has vanishing divergence: $\square_\nu \hat{T}^\nu{}_\mu = 0$. \mathbf{T} is the **canonical energy-momentum** tensor and the vanishing of its divergence expresses energy-momentum conservation.

Again using (6.13), we find for the canonical energy-momentum tensor of the e.m. field

$$\hat{T}^\nu{}_\mu = \frac{1}{\mu_0} \left(F^{\alpha\nu} \frac{\partial A_\alpha}{\partial x^\mu} + \frac{1}{4} F_{\alpha\beta} F^{\alpha\beta} \delta^\nu{}_\mu \right). \quad (6.37)$$

Even when we lower \mathbf{T} 's first index by premultiplying by $\eta_{\kappa\nu}$, this isn't symmetric like the \mathbf{T} of §1.4? We'd very much like $\hat{\mathbf{T}}$ to be symmetric, if only because Einstein's equations require it to be so. Also we'd like the energy-momentum tensor to depend on \mathbf{A} only through \mathbf{F} . We can attain both goals by adding into $\hat{\mathbf{T}}$ what's necessary to upgrade the derivative of \mathbf{A} in the first term into an \mathbf{F} . The required item is

$$\Delta^\nu{}_\mu = -\frac{1}{\mu_0} F^{\alpha\nu} \frac{\partial A_\mu}{\partial x^\alpha}. \quad (6.38)$$

In the absence of sources (which is when we would expect the energy-momentum tensor to be divergence-free) Δ is itself divergence free:

$$\square_\nu \Delta^\nu{}_\mu = -\frac{1}{\mu_0} \frac{\partial^2 (F^{\alpha\nu} A_\mu)}{\partial x^\nu \partial x^\alpha} = 0. \quad (6.39)$$

So if we define $\mathbf{T} \equiv \hat{\mathbf{T}} + \Delta$, \mathbf{T} will be symmetric and divergence-free in vacuo. The energy-momentum tensor of the e.m. field is then

$$\begin{aligned} T^\nu{}_\mu &= \frac{1}{\mu_0} \left(F^{\alpha\nu} \frac{\partial A_\alpha}{\partial x^\mu} - F^{\alpha\nu} \frac{\partial A_\mu}{\partial x^\alpha} + \frac{1}{4} F^{\alpha\beta} F_{\alpha\beta} \delta^\nu{}_\mu \right) \\ &= \frac{1}{\mu_0} \left(F_{\mu\alpha} F^{\alpha\nu} + \frac{1}{4} F^{\alpha\beta} F_{\alpha\beta} \delta^\nu{}_\mu \right). \end{aligned} \quad (6.40)$$

This is minus the energy-momentum tensor (1.28). The minus sign arises because we have here defined \mathbf{T} so that T_0^0 rather than $T_{00} = -T_0^0$ is the energy density.

Relativity, Electromagnetism & Mechanics III

1. A particle of mass m_1 hangs by a light string of length l from a rigid support, and a second mass, m_2 , hangs by an identical string from m_1 . The angles with the vertical of the strings supporting m_1 and m_2 are θ and ϕ , respectively. Show that the frequencies of the two normal modes of oscillation about equilibrium are ω_{\pm} , where

$$\omega_{\pm}^2 = g \frac{m_1 + m_2}{m_1} \left[1 \pm \sqrt{\frac{m_2}{m_1 + m_2}} \right].$$

Describe the motion in each of the normal modes in the cases (a) $m_1 \gg m_2$, and (b) $m_2 \gg m_1$.

2. A particle of mass m slides inside a smooth straight tube OA to which it is connected at point O by a light spring of natural length a and modulus mk . The system rotates in a horizontal plane with constant angular velocity ω about a fixed vertical axis through O. Determine the distance r of the particle from O at time t for the case when $\omega^2 < k/a$, if $r = a$ and $\dot{r} = 0$ at $t = 0$. Show also for this case that the maximum value of the reaction of the tube on the particle is $2ma\omega^3/b$, where $b^2 \equiv (k/a - \omega^2)$.

3. Use a Lagrangian to show that when referred to spherical polar coordinates, the equations of motion of a particle in a gravitational potential $\Phi(\mathbf{x})$ are

$$\begin{aligned} 0 &= \ddot{r} - r(\dot{\theta}^2 + \sin^2 \theta \dot{\phi}^2) + \frac{\partial \Phi}{\partial r} \\ 0 &= \frac{d}{dt}(r^2 \dot{\theta}) - r^2 \dot{\phi}^2 \sin \theta \cos \theta + \frac{\partial \Phi}{\partial \theta} \\ 0 &= \frac{d}{dt}(r^2 \sin^2 \theta \dot{\phi}) + \frac{\partial \Phi}{\partial \phi}. \end{aligned}$$

In the case in which $\Phi = \Phi(r)$ is spherically symmetric, show that

$$mr^2 \sqrt{\dot{\theta}^2 + \sin^2 \theta \dot{\phi}^2}$$

is a conserved quantity and interpret this result physically.

4. A circular hoop of mass m and radius a hangs from a point on its circumference and is free to oscillate in its own plane. A bead of mass m can slide without friction around the hoop. Choose a set of generalized coordinates and write down the Lagrangian for the system. Show that the natural frequencies for small oscillations about equilibrium are $\omega_1 = \sqrt{2g/a}$ and $\omega_2 = \sqrt{g/2a}$.

5. What is meant by the terms *symmetry principle* and *conservation law* as used in classical dynamics? Give simple examples to illustrate the symmetries underlying the conservation of linear and angular momentum.

A particle with position coordinates \mathbf{r} moves in a central potential $V(r)$. By considering the quantity $(\mathbf{r} \times \dot{\mathbf{r}})$ show that the orbit of the particle lies in a fixed plane.

Find all potential functions $V(r)$ and corresponding functions $\alpha(r)$ for which the vector

$$\mathbf{K} = \dot{\mathbf{r}} \times (\mathbf{r} \times \dot{\mathbf{r}}) + \alpha(r)\mathbf{r}$$

is conserved.

Find also the potentials $V(r)$ and functions $\beta(r)$ for which the components of the matrix

$$Q_{ij} \equiv \dot{r}_i \dot{r}_j + \beta(r)r_i r_j$$

are constants of the motion, where r_i, \dot{r}_i ($i = 1, 2, 3$) are the components of position and velocity of the particle along any three independent fixed axes.

6. A system with three degrees of freedom described by coordinates q_1, q_2, q_3 has Lagrangian

$$L = \frac{1}{2}(\dot{q}_1^2 + \dot{q}_2^2 + \dot{q}_3^2) - \frac{1}{2}(q_1^2 + q_2^2 + q_3^2) + \alpha(q_2q_3 + q_3q_1 + q_1q_2),$$

where $0 < \alpha < \frac{1}{2}$. Show that L is invariant under infinitesimal rotations about the $(1, 1, 1)$ axis in q -space, and hence find a constant of motion other than the total energy. Verify from the equations of motion that it is indeed constant.

7. A point charge q is placed at the origin in the magnetic field generated by a spatially confined current distribution. Given that

$$\mathbf{E} = \frac{q}{4\pi\epsilon_0} \frac{\mathbf{r}}{r^3}$$

and $\mathbf{B} = \nabla \times \mathbf{A}$ with $\nabla \cdot \mathbf{A} = 0$, show that the field's momentum

$$\mathbf{P} \equiv \epsilon_0 \int \mathbf{E} \times \mathbf{B} \, d^3\mathbf{x} = q\mathbf{A}(0).$$

Use this result to interpret the formula for the canonical momentum of a charged particle in an e.m. field.