COMMENTS ON JACKSON September 27, 2024

These notes are meant as a running commentary on Jackson, chapters 1 to 7 and chapter 11. They will replace the usual lectures, and contain a fairly large number of exercises and problems. The idea is that you should try those and send them to me. Then you will receive detailed comments on your attempts. This is actually an old idea associated to the name 'Hermods', meaning that we throw fifty years of pedagogical development to the winds.

Grades are given based on how many solutions you send in. If I have critical comments on them you can simply try again, but you must follow Feynman's recipe for problem solving:

- 1: Write down the problem.
- 2: Think very hard.
- 3: Write down the solution.

Steps 1 and 3 are very important.

Some of the contents of the course should be familiar from earlier courses, but Jackson treats things in more depth. His book is not easy to read. He presents the subject, and seems unconcerned about what parts are easy to explain. Excellent introductory books include

- E. M. Purcell: *Electricity and Magnetism*, Berkeley physics course 1965.
- R. P. Feynman, R. B. Leighton, and M. Sands: *The Feynman Lectures on Physics Vol II*, Addison-Wesley 1963.

The subject is presented from a modern (2022) point of view in

- R. M. Wald: *Advanced Classical Electromagnetism*, Princeton UP 2022.

Assuming for the sake of the argument that the University had not decided to abolish the library, a visit to it would reveal that many other excellent books exist.

INTRODUCTION

The partial differential equations first came to theoretical physics as a servant, but by degrees it became its master.

Albert Einstein

Maxwell's equations for the electromagnetic field are

$$\nabla \cdot \mathbf{B} = 0 \qquad \text{Gilbert's law} \tag{1}$$

$$\nabla \times \mathbf{E} + \partial_t \mathbf{B} = \mathbf{0} \qquad \text{Faraday's law} \tag{2}$$

$$\nabla \cdot \mathbf{E} = \frac{1}{\epsilon_0}\rho \qquad \text{Gauss' law} \tag{3}$$

$$\nabla \times \mathbf{B} - \frac{1}{c^2}\partial_t \mathbf{E} = \mu_0 \mathbf{J} \qquad \text{Ampère-Maxwell's law ,} \tag{4}$$

where $c^2 = 1/\epsilon_0\mu_0$. I use the notation

$$\partial_t = \frac{\partial}{\partial t} \ . \tag{5}$$

The charge density $\rho = \rho(\mathbf{x}, t)$ and the current density $\mathbf{J} = \mathbf{J}(\mathbf{x}, t)$ give a complete account of all charges. For consistency they must obey[1]

$$\partial_t \rho + \nabla \cdot \mathbf{J} = 0 \ . \tag{6}$$

This is known as the equation of *charge conservation*.

Given a solution of Maxwell's equations the force acting at position $\mathbf{x}$ on a test particle carrying a small charge q is given by *Lorentz' equation*

$$\mathbf{F}(\mathbf{x}, \dot{\mathbf{x}}, t) = q(\mathbf{E}(\mathbf{x}, t) + \dot{\mathbf{x}} \times \mathbf{B}(\mathbf{x}, t)) \ , \tag{7}$$

where $\dot{\mathbf{x}}$ is the velocity of the particle. A complete theory needs dynamical equations for ρ and $\mathbf{J}$, consistent with charge conservation and also with Lorentz' equation. Setting ρ and $\mathbf{J}$ to zero leads to an interesting theory in itself.

[1]Exercise: Prove this. If you have forgotten about div, grad, and curl, read on, then go back and prove this.

Units

The dimension of the constant c must be [metre/second], otherwise the left hand sides of eqs. (2) and (4) would be mutually inconsistent. Its numerical value in these units—close to $3 \cdot 10^8$—can be fixed by experiments.[2] On the other hand the constants ϵ_0, μ_0 depend on the unit system chosen for the electric and magnetic fields. To make this choice one begins by making contact with the mechanical units through Coulomb's law: using Maxwell's equations one can show (and we will) that the force between two static charges at distance r metres from each other is

$$F = \frac{1}{4\pi\epsilon_0}\frac{qq'}{r^2} \ . \tag{8}$$

Therefore the dimension of q^2/ϵ_0 is that of [kg×metre^3/second^2], but the dimension and numerical value of ϵ_0 will depend on the unit adopted for charge. There is a similar story for μ_0. Incidentally the dimension of q^2/ϵ_0 is the same as the dimension of hc, where h is Planck's constant. It follows that $q^2/\epsilon_0 hc$ is a dimensionless number—and an interesting one, if q stands for the charge of the electron.

In the SI system, which we will use, one makes the arbitrary choice

$$\epsilon_0 = \frac{10^7}{4\pi c^2} \ , \qquad \mu_0 = 4\pi \cdot 10^{-7} \ . \tag{9}$$

This comes from having adopted the Ampère as a basic unit for current, in terms of the force acting between two parallel wires. To reach the more intelligent Gaussian unit system one sets $F = qq'/r^2$, insists that the electric and magnetic fields have the same dimension, and makes sure that the only constants that appear in the equations are c and 4π. The factor of 4π is there essentially because the unit sphere has area 4π. Trying to suppress it in the equations will make it pop up somewhere else.

Some everyday quantities come out as everyday numbers in Gaussian units. In my home the geomagnetic field is about half a Gauss, and about 50

[2]Exercise: Show that the dimension of c must be that of length/time. Then start from Maxwell's equations in SI units and rescale $\mathbf{E}$ and $\mathbf{B}$ with dimensionful constants so that they get the same dimension. Can you do further rescalings so that ϵ_0 and μ_0 disappear from the equations?

microteslas in SI units. The Gaussian system also has the obvious advantage that the man-made constants ϵ_0 and μ_0 do not appear.

Jackson switches to Gaussian units later in the book. He has a nice discussion of units on pp. 775-784. The papers by Birge that he cites are also very readable, and you can consult more recent references if you wish.[3]

Vectors and vector fields

It is important to feel comfortable with vector analysis. A boldface letter denotes a vector, say

$$\mathbf{E} = \mathbf{e}_x E_x + \mathbf{e}_y E_y + \mathbf{e}_z E_z \ . \tag{10}$$

The alternative notation E_i to denote a vector does not enter Jackson's book until he comes to some of the trickier calculations involving magnetic fields, but—under the name of *Cartesian tensors*—it plays an essential role in special relativity. If the basis vectors are understood one can think of a vector as an array of numbers,

$$\mathbf{E} = \begin{pmatrix} E_x \\ E_y \\ E_z \end{pmatrix} = E_i \ . \tag{11}$$

As it stands this is sloppy but useful. The index i now runs from 1 to 3. Finally, $\mathbf{E}(\mathbf{x}) = E_i(x)$ is a *vector field*, a vector defined at each point of space, and can be partially visualized by its field lines, which go through the points in the direction of the vector defined there.

The differential operator ∇ is called "nabla", Hebrew for the musical instrument known in English as "psalter"—because this is what the symbol looked like to Maxwell, who knew his Bible. When it acts on a function it gives the *gradient* of the function, which is the vector field

$$\nabla f = \mathbf{e}_x \partial_x f + \mathbf{e}_y \partial_y f + \mathbf{e}_z \partial_z f = \partial_i f \ . \tag{12}$$

Nabla operates on a vector field $\mathbf{E}$ in two different ways. The *divergence* of $\mathbf{E}$ is the function

[3]Reference: J. de Boer, *On the history of the quantity calculus and the International System*, Metrologia **31** (1995) 405.

$$\nabla \cdot \mathbf{E} = \partial_x E_x + \partial_y E_y + \partial_z E_z = \partial_i E_i \ . \tag{13}$$

I am assuming a Cartesian coordinate system. If you wish to use spherical or cylindrical coordinates you can consult Jackson's back pages. For the index notation we have adopted the *Einstein summation convention*, according to which a repeated index occurring in some term of an equation is always summed over. The *rotation* or *curl* of $\mathbf{E}$ is the vector field

$$\nabla \times \mathbf{E} = \begin{pmatrix} \partial_y E_z - \partial_z E_y \\ \partial_z E_x - \partial_x E_z \\ \partial_x E_y - \partial_y E_x \end{pmatrix} = \epsilon_{ijk} \partial_j E_k \ . \tag{14}$$

Note the cyclic pattern. Note also that the notation has been streamlined to fit a three dimensional space, and that we sum over j and k in the index notation. The object ϵ_{ijk} is defined by $\epsilon_{123} = 1$ and by being totally anti-symmetric, that is to say it changes its sign whenever the order of two indices is switched (so that for example $\epsilon_{132} = -1$, $\epsilon_{312} = 1$, and $\epsilon_{112} = 0$).

Applying nabla repeatedly we obtain the *Laplace operator*

$$\nabla^2 = \nabla \cdot \nabla = \partial_x^2 + \partial_y^2 + \partial_z^2 = \partial_i \partial_i \ . \tag{15}$$

It turns functions into functions and vectors into vectors. (It could look quite complicated, had we not agreed to use a Cartesian coordinate system.) In a purely automatic manner one proves the identities

$$\nabla \times (\nabla f) = 0 = \epsilon_{ijk} \partial_j \partial_k f \ , \qquad \nabla \cdot (\nabla \times \mathbf{E}) = 0 = \partial_i (\epsilon_{ijk} \partial_j E_k) \tag{16}$$

$$\nabla \times (\nabla \times \mathbf{E}) = \nabla (\nabla \cdot \mathbf{E}) - \nabla^2 \mathbf{E} \ . \tag{17}$$

Because they are written in boldface they hold in all coordinate systems. It is instructive to prove the identity (17) in the tensor formalism. We then rely on the ϵ–δ identity

$$\epsilon_{ijk} \epsilon_{kmn} = \delta_{im} \delta_{jn} - \delta_{in} \delta_{jm} \ , \tag{18}$$

where the Kronecker delta was introduced.[4] The calculation goes

[4]Exercise: Prove the ϵ–δ identity (preferably without thinking). To see if it is useful or not, prove eq. (17) without relying on this identity.

$$\epsilon_{ijk}\partial_j(\epsilon_{kmn}\partial_m E_n) = \epsilon_{ijk}\epsilon_{kmn}\partial_j\partial_m E_n =$$
$$= (\delta_{im}\delta_{jn} - \delta_{in}\delta_{jm})\partial_j\partial_m E_n = \partial_i\partial_j E_j - \partial_j\partial_j E_i \ . \tag{19}$$

Vector analysis

We will be interested in three theorems that relate the integral over a region to an integral over its boundary. The first theorem, which has no name, states that the integral of a gradient of a function along a curve C depends only on the end points $\mathbf{x}_1$ and $\mathbf{x}_2$ of the curve, not on where the curve goes in between. So

$$\int_C \nabla f \cdot \mathbf{dl} = f(\mathbf{x}_2) - f(\mathbf{x}_1) \ . \tag{20}$$

(You are supposed to remember how curve and surface integrals are defined, in particular that **dl** is a vector directed along the tangent vector of the curve.[5]) An equivalent way of stating this result is to say that the integral of a gradient along a closed curve vanishes. (Why is it equivalent?) The wonderful thing is that the converse holds: If the integral along *every* closed curve vanishes the vector field is necessarily a gradient,

$$\oint_C \mathbf{A} \cdot \mathbf{dl} = 0 \quad \Leftrightarrow \quad \mathbf{A} = \nabla f \ , \tag{21}$$

or more completely in words: *The integral of a vector field* $\mathbf{A}$ *vanishes around every closed loop if and only if that vector field is a gradient of a function.* This function is determined up to an arbitrary constant. In the proof one picks an arbitrary point $\mathbf{x}_1$ and defines the function f by means of

$$f(\mathbf{x}) = \int_C^{\mathbf{x}} \mathbf{A} \cdot dl \ , \tag{22}$$

where the curve starts at $\mathbf{x}_1$ and ends at $\mathbf{x}$. The definition does not depend on the choice of the curve because $\mathbf{A}$ is such that its integral along any

[5]Exercise: If you feel unsure, let $f = x^2ye^z$ and let C be the curve $\sigma(t) = (\sin t, \cos t, t)$, $0 \le t \le 2\pi$. Compute $\int_C \nabla f \cdot \mathbf{dl}$, as a curve integral, and check that eq. (20) comes out.

closed curve vanishes, and the dependence on $\mathbf{x}_1$ is only through an additive constant that leaves the gradient of f unaffected.

There is some fine print that we ignore: The functions have to be differentiable and the curve has to be smooth enough so that the integrals exist.

The next theorem is

Stokes' theorem: *For every surface S with boundary C and for every vector field $\mathbf{A}$ there holds*

$$\int_S (\nabla \times \mathbf{A}) \cdot \mathbf{da} = \oint_C \mathbf{A} \cdot \mathbf{dl} \ . \tag{23}$$

The surface area element is $\mathbf{da} = \mathbf{n} \mathrm{d}S$, where $\mathbf{n}$ is the unit normal vector of the surface. There is a sign which is determined by the right hand thumb rule. A closed surface is one without any boundary, and it follows that the integral of the curl of a vector field over any closed surface S vanishes. Again the converse holds: *If the integral over every closed surface vanishes the vector field is necessarily a curl,*

$$\forall S: \ \oint_S \mathbf{B} \cdot \mathbf{da} = 0 \quad \Leftrightarrow \quad \mathbf{B} = \nabla \times \mathbf{A} \ . \tag{24}$$

The vector field $\mathbf{A}$ is determined up to a gradient of an arbitrary function.

You may think that these theorems are useless in practice: for a given vector field we will never be able to test if the integral vanishes around every closed loop, or over every closed surface. But sometimes we can. Stokes' theorem says that the integral over every closed loop will vanish if the vector field has vanishing curl. Then the converse of the first theorem ticks in to show (modulo the fine print) that

$$\nabla \times \mathbf{A} = 0 \quad \Leftrightarrow \quad \mathbf{A} = \nabla f \ . \tag{25}$$

Here there is some fine print that can be important. We have tacitly assumed that every closed curve bounds a surface lying entirely inside the space. This can fail. If space is like the surface of a cylinder then curves winding around the cylinder cannot be shrunk to a point without leaving the cylinder. Such closed curves are said to be *non-contractible*. This is an interesting caveat, and we will actually come across it in our discussion of magnetism.

The third and final theorem is

Gauss' theorem: *For every volume V with boundary S and for every vector field* $\mathbf{B}$ *there holds*

$$\int_V \nabla \cdot \mathbf{B} \, \mathrm{d}V = \oint_S \mathbf{B} \cdot \mathbf{da} \, . \tag{26}$$

The right hand side is known as the *flux* of $\mathbf{B}$ through S.[6] [7] If a vector field $\mathbf{B}$ has vanishing divergence its flux through every closed surface that bounds a volume vanishes, and the converse of Stokes' theorem then ensures that there exists a vector field $\mathbf{A}$ such that

$$\nabla \cdot \mathbf{B} = 0 \quad \Leftrightarrow \quad \mathbf{B} = \nabla \times \mathbf{A} \, . \tag{27}$$

Using Stokes' and Gauss' theorems one can formulate Maxwell's equations in words, and recapture Faraday's intuitions about the electromagnetic field. For instance we can say that "the integral of the electric field along any closed loop plus the time derivative of the flux of a magnetic field through any surface spanned by that loop equals zero".[8]

Field lines and gauge theories

Faraday's intuitions were based on the field lines that are used to visualize the electric and magnetic fields. Here some caution is called for. The electric field (say) does give a vector attached to every point in space, but that vector does not sit "in" space. When we draw it as if it did, we are really depicting the vector along which an electrically charged test particle would accelerate. Still, all of Faraday's intuitions have proved true. It would be hard to convince a plasma physicist that a magnetic field line, vibrating under tension, is not really there.

The mysterious unobservable potentials $\mathbf{A}$ and Φ that lie behind the fields moved to the centre stage when the ideas of electrodynamics developed into Yang–Mills theories, providing accurate theories for the strong and weak

[6] Exercise: Again, if you feel unsure about this, compute the flux of the vector field $\mathbf{v} = 3xy\mathbf{e}_x + xz^2\mathbf{e}_y + y^3\mathbf{e}_z$ through the unit sphere.

[7] Exercise: Assume Gauss' theorem. Prove that $\int \nabla\Phi \, \mathrm{d}V = \oint \Phi \mathbf{n} \mathrm{d}A$, where $\mathbf{n}$ is the normal vector of the (closed) boundary of V.

[8] Exercise: Formulate all of the Maxwell equations as well as the equation of charge conservation in words like this, and draw suitable illustrations.

interactions of elementary particle physics. The precise mathematical language that makes complete sense of all this is provided by the theory of fibre bundles.[9]

Preview: Statics and media

The first five chapters in Jackson are concerned with the time independent field generated by static charges and steady currents. Statics is important because it decides the possible equilibrium states of a dynamical theory. And it can get quite intricate, because we may have only limited knowledge of where the charges inside a piece of material are located. This becomes particularly clear if we consider charges placed on an idealized conductor, inside of which the charges can move freely. In an equilibrium situation the charges do not move, which means that they cannot be subject to any net forces. But the forces are given by Lorentz' equation (7) and depend on the electric field that we are trying to solve for. The key to this problem is in Jackson's Figure I.4, where the integral form of Maxwell's equations is used to study discontinuities in the fields across the boundary of some medium. In equilibrium we must have $\mathbf{E} = \mathbf{0}$ inside the conductor. Charges will pile up on its surface, and the analysis shows that the electric field can have discontinuities there. The details, including the explanation of why the charges stay on the surface of the conductor, are in chapters 1 and 2.[10]

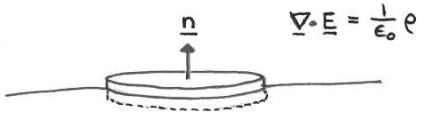

Figure 1: A discontinuity in the electric field at the surface of a conductor, manifesting itself as a surface charge.

Going from idealized conductors to other naturally occurring materials

[9]Reference: For a first introduction see the Appendix in R. Penrose: *Fashion, Faith, and Fantasy*, Princeton UP, 2016. See also C. Rovelli: *Why gauge?*, Found. Phys. **44** (2014) 91.

[10]Exercise: Derive the connection between surface charges and discontinuities in **E**, using eq. (3) and Figure 1.

things become more complex. Only a part of the charge distribution can be controlled externally, while another part is distributed throughout some piece of matter in some manner not known in detail. This leads to the introduction of the fields $\mathbf{D}$ and $\mathbf{H}$, discussed by Jackson in his chapters 4 and 5.

Preview: Dynamics

It is important to recognize that the electromagnetic field lives a life of its own, also when the right hand sides vanish. But to see this we have to move beyond statics. If we set $\rho = 0$ and $\mathbf{J} = 0$ and take a quick look at the equations, something remarkable emerges: Applying a curl to Faraday's law, and recalling that all the right hand sides of Maxwell's equations now vanish, we can derive the equation

$$\nabla^2\mathbf{E} - \frac{1}{c^2}\partial_t^2\mathbf{E} = 0 \ . \tag{28}$$

The components of $\mathbf{E}$ (and by a similar calculation the components of $\mathbf{B}$) obey the *wave equation*, describing the propagation of waves with velocity $c \approx 3 \cdot 10^8$ m/s. The numerical coincidence between the constant c and the velocity of light was noted by Kirchhoff. Maxwell analyzed things further, and then made the comment:

> We can scarcely avoid the inference that light consists in the transverse undulations of some medium which is the cause of electric and magnetic phenomena.

The idea of a medium as a cause of electromagnetic phenomena was eventually dropped.[11] Hertz took a minimalist position:

> To the question "What is Maxwell's theory ?" I know of no shorter or more definite answer than the following: —Maxwell's theory is Maxwell's system of equations.

So equations (1)–(4) and (28) still stand, and optics has become a part of electrodynamics.[12]

[11] Reference: For an interesting account by a participant in the revolution, see A. Einstein and L. Infeld: *The Evolution of Physics*, Simon and Schuster, New York 1938.

[12] Exercise: Derive eq. (28) from eqs. (1)–(4). You will not feel quite the thrill that

We get some insight into the difference between static and dynamic situations if we return to Jackson's Figure I.4. We can just as well apply the argument not to the boundary of a medium but to an arbitrary surface in space, letting ρ and $\mathbf{J}$ vanish for the sake of the argument. Applying Jackson's argument when $\rho = 0$ we find that

$$\nabla \cdot \mathbf{E} = 0 \quad \Rightarrow \quad \mathbf{E}_\perp \text{ is continuous} , \tag{29}$$

where $\mathbf{E}_\perp$ stands for the electric field along the normal of the surface. It is a more delicate matter to study

$$\nabla \times \mathbf{E} = -\partial_t \mathbf{B} . \tag{30}$$

In a static situation the right hand side vanishes, and the conclusion will be that the tangential component $\mathbf{E}_{||}$ is continuous as well. But in a dynamic situation the term $\partial_t \mathbf{B}$ needs watching. It can become infinite, and in fact proportional to a Dirac delta function. If so there will be a contribution to the surface integral over the thin strip in Jackson's argument, and the tangential component of $\mathbf{E}$ will be discontinuous. The conclusion is that Maxwell's vacuum equations allow discontinuous solutions, but only in the time dependent case. A closer investigation will show that these discontinuities move with the speed c.[13]

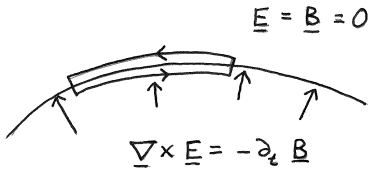

Figure 2: How a discontinuity in the tangential component of the electric field can arise across a moving boundary.

In fact we are familiar with this kind of discontinuities. Switch on the light in a dark room. There will then be a surface in space, moving outwards from the light bulb with the speed of light, along which the electric and magnetic fields are discontinuous. Moreover the discontinuity is of a special

Maxwell felt, but it is nice anyway.

[13]Exercise: Spell out this argument as far as you can, relying on Jackson's page 26.

kind. According to our analysis the normal components of the fields must be continuous, while the components of the fields along the surface, that is to say orthogonal to the direction of the propagation of the wave, can be discontinuous. This is why Maxwell talked about transverse undulations, rather than longitudinal undulations as in sound waves.

Idealizations

Jackson gives a thoughtful account on the limits of applicability of the theory. Let me add that there have been many attempts to make sense of the notion of an electrically charged point particle. The idea is to set

$$\rho(\mathbf{x}, t) = q\delta(\mathbf{x} - \mathbf{x}(t)) \ , \tag{31}$$

where δ denotes Dirac's delta function. This can be regarded as describing a test particle not affecting the electromagnetic field. The trajectory $\mathbf{x}(t)$ of the particle is then determined by Lorentz' equation (7). Alternatively we can regard the trajectory $\mathbf{x}(t)$ of the particle as given and use our expression for ρ, together with a suitable current $\mathbf{J}$, on the right hand side of Maxwell's equations. This works well, for instance when describing the electromagnetic waves emitted by a charge driven by external forces. However, trying to solve Maxwell's and Lorentz' equations together as a coupled system, using this Ansatz, leads to inconsistencies. The theory demands a continuous model for ρ and $\mathbf{J}$. Still we will find the "point particle" useful as a mathematical tool. What we cannot do is to use classical electrodynamics to explain why electrical charge appears in discrete amounts in Nature.

In recent years the centre of gravity in fundamental physics has moved towards classical field theory (and more specifically to general relativity) . As a result there has been significant progress on these questions lately, making the final chapter in Jackson's book look somewhat dated.[14]

Apology

[14]Reference: S. E. Gralla, A. I. Harte, and R. M. Wald, *A rigorous derivation of the electromagnetic self-force*, Phys. Rev. **D80** (2009) 024031.

Still classical electrodynamics is not a very active research area. On the other hand Maxwell's equations, together with a proper (perhaps quantum mechanical) description of charged matter, covers basically all of physics from the scale of atoms to the scale of bacteria, inclusive. At smaller scales the nuclear forces begin to be important. At larger scales gravity kicks in. But if you want to explore any phenomenon in between electrodynamics is the only force that counts. Frictional forces, van der Waals forces, light, electronics, and more, are manifestations of electromagnetism. As a service science, electrodynamics stands supreme. Also, among the fundamental forces, electrodynamics is unique in that there is a large range of observed phenomena for which the classical version is relevant, and a large range for which the quantum version is needed. With the nuclear forces all is quantum, with gravity all is classical.

Fundamental research is often defended by quoting the famous answer to a politician who asked what use one can make of it: "What is the use of a new-born baby?" (Or maybe the answer was: "Sir, one day you will tax it!" The story is often told of Gladstone and Faraday, and if you cannot recall who Gladstone was, never mind). The point is that the story concerns research into the electromagnetic field—and the range of practical applications that this research eventually turned out to have is simply fantastic. Our technological civilization rests on it.

Problem 1: Write a careful account of Cavendish's experiment, described in Jackson's Figure I.1. Emphasize the theory behind it. You may be interested to know that the "draughtsman", mentioned in the caption, was commissioned by Maxwell himself.

ELECTROSTATICS

Statics is important to dynamics because it decides the possible equilibrium states of the theory. The basic ideas of electrostatics are spelled out in Jackson's chapter 1.

In electrostatics the electric field can be regarded as simply a device encoding the force that acts on a static test charge. This force is due to the presence of some other static charges. For one charge placed at the origin there is a radial electric field

$$E_r = \frac{q}{4\pi\epsilon_0}\frac{1}{r^2} \quad \Leftrightarrow \quad \mathbf{E} = \frac{q}{4\pi\epsilon_0}\frac{\mathbf{x}}{r^3} \ . \tag{32}$$

If there are several charges, at positions $\mathbf{x}_i$, one obtains the total electric field at position $\mathbf{x}$ by just adding the electric fields due to each individual charge,

$$\mathbf{E}(\mathbf{x}) = \frac{1}{4\pi\epsilon_0}\sum_i q_i \frac{\mathbf{x}-\mathbf{x}_i}{|\mathbf{x}-\mathbf{x}_i|^3} \ . \tag{33}$$

This is called the Principle of Superposition. For a continuous charge distribution the sum becomes an integral,

$$\mathbf{E}(\mathbf{x}) = \frac{1}{4\pi\epsilon_0}\int d^3x' \ \rho(\mathbf{x}')\frac{\mathbf{x}-\mathbf{x}'}{|\mathbf{x}-\mathbf{x}'|^3} \ . \tag{34}$$

These formulas are well supported by experiments. But since we have already denied the existence of "point charges" it is legitimate to ask what the status of this derivation is. The answer will turn out to be that eq. (34) results from taking the inverse of the Laplace operator using a Green function consistent with the boundary conditions at hand, and then applying this inverse operator to the charge distribution. To see all this we turn to Maxwell's equations.

A static solution of Maxwell's equations

In the time independent case the electric and magnetic fields decouple, so we only have to deal with

$$\begin{cases} \nabla \times \mathbf{E} = \mathbf{0} \\ \\ \nabla \cdot \mathbf{E} = \frac{1}{\epsilon_0}\rho \ . \end{cases} \tag{35}$$

Stokes' theorem says that because the electric field has vanishing curl it is the gradient of some function $\Phi = \Phi(\mathbf{x})$, known as the *electric potential*. We can then rewrite the equations as

$$\begin{cases} \mathbf{E} = -\nabla\Phi \\ \\ \nabla^2\Phi = -\frac{1}{\epsilon_0}\rho \ . \end{cases} \tag{36}$$

The equation for Φ is known as *Poisson's equation*. It is the basic equation in Newtonian gravitation too, except that then ρ has a sign since the mass density is never negative.

The equations are linear in the unknown $\mathbf{E}$, or Φ. Therefore solutions can be added, the Principle of Superposition holds, and it is enough to understand the solution for a single charge, conveniently placed at the origin. We know what $\mathbf{E}$ we want, and a small calculation shows that

$$-\nabla\left(\frac{1}{r}\right) = \frac{\mathbf{x}}{r^3} \ . \tag{37}$$

So we have found the electric potential for the electrostatic field from a point charge. Thus what we need to show is that

$$\Phi = \frac{q}{4\pi\epsilon_0}\frac{1}{r} \quad \Leftrightarrow \quad \nabla^2\Phi = -\frac{1}{\epsilon_0}\rho \ , \tag{38}$$

where ρ is a peculiar charge density describing a single charge q placed at the origin. It vanishes everywhere except at the origin itself, where it is presumably infinite.

It is easy to see that

$$r \neq 0 \quad \Rightarrow \quad \nabla^2\left(\frac{1}{r}\right) = 0 \ . \tag{39}$$

What I mean by "easy to see" is that it is easy to check. We apply the derivatives to the function, collect terms, and find that they cancel.[15] The

[15]Exercise: Check equations (37) and (39), using Cartesian coordinates for the calculation.

terms will have some power of r in the denominator, which means that our equations are ill-defined when $r = 0$. But then for a point charge, ρ is "infinite" at the origin, and this is not (yet) well defined either. We have to do some work in order to fully define ρ for the point charge.

What we do know about ρ is that

$$\begin{cases} r \neq 0 \quad \Rightarrow \quad \rho(\mathbf{x}) = 0 \\ \\ \int_V \rho \, \mathrm{d}V = q \ , \end{cases} \tag{40}$$

where V is any volume containing the origin. So we have to show that $-\epsilon_0 \nabla^2 \Phi$ has the same properties, for the potential Φ we have chosen. But this is easy. We can safely assume that V is a round ball centred at the origin, since the integrand vanishes elsewhere. Then we apply Gauss' theorem to convert the volume integral to an integral over the surface of the ball:

$$\begin{gathered} \int_{\odot} \left(-\epsilon_0 \nabla^2 \Phi\right) \mathrm{d}V = -\frac{q}{4\pi} \int_{\odot} \nabla^2 \left(\frac{1}{r}\right) \mathrm{d}V = \\ = -\frac{q}{4\pi} \oint_S \nabla \left(\frac{1}{r}\right) \cdot \mathbf{da} = \frac{q}{4\pi} \int_0^{\pi} \mathrm{d}\theta \sin\theta \int_0^{2\pi} \mathrm{d}\phi \frac{r^2}{r^2} = q \ . \end{gathered} \tag{41}$$

This does it. The expression $(-\epsilon_0 \nabla^2 \Phi)$ has exactly the same properties as does the charge density ρ we postulated, hence they are equal, and eqs. (35) hold.

Delta functions and distributional solutions

At this point it is convenient to introduce *delta functions*. A delta function, in one dimension, is defined as an object $\delta(x - a)$ such that

$$\int_{-\infty}^{\infty} f(x)\delta(x - a)\mathrm{d}x = f(a) \ , \tag{42}$$

for all functions $f(x)$ subject to some restrictions. A delta function is in itself not a function, it is a *distribution* and always occurs inside an integral. Notice that a delta function does not exist in isolation. It is always defined relative to some particular class of functions $f(x)$ that one allows in the integral.

It is assumed that an integral in which a delta function occurs can be subject to all the usual manipulations such as changes of variables,

$$\int_{-\infty}^{\infty} f(x)\delta(ax)\mathrm{d}x = \int_{-\infty}^{\infty} f(x/a)\delta(x)\frac{\mathrm{d}x}{|a|} = \frac{1}{|a|}f(0) \ . \tag{43}$$

This is written symbolically as $\delta(ax) = \delta(x)/|a|$. We also insist that partial integration is as simple as it can be,

$$\int_{-\infty}^{\infty} \frac{df}{dx}\delta(x)\mathrm{d}x = -\int_{-\infty}^{\infty} f(x)\frac{d}{dx}\delta(x)\mathrm{d}x \ . \tag{44}$$

Reading this backwards gives a meaning to the derivative of a delta function. In fact it can be differentiated an arbitrary number of times. For our purposes the properties of the delta function are given by Jackson on p. 26.[16]

The calculation we did can now be described by saying that we set the charge density of a point charge placed at the origin to

$$\rho(\mathbf{x}) = q\delta(\mathbf{x}) = q\delta(x)\delta(y)\delta(z) \ . \tag{45}$$

(In the second step Cartesian coordinates are used. Changing variables to spherical polar coordinates is of course possible but requires a little care, and we will come back to it.) We then carefully checked that this is exactly what appears on the right hand side of Gauss' law.

We proceeded to show that the function

$$G(\mathbf{x}) = \frac{1}{r} \tag{46}$$

obeys

$$\nabla^2 G(\mathbf{x}) = -4\pi\delta(\mathbf{x}) \ . \tag{47}$$

This is called the *fundamental solution* of Poisson's equation. If we shift the origin it is usually called a *Green function*:

$$G(\mathbf{x},\mathbf{x}') = \frac{1}{|\mathbf{x}-\mathbf{x}'|} \quad \Rightarrow \quad \nabla^2 G(\mathbf{x},\mathbf{x}') = -4\pi\delta(\mathbf{x}-\mathbf{x}') \ . \tag{48}$$

[16]Exercise: Prove entry 5 in Jackson's list of properties of the delta function (p. 26). Do it for some some simple function first, say for $f(x) = ax$.

This is now a function of two points in space, but like a delta function it is an object that is really supposed to sit inside an integral. The point is that with a Green function in hand we can solve Poisson's equation for an arbitrary charge density ρ. The solution is

$$\Phi(\mathbf{x}) = \int \mathrm{d}^3x' \frac{1}{4\pi} G(\mathbf{x}, \mathbf{x}') \frac{\rho(\mathbf{x}')}{\epsilon_0} = \int \mathrm{d}^3x' \frac{1}{4\pi\epsilon_0} \frac{\rho(\mathbf{x}')}{|\mathbf{x} - \mathbf{x}'|} \ . \tag{49}$$

This is an integral that we can do, given ρ. In this sense it is explicit. Applying the gradient operator to it we recover the electrostatic field (34). And if we apply the Laplace operator we get

$$\nabla^2\Phi(\mathbf{x}) = \int \mathrm{d}^3x' \frac{\nabla^2 G(\mathbf{x}, \mathbf{x}')\rho(\mathbf{x}')}{4\pi\epsilon_0} = -\int \mathrm{d}^3x' \frac{\delta(\mathbf{x} - \mathbf{x}')\rho(\mathbf{x}')}{\epsilon_0} = -\frac{\rho(\mathbf{x})}{\epsilon_0} \ . \tag{50}$$

We have found a solution for the electric field generated by a static charge distribution, but it remains to show that the solution is unique. Concerning Green functions we will have more to say later, for now we just observe that—like the delta function—a Green function is really a distribution, not a function. It is always meant to occur inside an integral.

Lower dimensional electrostatics

We take time out for a look in lower dimensions, since it provides useful experience with delta functions. It is also useful when we consider electric charges confined to a very thin wire or a very thin surface. We approximate 'very thin' with 'infinitely thin', and suppose that we are close enough to the line or surface where the charge is concentrated so that we can assume the line to be straight and the surface to be plane. When the only relevant directions are normal to the line, or surface, we are in effect doing electrostatics in a space of only two, or one, dimensions.

Gauss' theorem holds in all dimensions, so for a point charge q (placed at the origin) it remains true that

$$\oint_S \mathbf{E} \cdot \mathbf{da} = \frac{q}{\epsilon_0} \ , \tag{51}$$

where the integration is over a "sphere" surrounding the origin. If the dimension $d = 2$ this sphere is a circle, if $d = 1$ it consists of two points equidistant from the charge. Since the area of the sphere grows like r^{d-1} we conclude that the strength of the electric field falls like

$$E \sim \frac{1}{r^{d-1}} \ . \tag{52}$$

If $d = 2$ we get for the electric field and the potential from a point charge that

$$E \sim \frac{1}{r} \ , \qquad \Phi \sim \ln r \ , \qquad r = \sqrt{x^2 + y^2} \ . \tag{53}$$

This is singular at the origin, but less so than in three dimensions. The same behaviour will be observed close to a line charge in $d = 3$.

If $d = 1$, or if we are close to a surface charge in three dimensions, the electric field does not fall off with distance at all. It still points away from (or towards) the charge though, so it changes sign there. Indeed

$$E = \begin{cases} \frac{q}{2\epsilon_0} & \text{if } x > 0 \\ \\ -\frac{q}{2\epsilon_0} & \text{if } x < 0 \ . \end{cases} \tag{54}$$

This step function is the derivative of the potential, and hence the potential has a kink at the position of the charge. The charge density is the derivative of a step function, hence it vanishes everywhere except at the origin, and there it is "infinite". It looks suspiciously like a delta function, and in fact it is. If we have a step function $\epsilon(x)$ which jumps from $-1/2$ to $+1/2$ at the origin we can do a partial integration to see that

$$\begin{aligned} &\int_{-\infty}^{\infty} f(x) \frac{d\epsilon(x)}{dx} \mathrm{d}x = \frac{1}{2}(f(\infty) - f(-\infty) - \int_{-\infty}^{\infty} \frac{df}{dx} \epsilon(x) \mathrm{d}x = \\ &= \frac{1}{2}(f(\infty) - f(-\infty)) - \frac{1}{2}\left(\int_0^{\infty} \frac{df}{dx} \mathrm{dx} - \int_{-\infty}^{0} \frac{df}{dx} \mathrm{d}x \right) = f(0) \ . \end{aligned} \tag{55}$$

This calculation proves that the derivative of the step function is a perfectly respectable one dimensional delta function, provided only that we restrict

ourselves to functions $f(x)$ that have finite limits as $x \to \pm\infty$. This restriction on the functions that are allowed in the integral is actually part of the definition of the delta function.[17]

Uniqueness of the solution

We now come back to the main track. The physical fact that we build on is Coulomb's law, and nothing more. We want to show that this is equivalent to Maxwell's static equations (35). We have shown consistency. But when we bring in differential equations we also raise the spectre of unphysical solutions. Do Maxwell's equations admit solutions that contradict Coulomb's law, as well as solutions that agree with it?

To investigate this, suppose that we have two different solutions for the same charge distribution. The linearity of the equation then admits the deduction

$$\begin{cases} \nabla^2\Phi_1 = -\frac{1}{\epsilon_0}\rho \\ \\ \nabla^2\Phi_2 = -\frac{1}{\epsilon_0}\rho \end{cases} \quad \Rightarrow \quad \nabla^2(\Phi_1 - \Phi_2) = 0 \ . \tag{56}$$

Thus what we need to prove uniqueness is that

$$\nabla^2\Phi = 0 \quad \Rightarrow \quad \Phi = 0 \ . \tag{57}$$

It will turn out that this is true provided we add the extra condition that $\Phi \to 0$ as $r \to \infty$. This assumption makes physical sense ("we can ignore any charges or electric fields sitting far away from the region of interest"). The fact that extra assumptions have to be brought in is typical of laws formulated in terms of differential equations. Newton's laws do not in themselves predict that the orbit of Mars is an ellipse. You have to add initial conditions saying that at this moment in time Mars is at this position and travels with this velocity. Similarly, extra conditions will be needed to get definite statements out of Poisson's equation. One has to analyze the partial differential equation

[17]Exercise: The Poisson equation $\nabla^2\Phi = 4\pi G\rho_{\rm m}$, where $\rho_{\rm m}$ is the mass density, occurs in Newton's theory of gravity. Choose $\Phi(x,y,z) = -GM(x^2+y^2+(|z|+a)^2)^{-1/2}$, and show that this potential arises from an infinitely thin disk of matter (that is, roughly, a spiral galaxy).

in detail in order to see precisely what kind of conditions that will make the solution unique. In this way partial differential equations (PDEs) differ from ordinary differential equations (ODEs). There are general theorems about existence and uniqueness of solutions that apply to all ODEs. There is nothing quite like that for PDEs. To quote Fritz John, one of the great experts on PDEs:

The constraints imposed by a partial differential equation on its solutions (like those imposed by the environment on a living organism) have an infinite variety of consequences, local and global, equalities and inequalities.

After this dark remark we turn to a study of the *Laplace equation*

$$\nabla^2 \phi = 0 \ . \tag{58}$$

Solutions of this equation are known as *harmonic functions*. We are going to establish a very important property that will enable us to produce the required uniqueness proof with both hands in our pockets.

- The Laplace equation has the Mean Value Property that

$$\nabla^2 \Phi = 0 \quad \Leftrightarrow \quad \Phi(\mathbf{x}) = \frac{\int_S \Phi \mathrm{d}S}{\int_S \mathrm{d}S} \ , \tag{59}$$

  where the integration is over any sphere S centred at $\mathbf{x}$.

At any point the value assumed by Φ equals the average of the values it assumes on a sphere surrounding the point, and this is true for every such sphere.

Suppose first that we have proved this. It follows that

- The Laplace equation obeys the Maximum Principle: if the Laplace equation holds within a bounded volume V then any local maximum or minimum of the function must occur on the boundary of the volume.

This is obvious: if a local maximum occurs at a point in the interior we could surround that point with a very small sphere. On that sphere the values assumed by the function are smaller than the maximum, hence the average over that sphere cannot be equal to the value at the point, which contradicts the Mean Value Property.

- If $\nabla^2\Phi = 0$ in a region and if $\Phi = 0$ on the boundary surrounding that region it must be true that $\Phi = 0$ throughout the region.

This is is true because the Maximum Principle says that the maximum and the minimum of the function are both equal to 0.

- The solution of the Laplace equation in a region is uniquely determined by the values of the function on the boundary surrounding the region.

This follows because the difference of two such solutions vanishes at the boundary, and then it vanishes everywhere.

The values of Φ on the boundary surrounding a region in which the Laplace equation is to be solved are known as *Dirichlet conditions*. Dirichlet conditions determine the solution uniquely. Moreover it turns out that the solution is always very regular (infinitely differentiable, and expressible by convergent power series), except possibly at the very boundary where the Dirichlet conditions are imposed. The boundary conditions themselves are allowed to be quite irregular functions, for instance they need not be continuous.[18]

The problem we want to solve is only a little different. We were imposing the Laplace equation over all space. The claim is that the solution $\Phi = 0$ is unique if the function goes to zero at infinity. To prove that the value of $\Phi(\mathbf{x})$ is smaller than any given ϵ, surround the point $\mathbf{x}$ with a very large sphere. If the sphere is large enough the values assumed by Φ on that sphere will be smaller than ϵ. So will the average over that sphere. By the Mean Value Property it follows that $\Phi(\mathbf{x}) < \epsilon$, and this is true for all ϵ, hence $\Phi = 0$.

It remains to prove the Mean Value Property. But this is easy. First of all it is trivial in one dimension, where vanishing second derivative means that the function is linear. When we average a function Φ over a sphere we obtain a function of the position $\mathbf{x}$ of the centre of the sphere and of its radius r, namely

$$M_\Phi(\mathbf{x}, r) = \frac{\int_S \Phi \mathrm{d}S}{\int_S \mathrm{d}S} = \frac{1}{4\pi} \int \Phi \mathrm{d}\Omega \ , \qquad \mathrm{d}\Omega = \mathrm{d}\theta \mathrm{d}\phi \sin\theta \ . \tag{60}$$

[18]Exercise: An analytic function is a differentiable complex valued function on a two dimensional plane such that it depends only on the particular combination $z = x + iy$, $\Phi(x, y) = \Phi(x + iy) = \Phi(z)$. Prove that every such function is a solution of the two dimensional Laplace equation.

Factors of r cancel since the surface element of the sphere S contains a factor of r^2. To see how the dependence on $\mathbf{x}$ and r comes in we write this formula in full detail,

$$M_\Phi = \frac{1}{4\pi}\int_0^\pi \int_0^{2\pi} \Phi(x + r\cos\phi\sin\theta, y + r\sin\phi\sin\theta, z + r\cos\theta)\mathrm{d}\Omega \ . \quad (61)$$

If we make the sphere very small the average will approach the value at the centre arbitrarily closely,

$$\lim_{r\to 0} M_\Phi(\mathbf{x}, r) = \Phi(\mathbf{x}) \ . \quad (62)$$

Now we take the derivative with respect to r, and afterwards we reinstate the factors of r to have everything written in proper vector form:

$$\partial_r M_\Phi(\mathbf{x}, r) = \frac{1}{4\pi}\int \partial_r \Phi \mathrm{d}\Omega = \frac{1}{4\pi r^2}\int_S \nabla\Phi \cdot \mathbf{da} \ . \quad (63)$$

Inside the integral the derivative with respect to r was traded for the gradient scalared with the unit normal vector of the sphere. Now we apply Gauss' theorem to convert this into an integral over a ball and use the fact that Φ solves the Laplace equation:

$$\partial_r M_\Phi(\mathbf{x}, r) = \frac{1}{4\pi r^2}\int_\odot \nabla^2\Phi \mathrm{d}V = 0 \ . \quad (64)$$

Hence the average does not depend on the radius of the sphere—it equals the average over an arbitrarily small sphere, and hence it equals the value of Φ at the centre. The proof is complete.[19]

Green functions for pedestrians

Green functions play an important role throughout Jackson's book, and a pedestrian look at what they are may be helpful. They are, in fact, inverses of linear partial differential operators. Suppose we are given the equation

[19]Exercise: Create an electric field inside a spherical cavity. You can choose the boundary conditions freely. Can you choose them so that a test charge has a stable equilibrium position at the centre of the cavity? Anywhere in the interior of the cavity?

$$\nabla^2\Phi(\mathbf{x}) = f(\mathbf{x}) \ , \tag{65}$$

where $f(\mathbf{x})$ is a known function. That is, we continue to use the Laplace operator as an example even though the ideas are more general. We want to find an inverse of the operator so that

$$\Phi(\mathbf{x}) = \text{“ } \frac{1}{\nabla^2} f(\mathbf{x}) \text{ ”} = \int' G(\mathbf{x},\mathbf{x}') f(\mathbf{x}') \mathrm{d}^3x' \ . \tag{66}$$

This requires that the Green function G obeys

$$\nabla^2 G(\mathbf{x},\mathbf{x}') = \delta(\mathbf{x},\mathbf{x}') \ . \tag{67}$$

The question is whether such an inverse exists, and if so if it is unique.[20]

The operators are linear, and the functions on which the operators act are best thought of as vectors in some (infinite dimensional) function space. This means that we can get a feeling for things by recalling the conditions under which a matrix M has a unique inverse, that is when we can write

$$\sum_{j=1}^{n} M_{ij} v_j = f_i \quad \Leftrightarrow \quad v_i = \sum_{j=1}^{n} M^{-1}_{ij} f_j \ . \tag{68}$$

Instead of integrals we have finite sums, but the ideas are the same.

The inverse will fail to exist if the determinant is zero. Another way of saying this is that the inverse will fail to exist if there exist vectors $\mathbf{u}$ such that

$$\sum_{j=1}^{n} M_{ij} u_j = 0 \ . \tag{69}$$

What do we do if this happens? If the eigenvectors of M are orthogonal to each other we can restrict ourselves to a *subspace* of the original vector space. We can define a unique inverse matrix M^{-1} restricted to the subspace orthogonal to all zero eigenvectors. With this understanding eq. (68) is perfectly legitimate whenever $\mathbf{f}$ belongs to that subspace. Having defined M^{-1} we get solutions to $M\mathbf{v} = \mathbf{f}$ of the form

[20]Exercise: Rewrite eq. (65) as $\int' M(\mathbf{x},\mathbf{x}')\Phi(\mathbf{x}')\mathrm{d}^3x' = f(\mathbf{x})$ for some distribution $M(\mathbf{x},\mathbf{x}')$. This distribution is the ‘matrix’ that we are trying to invert.

$$v_i = \sum_{j=1}^{n} M^{-1}_{ij} f_j + u_i \ . \tag{70}$$

Here $\mathbf{u}$ is any vector obeying $M\mathbf{u} = \mathbf{0}$, but this ambiguity disappears if we insist that $\mathbf{v}$ must lie in a subspace that contains no such vectors (excepting the zero vector).

For the Laplace operator we do have a problem with zero eigenvalues, because there do exist functions F such that

$$\nabla^2 F(\mathbf{x}) = 0 \ . \tag{71}$$

Therefore the Green function we have used so far is under suspicion. The most general Green function associated to the Laplace operator is of the form

$$G(\mathbf{x}, \mathbf{x}') = \frac{1}{|\mathbf{x} - \mathbf{x}'|} + F(\mathbf{x}, \mathbf{x}') \ . \tag{72}$$

Provided that $F(\mathbf{x}, \mathbf{x}')$ solves the Laplace equation this is a Green function obeying

$$\nabla^2 G(\mathbf{x}, \mathbf{x}') = \nabla'^2 G(\mathbf{x}, \mathbf{x}') = -4\pi\delta(\mathbf{x}, \mathbf{x}') \ . \tag{73}$$

The freedom in choosing the function F reflects the freedom in choosing boundary conditions. Previously we could afford to set $F = 0$ because we were interested in situations where the field vanishes at infinity, and we restricted ourselves to the space of regular functions falling off sufficiently fast at infinity. When you add two such functions you get a third function of this type, so this is indeed a linear subspace of the space of all functions.

In the general boundary value problem we may have to impose other restrictions on the space of allowed functions. Effectively we will use the homogeneous solution F to adapt the Green function to the problem at hand. This may sound obscure when you hear it for the first time, but it will become very clear when we come to the method of images in Jackson's chapter 2.

Another look at the boundary value problem

We discussed conditions under which solutions of the Laplace equation are unique using the Mean Value Principle. This is not only elegant, it is practically useful too, as becomes clear in Jackson's section 1.13 where he presents a numerical method for how to solve the equation. But Jackson himself discusses uniqueness of the solutions along different lines, using Green functions.

First we bring in *Green's identity*. Because of Gauss' theorem there holds, for arbitrary functions f and g on a region V with boundary S, that

$$\int_V (f\nabla^2 g - g\nabla^2 f)\mathrm{d}V = \int_V \nabla\cdot(f\nabla g - g\nabla f)\mathrm{d}V = \oint_S (f\nabla g - g\nabla f)\cdot\mathbf{da}\ . \quad (74)$$

Before making use of it we change the integration variable to $\mathbf{x}'$. Then we set $f(\mathbf{x}') = \Phi(\mathbf{x}')$ and $g(\mathbf{x}') = G(\mathbf{x},\mathbf{x}')$, where

$$\nabla'^2\Phi(\mathbf{x}') = -\frac{1}{\epsilon_0}\rho(\mathbf{x}')\ , \qquad \nabla'^2 G(\mathbf{x},\mathbf{x}') = -4\pi\delta(\mathbf{x},\mathbf{x}')\ . \quad (75)$$

The left hand side of Green's identity then evaluates to

$$-4\pi\Phi(\mathbf{x}) + \frac{1}{\epsilon_0}\int_V G(\mathbf{x},\mathbf{x}')\rho(\mathbf{x}')\mathrm{d}V'\ . \quad (76)$$

The right hand side becomes

$$\oint (\Phi(\mathbf{x}')\nabla' G(\mathbf{x},\mathbf{x}') - G(\mathbf{x},\mathbf{x}')\nabla'\Phi(\mathbf{x}'))\cdot\mathbf{da}'\ . \quad (77)$$

Now comes the trick. We assume that we can choose the so far arbitrary function F in the definition of the Green function so that the Green function vanishes on the boundary S of the region we consider. Then one of the terms in the surface integral vanishes. When we collect things together Green's identity gives us the equation

$$\Phi(\mathbf{x}) = \frac{1}{4\pi\epsilon_0}\int G(\mathbf{x},\mathbf{x}')\rho(\mathbf{x}')\mathrm{d}V' - \frac{1}{4\pi}\oint_S \Phi(\mathbf{x}')\nabla' G(\mathbf{x},\mathbf{x}')\cdot\mathbf{da}'\ . \quad (78)$$

Provided we can actually find the function F that we want, this is a fully explicit formula for the potential inside the volume in terms of the known charge density and the Dirichlet boundary conditions on the potential.

A variation is possible, where we choose the function F so that the normal derivative of the Green function vanishes on the boundary. Then we obtain

$$\Phi(\mathbf{x}) = \frac{1}{4\pi\epsilon_0}\int G(\mathbf{x},\mathbf{x}')\rho(\mathbf{x}')\mathrm{d}V' + \frac{1}{4\pi}\oint_S G(\mathbf{x},\mathbf{x}')\nabla'\Phi(\mathbf{x}')\cdot\mathbf{da}' \ . \qquad (79)$$

This is a fully explicit formula where the potential is determined by the normal derivative of the potential on the boundary, that is to say by *Neumann boundary conditions*. In electrostatics the physics typically requires Dirichlet conditions, but Neumann conditions are important in other theories where the Laplace equation figures (for instance in hydrodynamics, if the velocity field of the fluid can be described as the gradient of a velocity potential).

The bounding surface S may be disconnected—it may for instance consist of two concentric spheres, in which case the region lies in between. And in fact the outer sphere may be moved to infinity, in which case the condition that the potential vanishes at infinity serves as the appropriate Dirichlet condition at that end. We just have to insist that the boundary on which the conditions are set really does enclose the region where we want to solve the equation. We can impose Dirichlet conditions on one part of the boundary and Neumann conditions on another, but trying to impose both conditions at the same point will in general lead to inconsistencies. Evidently so, because once Dirichlet conditions (say) are imposed, the field including the Neumann conditions are completely determined so the latter can no longer be chosen freely.

In practice the explicit formulas in terms of the adapted Green function are difficult to use. If the boundary has an involved shape it will become very hard to solve for the function F, which is what we need to adapt the Green function. But for 'nice' boundaries we will see that the method of images can make life very easy.

Meanwhile we repeat the main mathematical message of this chapter:

Consider a closed surface S enclosing a region. Suppose $\nabla^2\Phi = 0$ throughout this region. Then Φ is uniquely determined by either Dirichlet or Neumann conditions on S.

Problem 2: Consider an arbitrary vector field $\mathbf{V} = \mathbf{V}(\mathbf{x})$. By suitably restricting its behaviour at infinity prove that you can find a function f and a vector field $\mathbf{A}$ such that the equation $\mathbf{V} = \nabla f + \nabla\times\mathbf{A}$ holds.

ELECTROSTATIC AND MAGNETOSTATIC ENERGY

We now start an argument that will add some 'reality' to the electric field. The idea is to associate energy, and eventually momentum, to the field itself. To justify this idea we must go beyond statics, so we include a few things from chapters 5 and 6 as well—although the full justification will have to wait until we have introduced some ideas from relativity theory.

Electrostatic self-force

First let us consider the force experienced by a charged body placed in an electrostatic field $\mathbf{E}(\mathbf{x})$. The body is described by a charge density $\rho(\mathbf{x})$, vanishing outside the body. Lorentz' force law tells us that the total force acting on the body is

$$\mathbf{F} = \int \rho(\mathbf{x})\mathbf{E}(\mathbf{x})\ \mathrm{d}^3x\ . \tag{80}$$

The total electric field is a superposition $\mathbf{E} = \mathbf{E}_{\text{ext}} + \mathbf{E}_{\text{self}}$ of an external field $\mathbf{E}_{ext}$ and the field $\mathbf{E}_{\text{self}}$ created by the body itself. But it is easily seen that the self-field does not contribute to the force. Equation (34) tells us that the self-force is

$$\mathbf{F}_{\text{self}} = \int \rho(\mathbf{x})\mathbf{E}_{\text{self}}(\mathbf{x})\ \mathrm{d}^3x = \frac{1}{4\pi\epsilon_0}\int \frac{\rho(\mathbf{x})\rho(\mathbf{x}')(\mathbf{x}-\mathbf{x}')}{|\mathbf{x}-\mathbf{x}'|^3}\ \mathrm{d}^3x\mathrm{d}^3x' = \mathbf{0}\ . \tag{81}$$

The last step follows because the integrand changes sign when we exchange $\mathbf{x}$ and $\mathbf{x}'$. Hence there is no self-force in electrostatics. The same argument shows that there is no self-force in Newtonian gravity either. In full electrodynamics, where moving charges can radiate, there is a non-trivial self-force that has been understood only fairly recently. But as long as we stay within the confines of electrostatics we can ignore self-forces and self-fields.

Electrostatic energy

We want an expression for the amount of energy stored in a configuration of electric charges. To begin with, the work done on an electric point charge when we move it—slowly, so that it does not radiate—from A to B in an external static electric field $\mathbf{E}$ is easily computed. We do not have to worry about any self-force, and Lorentz' force law gives

$$W = -\int_A^B \mathbf{F}\cdot\mathbf{dl} = -q\int_A^B \mathbf{E}\cdot\mathbf{dl} = q\int_A^B \nabla\Phi\cdot\mathbf{dl} = q(\Phi_B - \Phi_A)\ . \qquad (82)$$

This gives physical meaning to potential differences, but not to the actual value of the potential itself (a point that worried Maxwell somewhat). Note the key assumption that the internal energy of the point charge itself is unchanged while it is being transported from A to B.

We can now ask how much energy one has to supply in order to create a configuration of several point charges by moving them in from infinity (far away). The answer becomes a sum over all pairs of charges,

$$W = \frac{1}{4\pi\epsilon_0}\sum_{\text{pairs}}\frac{q_i q_j}{|\mathbf{x}_i - \mathbf{x}_j|} = \frac{1}{8\pi\epsilon_0}\sum_{i\neq j}\frac{q_i q_j}{|\mathbf{x}_i - \mathbf{x}_j|}\ . \qquad (83)$$

But how to compute the energy stored in a continuous charge distribution? This question leads through deep waters to a suggestive answer. We plunge in with the Ansatz

$$W = \frac{1}{8\pi\epsilon_0}\int\int'\frac{\rho(\mathbf{x})\rho(\mathbf{x}')}{|\mathbf{x}-\mathbf{x}'|}\mathrm{d}^3x\mathrm{d}^3x'\ . \qquad (84)$$

This is the analogue of our double sum for point charges, except that we did not restrict the integrals so as to exclude the interaction of a charge with itself. For the point charges these terms must be excluded if one is to avoid a meaningless answer. In the continuous case we just continue to swim. Using the solution (49) for the electric potential in the first step, we transform the expression to

$$\begin{aligned} W &= \frac{1}{2}\int \rho\ \Phi\ \mathrm{d}^3x = -\frac{\epsilon_0}{2}\int \nabla^2\Phi\Phi\ \mathrm{d}^3x = \\ &= \frac{\epsilon_0}{2}\int \nabla\Phi\cdot\nabla\Phi\ \mathrm{d}^3x = \frac{\epsilon_0}{2}\int \mathbf{E}\cdot\mathbf{E}\ \mathrm{d}^3x\ . \end{aligned} \qquad (85)$$

In the middle of the calculation a partial integration was performed, and it was assumed that the surface term vanishes provided it is evaluated on a very large sphere. Thus there was a faintly non-local flavour to the calculation.

It seems that the electric field *in itself* carries energy, and indeed that there is a *local* energy density

$$w(\mathbf{x}) = \frac{\epsilon_0}{2}\mathbf{E}(\mathbf{x}) \cdot \mathbf{E}(\mathbf{x}) \ . \tag{86}$$

Moreover the new answer is better than eq. (83), because it is never negative.

It is easy to track down the reason for the discrepancy. Equation (83) takes only the interaction energy into account, while our final expression includes self-energy contributions. Indeed, the electric field from two point charges is

$$\mathbf{E} = \frac{q_1}{4\pi\epsilon_0}\frac{\mathbf{x}-\mathbf{x}_1}{|\mathbf{x}-\mathbf{x}_1|^3} + \frac{q_2}{4\pi\epsilon_0}\frac{\mathbf{x}-\mathbf{x}_2}{|\mathbf{x}-\mathbf{x}_2|^3} \ . \tag{87}$$

Squaring that, eq. (86) gives the energy density as

$$w = \frac{1}{32\pi^2\epsilon_0}\left(\frac{q_1^2}{|\mathbf{x}-\mathbf{x}_1|^4} + \frac{q_2^2}{|\mathbf{x}-\mathbf{x}_2|^4} + \frac{2q_1q_2(\mathbf{x}-\mathbf{x}_1)\cdot(\mathbf{x}-\mathbf{x}_2)}{|\mathbf{x}-\mathbf{x}_1|^3|\mathbf{x}-\mathbf{x}_2|^3}\right) \ . \tag{88}$$

The last term is the interaction energy density. Integrating it over all space gives the total interaction energy[21]

$$\begin{aligned} W_{\text{int}} &= \frac{q_1q_2}{16\pi^2\epsilon_0}\int \frac{(\mathbf{x}-\mathbf{x}_1)\cdot(\mathbf{x}-\mathbf{x}_2)}{|\mathbf{x}-\mathbf{x}_1|^3|\mathbf{x}-\mathbf{x}_2|^3}\mathrm{d}^3x = \\ &= \frac{q_1q_2}{16\pi^2\epsilon_0}\int \nabla\left(\frac{1}{|\mathbf{x}-\mathbf{x}_1|}\right)\cdot\nabla\left(\frac{1}{|\mathbf{x}-\mathbf{x}_2|}\right)\mathrm{d}^3x = \\ &= -\frac{q_1q_2}{16\pi^2\epsilon_0}\int \frac{1}{|\mathbf{x}-\mathbf{x}_1|}\nabla^2\left(\frac{1}{|\mathbf{x}-\mathbf{x}_2|}\right)\mathrm{d}^3x = \\ &= \frac{q_1q_2}{4\pi\epsilon_0}\int \frac{1}{|\mathbf{x}-\mathbf{x}_1|}\delta^{(3)}(\mathbf{x}-\mathbf{x}_2)\mathrm{d}^3x = \frac{1}{4\pi\epsilon_0}\frac{q_1q_2}{|\mathbf{x}_1-\mathbf{x}_2|} \ . \end{aligned} \tag{89}$$

[21]Exercise: Justify the step between line two and line three in the calculation that follows.

Here the partial integration led to a surface term that obviously vanishes. The result agrees with eq. (83). If we decide to include the self-energy terms in the energy density, the total energy will diverge. In fact you will encounter the integral

$$W \sim \int_0^\infty \frac{r^2}{r^4} \mathrm{d}r = \infty \ . \tag{90}$$

It would take an infinite amount of energy to squeeze a sensible charge distribution into a point charge. Which is as good a reason as any to exclude point charges from the theory.

It then seems reasonable to adopt the expression

$$W = \frac{\epsilon_0}{2} \int_V \mathbf{E} \cdot \mathbf{E} \ \mathrm{d}V \tag{91}$$

as the energy stored in the charge density that gives rise to the electrostatic field in question.

The capacitance matrix

Here is an example to illustrate some of the preceding ideas. Consider N ideal conductors placed somehow in space, held at constant potentials V_i that can be chosen arbitrarily. We know that outside the conductors there exists a unique solution for a static electric field vanishing at infinity. This can be realized by placing appropriate amounts of charge Q_i on the conductors. We would have to know the detailed geometry of the conductors in order to calculate these charges, but we can make some general statements about them and about the energy stored in the configuration. To begin with, let Φ_j be a solution of the Laplace equation in the empty space outside the conductors, such that $\Phi_j = 1$ on the jth conductor and $\Phi_j = 0$ on the remaining $N-1$ conductors (and such that $\Phi_j \to 0$ at infinity). Then we denote the total charge on the ith conductor in that solution by C_{ij}. The solution we actually want can be found by superposing such solutions. Indeed, if the conductors are held at arbitrary potentials V_j then the charge on the ith conductor is

$$Q_i = \sum_{j=1}^{N} C_{ij} V_j \ . \tag{92}$$

One of our expressions for electrostatic energy tells us that the energy stored in this configuration is

$$W = \frac{1}{2}\sum_{i,j} C_{ij} V_i V_j \ . \tag{93}$$

It is then an exercise to show that the elements of the capacitance matrix obey[22]

$$C_{ij} = C_{ji} \ , \quad C_{ij} = \begin{cases} \geq 0 & \text{if} \quad i = j \\ \\ \leq 0 & \text{if} \quad i \neq j \end{cases} \ , \quad \sum_{j=1}^{N} C_{ij} \geq 0 \ . \tag{94}$$

To prove that the *capacitance matrix* whose matrix elements are the C_{ij} is symmetric you may use Green's identity (74), and you will then understand why this identity is sometimes referred to as *Green's reciprocation theorem.*

A variational principle

The notion of electrostatic energy can be put to immediate use. Let us have another look at the energy or *Dirichlet* integral

$$W[\Phi] = \frac{\epsilon_0}{2}\int_V \mathbf{E}\cdot\mathbf{E}\ \mathrm{d}V = \frac{\epsilon_0}{2}\int_V \nabla\Phi\cdot\nabla\Phi\ \mathrm{d}V \ . \tag{95}$$

In a wide variety of circumstances the static equilibrium state of some system can be found by minimizing its energy function. Suppose that the volume V is free of charges, and that the potential on its boundary S is fixed by Dirichlet boundary conditions. Then we can ask what choice of the function Φ inside V minimizes the energy integral?[23] If you know how to do it, the question is easily answered. We are looking for that function Φ for which

$$I[\Phi + \delta\Phi] - I[\Phi] = 0 \ , \tag{96}$$

[22]Exercise: Prove these statements. To prove the inequalities you can appeal to the maximum principle.

[23]Exercise: Show that a function that minimizes the energy integral cannot have a strict maximum inside the domain V. Then suppose that $\Phi = 0$ on the boundary of V, and show that $\Phi(\mathbf{x}) = 0$ is the only function that minimizes the energy integral.

where the variation $\delta\Phi$ is a small but otherwise arbitrary function on V, except that we assume that $\delta\Phi = 0$ on the boundary because there the value of Φ is fixed by the boundary conditions. With this understanding it is easy to bring the variation to the form

$$I[\Phi + \delta\Phi] - I[\Phi] = \epsilon_0 \int_V \Big(\nabla \cdot (\delta\Phi \nabla\Phi) - \delta\Phi \nabla^2 \Phi \Big) \, \mathrm{d}V \ . \tag{97}$$

(We calculate to first order in $\delta\Phi$. If you are unfamiliar with this kind of calculation, just ignore it and go directly to the result.) Using Gauss' law, and then the fact that $\delta\Phi = 0$ on the boundary S, this is

$$I[\Phi + \delta\Phi] - I[\Phi] = \epsilon_0 \int_S \delta\Phi \nabla\Phi \cdot \mathbf{da} - \epsilon_0 \int_V \delta\Phi \nabla^2 \Phi \mathrm{d}V = - \int_V \epsilon_0 \delta\Phi \nabla^2 \Phi \mathrm{d}V. \tag{98}$$

The energy integral has a minimum for those functions Φ such that the left hand side vanishes. The only way that this can happen for arbitrary choices of the variation $\delta\Phi$ inside the volume is that

$$\nabla^2 \Phi = 0 \ . \tag{99}$$

In this way we see that Maxwell's electrostatic vacuum equation follows from minimization of the electrostatic energy.

In this calculation we assumed Dirichlet boundary conditions. Neumann boundary conditions, in which the normal derivative of the potential on the boundary are specified, can be handled by a slight modification of the argument.[24]

Now let us add a fixed charge density ρ inside the volume V, and look for the minimum of the integral

$$W[\Phi] = \frac{1}{2} \int_V \left(2\Phi\rho - \epsilon_0 \nabla\Phi \cdot \nabla\Phi \right) \mathrm{d}V \ . \tag{100}$$

Assume Dirichlet boundary conditions and repeat the previous argument. We then find that the integral takes its minimum value for those functions Φ that obey Poisson's equation

$$\epsilon_0 \nabla^2 \Phi + \rho = 0 \ . \tag{101}$$

[24]Exercise: In section 1.12 Jackson handles the Neumann conditions by adding a surface term to the energy integral. Spell this out in detail.

This way of looking at things is very useful in situations where V or ρ are such that we are unable to solve Poisson's equation. We just choose *any* function Φ depending on a number of free parameters $\alpha_1, \alpha_2, \ldots, \alpha_n$. Then we compute the integral numerically and get an answer as a function of these parameters. Finally we minimize the resulting function $I = I(\alpha_1, \alpha_2, \ldots, \alpha_n)$ with respect to those parameters. The result will give our best approximation of the true Φ. With a computer we can afford a fairly large number of parameters, and some judgment in the choice of trial functions will often result in strikingly good approximations (as you can see in the example illustrated by Jackson's Figure 1.9.).

The energy integral is useful in another way. You may have noticed that we have avoided the question of the existence of a solution. How general can we make the boundary conditions, and still obtain an analytic harmonic function in the interior? One way to address this question is to focus on the existence of a minimum of the energy integral (95). But it required eminent mathematicians to work out the details, so we stay away from them.

The energy in the magnetic field

In the vacuum case, assuming all time derivatives vanish, the magnetic field obeys

$$\begin{cases} \nabla \times \mathbf{B} = 0 \\ \\ \nabla \cdot \mathbf{B} = 0 \ . \end{cases} \tag{102}$$

This is exactly the same equations as the electrostatic ones (35) when $\rho = 0$, so we expect the energy stored in the magnetic field to be

$$W = \frac{1}{2\mu_0} \int \mathbf{B} \cdot \mathbf{B} \ \mathrm{d}V \ . \tag{103}$$

We have to argue for this conclusion in a different way though, because the electric and magnetic fields couple to charges in very different ways. The magnetic component of the Lorentz force is orthogonal to the direction of motion, and hence it does no work. It follows that eq. (82) remains valid in the presence of a magnetic field, so it does not cost any energy to move a charge around in a purely magnetic field. On the other hand

Ampère's law (4) shows that magnetic fields are created by currents, that is by charges in motion. A time independent magnetic field is not really in a static equilibrium state, it is in a *steady state*. It costs work to build up the currents that are responsible for a magnetic field.

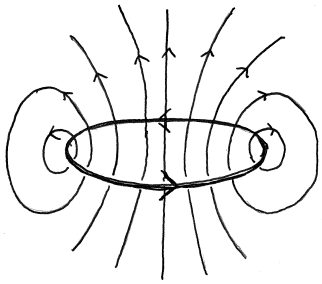

Figure 3: The magnetic field created by a circular current.

We can produce an argument for eq. (103) by considering magnets composed of small circular current loops, with field lines as in Figure 3. When we add to an existing magnetic field by moving such a current loop in from infinity the magnetic field threading through the loop has to change. By Faraday's law (2) this creates an electric field around the loop, which means that there is an emf—an *electromotive force*—acting on the current. To keep the current stable we have to do work against this force. It is this work that is eventually stored in the magnetic field that we are creating.[25] This time we do not have to worry about the self-energy of the charges, but we can no longer make the comfortable assumption that the internal energy of the transported objects stays unchanged during transport. Jackson provides the quantitative details in his section 6.2.

Local conservation of electromagnetic energy

Jackson's section 6.7 gives a simple and convincing argument for the energy densities we have put on the table. Let us look at the local energy density

[25]Exercise: Argue that the field from a current loop must look qualitatively as shown in Figure 3. Also consider the sign of the effect described, that is show that it costs energy to strengthen a magnetic field by moving a current loop in from infinity.

$$w(\mathbf{x},t) = \frac{1}{2}\left(\epsilon_0 \mathbf{E}\cdot\mathbf{E} + \frac{1}{\mu_0}\mathbf{B}\cdot\mathbf{B}\right) \ . \tag{104}$$

To deserve the name "energy density" this quantity must be *locally* conserved, that is to say that any change in total energy inside an arbitrary volume must happen either because work is done on it, or because energy is flowing in or out through the boundary of the volume.

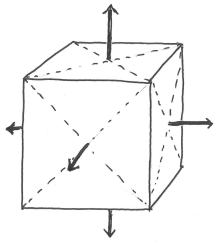

Figure 4: Local energy conservation: energy can flow in or out of every volume element.

A quick calculation using Maxwell's equations shows that[26]

$$\partial_t w = \frac{1}{\mu_0}\nabla\cdot(\mathbf{B}\times\mathbf{E}) - \mathbf{E}\cdot\mathbf{J} \ . \tag{105}$$

Judging from Figure 4 this is exactly what we want. We define *Poynting's vector field*

$$\mathbf{S}(\mathbf{x},t) = \frac{1}{\mu_0}\mathbf{E}\times\mathbf{B} = \epsilon_0 c^2 \mathbf{E}\times\mathbf{B} \ . \tag{106}$$

This can be regarded as quantifying the *energy flow* in a time dependent situation. In this way we obtain local energy conservation in the form

$$\partial_t w + \nabla\cdot\mathbf{S} = -\mathbf{J}\cdot\mathbf{E} \ . \tag{107}$$

The term on the right hand side represents work done *on* the electromagnetic field at a point.

To see the Poynting vector field in action, consider an electric current j through a wire with radius a and conductivity σ. See Figure 5. The electric

[26] Exercise: Do it!

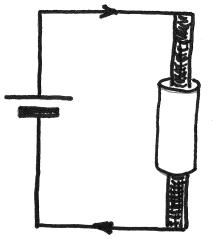

Figure 5: The energy budget in an electric wire. (The radius of the cylinder is slightly exaggerated.) Please add the Poynting vector field to the picture!

field is directed along the wire and obeys *Ohm's law* in the form $j = \sigma E$. The strength of the magnetic field is determined by the flux through the cylinder bounding the wire, and is given by $B = \mu_0 ja/2$. The magnitude of the normal component of the Poynting vector at the surface of the cylinder is then $aj^2/2\sigma$.[27] Integrating over the surface of a cylinder of length l we find

$$\int_C \mathbf{S} \cdot \mathbf{da} = -\frac{aj^2}{2\sigma} 2\pi al = -\frac{j^2}{\sigma}\pi a^2 l = -\int_V \frac{j^2}{\sigma}\, \mathrm{d}V \ . \tag{108}$$

This is the Joule heat dissipated in the wire, and indeed the energy is supplied by the Poynting vector field. It should be noted though that the energy is flowing into the wire from a direction that you may find unexpected.[28] For such reasons one can contemplate some changes in the definition of the Poynting vector fields. Adding the curl of some arbitrary vector field to the Poynting vector will not affect the conservation law (107), since the divergence of a curl is zero. Later on in the course we will appeal to an external judge—general relativity theory—to support Poynting's unchanged definition.

The case for energy carried by the fields themselves is already quite strong. For Maxwell it was natural to think that energy was carried by an elastic ether. When the ether theory was dropped it was suggested by some that energy can be attributed only to a configuration of charges as a whole, but

[27]Exercise: Verify this, and redraw Figure 5 so that it shows how electromagnetic energy is entering the cylinder. Be careful with the logic behind the picture.

[28]Exercise: Place an electric point charge at the centre of the circular current loop shown in Figure 3. Sketch the energy flow, as given by the Poynting vector field.

the notion of localized energy eventually carried the day.[29] For gravity, as described by general relativity, the corresponding questions are still open.

Problem 3: The earth has a negative surface charge, giving rise to an average surface electric field of around 100 V/m reaching some way up in the atmosphere. It also has a dipole magnetic field, and it rotates around an axis aligned with that dipole. Suppose the magnetic field reverses. (Apparently it does, now and then.) How is the rotation rate affected by the surface electric field?

[29]Reference: Two brilliant textbooks taking different sides of the issue are M. Mason and W. Weaver: *The Electromagnetic Field*, Chicago 1929, and M. Abraham and R. Becker: *Classical Electricity and Magnetism*, London 1937.

SOLVING BOUNDARY VALUE PROBLEMS

Jackson devotes chapter 2 to a small catalogue of exact solutions of electrostatic boundary value problems. This is a useful thing to have, before turning to numerical solutions.

Ideal conductors

The method of images illuminates almost all the issues we came across when discussing Green functions and boundary conditions. It is easy to see how boundary conditions can arise. In an electrolyte some of the charges are free to move, and in a metal some of the negative charges are free to move. In an *ideal conductor* all charges are completely free to move. We place a charge outside it, and wait for an equilibrium to set in. But we do not know in advance what the charge distribution will look like on the conductor. We do know, however, that in equilibrium there cannot be any electric field inside, otherwise the charges would start to move and equilibrium would be lost. To prevent this from happening the free charges will move to the surface of the conductor, and stay there. (We will see later why they do not leave the conductor.) In any solution of Maxwell's equations the tangential component of the electric field must be continuous on the surface. This means that once equilibrium has set in the potential will be constant all over the conductor, including its boundary. The surface is now an *equipotential surface*, and we have arrived at a Dirichlet problem. Once we have the unique solution the surface charge distribution can be calculated from the known discontinuities in the normal component of the electric field. (In a real metal, say copper, the surface charge is concentrated not to an infinitely thin shell but to a surface layer that is a few Ångströms across.)

The method of images

The problem is: We place a distribution of charges outside an infinite conducting plane. When the charges inside the conductor have moved to their equilibrium position, what is the electric field?

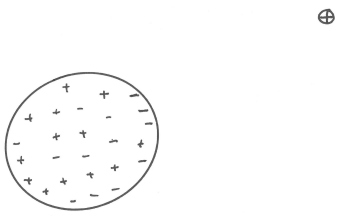

Figure 6: The external charge is fixed. Equilibrium sets in when the charges on the conductor have moved to positions that we do not know in advance.

As stated this looks difficult, because we don't know the equilibrium distribution of the charges. We reformulate the question as: Solve the Poisson equation in the half-space $x > 0$, with the boundary condition $\Phi(0) = 0$. To solve this, it is enough to find the appropriate Green function, or in other words to solve the problem for a single point charge sitting outside the conductor. Recall that the potential $\Phi(\mathbf{x})$ from a point charge placed at $\mathbf{x}'$ is

$$\Phi(\mathbf{x}) = \frac{q}{4\pi\epsilon_0} G(\mathbf{x}, \mathbf{x}') \ , \tag{109}$$

up to an irrelevant additive constant. In this formula the Green function $G(\mathbf{x}, \mathbf{x}')$ must obey the appropriate boundary conditions, that is to say we have to choose the function $F(\mathbf{x})$ in eq. (72) so that the Green function vanishes on the boundary of the region we consider, in particular at $x = 0$. But this is trivial in this example. Let the source point $\mathbf{x}'$ be placed at $x = a$, $y = z = 0$. Then the appropriate inverse of the Laplace operator is

$$G(\mathbf{x}, \mathbf{x}') = \frac{1}{\sqrt{(x-a)^2 + y^2 + z^2}} - \frac{1}{\sqrt{(x+a)^2 + y^2 + z^2}} \ . \tag{110}$$

This vanishes at $x = 0$, it falls to zero at infinity, and the second term is in fact a well defined function solving the Laplace equation throughout the region of space that we are interested in (that is, $x > 0$).

Equation (110) is exactly what we arrive at if we imagine that the conductor is a mirror, with an *image charge* placed behind it. As a matter of fact the electric field vanishes behind the mirror, but we use the field from

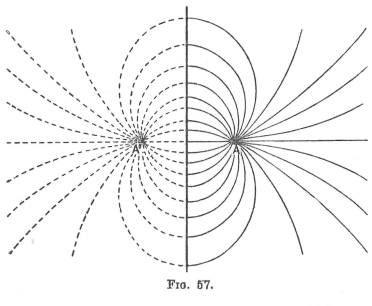

Figure 7: An electric field outside a conducting plane. Just set the part to the left of the plane to zero. (This is Figure 57 in Jeans's book.)

the image charge only in the region outside the conductor. Figure 7 says it all.

With the solution in hand we can easily calculate the equilibrium surface charge density σ on the conductor. Let the external point charge have charge q. Then we simply evaluate

$$\sigma = \epsilon_0(\text{discontinuity of electric field}) = -\epsilon_0 \partial_x \Phi \tag{111}$$

at $x = 0$, obtaining[30]

$$\sigma = -\frac{1}{2\pi} \frac{aq}{(a^2 + y^2 + z^2)^{3/2}} \ . \tag{112}$$

The problem is solved.[31]

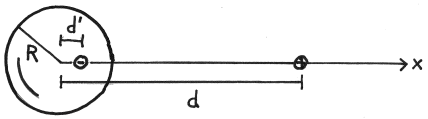

Figure 8: The method of images applied to a spherical conductor.

[30]Exercise: Do the derivation in detail. Also calculate the total surface charge on the conductor.

[31]Exercise: Two conducting planes are placed at $x = 0$ and at $y = 0$. The potential on the planes is kept at zero. Place a single charge somewhere in the positive quadrant of the (x, y)-plane. Use the method of images to calculate the electric potential in that quadrant.

A slightly more involved example is that of a charge placed outside a spherical conductor. Here we proceed using a little guesswork, armed with the knowledge that if we find a solution such that $\Phi = 0$ on the surface of the sphere (and $\Phi \to 0$ at infinity) then we know that it is the only solution. That is to say, a wrong guess will reveal itself, and force us to try again. It seems clear that we should place the image charge on the line connecting the position of the true charge to the centre of the sphere. Keeping both the position and the charge of the image otherwise open, we will then have the potential

$$\Phi = \frac{1}{4\pi\epsilon_0}\left(\frac{q}{\sqrt{(x-d)^2+y^2+z^2}} + \frac{q'}{\sqrt{(x-d')^2+y^2+z^2}}\right) . \tag{113}$$

Then we must choose d' and q' such that $\Phi = 0$ at the surface of the sphere. Since Φ is a function of the position on the sphere, while we have only two parameters to choose, it is not obvious that a solution exists at all. Still we make one further guess. Clearly d' must grow with R and shrink with d. It must also have the dimension of length. The simplest choice is

$$d' = \frac{R^2}{d} . \tag{114}$$

Plug this into eq. (113) and set $x^2 + y^2 + z^2 = R^2$. This will lead to the equation

$$\frac{q}{\sqrt{R^2+d^2-2dx}} = -\frac{q'}{\sqrt{R^2+\frac{R^4}{d^2}-2\frac{R^2}{d}x}} = -\frac{dq'}{R}\frac{1}{\sqrt{d^2+R^2-2dx}} . \tag{115}$$

We are in luck. The solution is

$$q' = -\frac{R}{d}q . \tag{116}$$

A certain resemblance to the theory of Möbius transformations acting on analytic functions will suggest itself here, if you know about these things.[32]

[32]Exercise: Show that a sphere is like a plane. That is, given a unit sphere, place the origin of the coordinate system on its surface, perform the transformation $x_i \to x_i' = x_i/r^2$, and see where the sphere ends up.

There are many other examples where the method of images is useful.[33] Catalogues of problems that can be solved this way have been constructed by taking any electrostatic field, finding its equipotential surfaces, and declaring that they be the surface of some conductor.

The work function

We are now in position to understand why the electric charges on a conductor do not leave its surface to disappear into space. In fact, a piece of metal has a 'work function' telling us how much work that has to be supplied in order to remove electrons from it. Jackson's section 2.3 tells the story in brief.

Consider a spherical conductor with radius R and net charge Q, and a charge q (of the same sign) outside it, sitting fixed at a distance d from the centre of the sphere. The idea is to use the method of images to show that if q is sufficiently close to the surface of the conductor then it is in fact *attracted* by it. To do so we split the total charge on the conductor as

$$Q = q' + (Q - q') \ , \tag{117}$$

where q' is the image charge of the charge outside. The charge q is attracted by the image charge, and repelled by an effective charge

$$Q - q' = Q + \frac{R}{d} q \tag{118}$$

that we can regard as sitting at the centre of the sphere.[34] For the force acting on the charge outside Coulomb's law now gives

$$F = \frac{q}{4\pi\epsilon_0} \left(\frac{Q + \frac{R}{d} q}{d^2} - \frac{\frac{Rq}{d}}{(d - \frac{R^2}{d})^2} \right) \ . \tag{119}$$

We apply a little polish, and obtain

$$F = \frac{1}{4\pi\epsilon_0} \frac{q}{d^2} \left(Q - \frac{R^3 q}{d(d^2 - R^2)^2} (2d^2 - R^2) \right) \ . \tag{120}$$

[33] Exercise: Derive Jackson's eq. (2.13) from his eq. (2.12), in full detail.

[34] Exercise: Exactly why can we do this? Spell out the justification carefully.

When $d >> R$ the first term dominates, and we get the usual Coulomb repulsion. However, when d is close to R the second term dominates. The charge outside is attracted to the image charge, which in this case lies close to the surface of the conductor—or, in reality, the charge is attracted to the surface charge that has built up below it. See Jackson's Figure 2.5 for some quantitative details.[35]

Orthogonal functions and separation of variables

Now for something completely different. Recall that the set of all functions defined on some domain, and obeying suitable conditions, can be regarded as a vector space. In vector spaces the first move is always to define an orthonormal basis. The Dirac notation is particularly clear about this. Denote a column vector by $|V\rangle$, and the corresponding row vector by $\langle V|$. (In quantum mechanics the vectors are usually complex, and a complex conjugate is involved in the definition of the bras. Our vectors are usually real.) An orthonormal basis—or ON basis for short—is defined as a set of vectors $|e_i\rangle$ obeying

$$\langle e_i|e_j\rangle = \delta_{ij} \tag{121}$$

$$\sum_i |e_i\rangle\langle e_i| = \mathbf{1} \ , \tag{122}$$

where $\mathbf{1}$ is the identity operator and the sum has to extend over sufficiently many vectors to ensure that the second condition holds (namely over d vectors, if the dimension of the space is d). With this in place *any* vector can be expressed, uniquely, as a linear combination of the basis vectors:

$$|V\rangle = \sum_i |e_i\rangle\langle e_i|V\rangle = \sum_i c_i|e_i\rangle \ , \qquad c_i = \langle e_i|V\rangle \ . \tag{123}$$

In function spaces the vectors are actually functions $f, g, \ldots$ and so on. Functions (on some specified domain) are vectors because the linear combination $c_1 f + c_2 g$ is again a function (on the same domain). We also need

[35]Exercise: A small charge q sits outside a spherical conductor of large total charge Q and radius R. The charges have the same sign. At (roughly) what distance d from the centre will the charge feel no force?

a definition of the scalar product of two functions, and this will involve an integration over the domain. For concreteness, let us assume that we are dealing with functions defined on the interval $x \in [a, b]$, and define the scalar product as

$$f \cdot g = \int_a^b f^*(x) g(x) \mathrm{d}x \ . \tag{124}$$

Complications can arise here. We must restrict ourselves to functions such that this integral exists. This is still a large, and in fact infinite dimensional, vector space. An ON basis $\{u_n\}_{n=0}^{\infty}$ must therefore have an infinite number of elements. In perfect analogy with eqs. (121)–(122) they must be chosen so that

$$\int_a^b u_n^*(x) u_m(x) \mathrm{d}x = \delta_{n,m} \tag{125}$$

$$\sum_{n=1}^{\infty} u_n(x) u_n^*(x') = \delta(x, x') \ . \tag{126}$$

The delta function appears here because it is in fact the identity operator on the function space—and as we know we may have to impose some restrictions on the allowed functions in order to make sure that the delta function is well defined. If everything works out correctly it is easy to check that any function f can be expressed, uniquely, as[36]

$$f(x) = \sum_{n=1}^{\infty} c_n u_n(x) \ , \qquad c_n = \int_a^b u_n^*(x) f(x) \mathrm{d}x \ . \tag{127}$$

Concerning the various traps that can be there, to do with convergence of sums and integrals, we simply observe that in these notes we will be dealing only with function spaces and sets of orthonormal functions that have been very thoroughly studied, so that in fact no problems arise. (In Jackson's book, the main examples are trigonometric functions, spherical harmonics, and Bessel functions.)

The method of *separation of variables* uses suitable ON bases in the space of functions in order to convert the hard problem of solving a PDE to the comparatively easy problem of solving a set of ordinary differential equations.

[36] Exercise: If you feel the least bit unsure, check this.

It works only if it is possible to adapt the choice of basis functions to the problem—to the Laplace operator and its boundary conditions—in a suitable way. A case that works is that of Cartesian coordinates. So, consider

$$(\partial_x^2 + \partial_y^2 + \partial_z^2)\Phi(x, y, x) = 0 \ . \tag{128}$$

The idea is to look for special solutions, of the form

$$\Phi(x, y, z) = X(x)Y(y)Z(z) \ , \tag{129}$$

and hope that there are sufficiently many such solutions so that one can form an ON basis from them. If so the *general* solution of the Laplace equation can be expressed as a linear combination of such solutions, and the coefficients in the expansion are determined by the boundary conditions. As always we rely on the fact that if we find a solution to the boundary value problem then we already know that is is unique.

For Cartesian coordinates the story is simple. Using our Ansatz we observe that

$$0 = \frac{1}{\Phi}\left(\partial_x^2\Phi + \partial_x^2\Phi + \partial_x^2\Phi\right) = \frac{1}{X}\frac{d^2X}{dx^2} + \frac{1}{Y}\frac{d^2Y}{dy^2} + \frac{1}{Z}\frac{d^2Z}{dz^2} \ . \tag{130}$$

Of the three terms on the right hand side the first is a function only of x, the second only of y, and the third only of z. The only way the sum can vanish is for all three terms to be constant. Thus we arrive at three ordinary differential equations

$$\frac{1}{X}\frac{d^2X}{dx^2} = k_1 \ , \quad \frac{1}{Y}\frac{d^2Y}{dy^2} = k_2 \ , \quad \frac{1}{Z}\frac{d^2Z}{dz^2} = k_3 \ . \tag{131}$$

Plugging the solutions into the Ansatz (129) we obtain solutions of the Laplace equation provided only that the constants obey

$$k_1 + k_2 + k_3 = 0 \ . \tag{132}$$

This is useful whenever the boundary of the region where we want to find a solution to the boundary value problem is simply described in terms of Cartesian coordinates. A rectangular box is an example (discussed in Jackson's section 2.9).

In general the method can be applied if there is a coordinate system adapted to the boundary conditions such that the separation of the variables works when the Laplace operator is expressed in these coordinates. Cartesian coordinates, spherical polars, and quite a few others will do. Note also that the solutions come out as infinite power series, so their usefulness depends on how quickly these series converge.[37]

The electric field outside a sharp edge

An interesting application of the method of separation of variables can be found in Jackson's section 2.11. We imagine that there are two conducting half-planes kept at constant potential V and meeting at an angle β. See Figure 9. We will be interested in the behaviour of the electric field close to the corner, and we will find that the answer depends critically on whether the angle at the corner is acute or obtuse. To get a well defined Dirichlet problem we would need to specify boundary conditions also on a surface surrounding the corner, but we will permit ourselves to keep this vague. This means that our solution will contain some undefined constants.

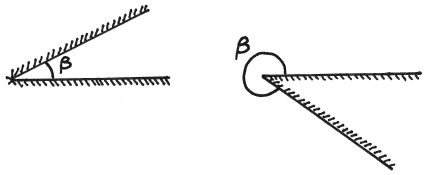

Figure 9: The behaviour of the electric field at the corner, or tip, will depend critically on whether the angle β is acute (at a corner) or obtuse (at a tip).

Cylindrical coordinates (r, ϕ, z) where $x = r\cos\phi$, $y = r\sin\phi$ clearly suggest themselves. The Laplace equation becomes[38]

[37]Exercise: In his Problem 2.16 Jackson gives an explicit solution for the electric potential inside a square, with a constant charge density there. Is this formula practically useful for plotting the potential? If it is, plot it!

[38]Exercise: If you feel unsure about this, derive this from the chain rule $\partial/\partial x = \partial\rho/\partial x\ \partial/\partial\rho + \partial\phi/\partial x\ \partial/\partial\phi$ etc. For your information, there is a better way to do it (using tensor calculus).

$$\nabla^2\Phi = (\partial_x^2 + \partial_y^2 + \partial_z^2)\Phi = \frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial\Phi}{\partial r}\right) + \frac{1}{r^2}\frac{\partial^2\Phi}{\partial\phi^2} + \frac{\partial^2\Phi}{\partial z^2} = 0\ . \tag{133}$$

Evidently the z-direction is of no interest, so we ignore it and treat the problem as two dimensional. Then we try the Ansatz

$$\Phi(x,y) = R(r)\Psi(\phi)\ . \tag{134}$$

With perhaps a little trial and error separation of variables works nicely:

$$\frac{r^2}{\Phi}\nabla^2\Phi = \frac{r}{R}\frac{d}{dr}\left(r\frac{dR}{dr}\right) + \frac{1}{\psi}\frac{d^2\Psi}{d\phi^2} = 0\ . \tag{135}$$

The first term depends only on r, the second only on ϕ, so we obtain two ODEs whose general solutions can be written down by inspection:

$$\frac{1}{\psi}\frac{d^2\Psi}{d\phi^2} = -\nu^2 \quad \Rightarrow \quad \begin{cases} \Psi = A_0 + B_0\phi & \text{if} \quad \nu = 0 \\ \\ \Psi = A_\nu\cos\nu\phi + B_\nu\sin\nu\phi & \text{if} \quad \nu \neq 0 \end{cases} \tag{136}$$

$$\frac{r}{R}\frac{d}{dr}\left(r\frac{dR}{dr}\right) = \nu^2 \qquad \Rightarrow \qquad R(r) = ar^\nu + br^{-\nu}\ . \tag{137}$$

To find the general solution for $R(r)$ we used trial and error, together with the theorem that says that the general solution of a second order ODE has two arbitrary constants, and no more.

The arbitrary constant coming from the separation of variables has been denoted by ν^2, because negative values are not of interest. We see this when we impose the boundary conditions

$$\Phi(r,0) = \Phi(r,\beta) = V\ . \tag{138}$$

First of all this forces $A_\nu = 0$. But it also forces us to assume that ν^2 is non-negative and ν is quantized,

$$\nu = \frac{n\pi}{\beta}\ . \tag{139}$$

Finally we insist that $\Phi(0,\phi) = V$, forcing $b = 0$. This leaves us with an infinite number of solutions of the form $R_n(r)\Psi_n(\phi)$. The solution we want is a superposition

$$\Phi(r,\phi) = V + \sum_{n=1}^{\infty} a_n r^{\frac{n\pi}{\beta}} \sin\left(\frac{n\pi\phi}{\beta}\right) . \tag{140}$$

Had we imposed boundary conditions also on the circle r = constant the constants a_n would have been fully determined.

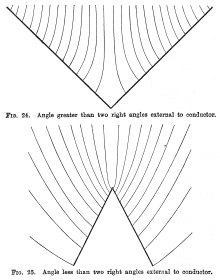

Figure 10: How the electric field lines behave at a corner, or at an edge. (These are Figures 24 and 25 in Jeans's book.)

However, we are interested in the situation close to the corner ($\rho \approx 0$). Unless the neglected boundary conditions are very special this means that the potential there will be given by

$$\Phi(\rho,\phi) \approx V + a_1 r^{\frac{\pi}{\beta}} \sin\frac{\pi\phi}{\beta} , \tag{141}$$

where a_1 is some non-vanishing constant. The potential is not observable, but the electric field is. It will be given by

$$E_r = -\partial_r\Phi \approx -\frac{\pi a_1}{\beta} r^{\frac{\pi}{\beta}-1} \sin\frac{\pi\phi}{\beta} \tag{142}$$

$$E_\phi = -\frac{1}{r}\partial_\phi\Phi \approx -\frac{\pi a_1}{\beta} r^{\frac{\pi}{\beta}-1} \cos\frac{\pi\phi}{\beta} . \tag{143}$$

This is where it matters if the angle is acute ($\beta < \pi$) or obtuse ($\beta > \pi$). In the former case we have a corner, and the electric field goes to zero as

$r \to 0$. In the latter case we have a sharp edge, and the electric field *diverges*. Indeed, by inspection we see that

$$\beta > \pi \quad \Rightarrow \quad \lim_{r\to 0} E_r = \lim_{r\to 0} = \infty \ . \tag{144}$$

The sharper the edge, the stronger the divergence. This can be intuitively understood in terms of the behaviour of the electric field lines. See Figure 10.

Strong electric fields are anything but harmless if they arise in air. A strong electric field will ionize air and cause currents to flow. This is how lightning rods work. Of course lightning rods are three dimensional objects, but we expect the general conclusions to survive also in this dimension. (They do, but the details are a bit complicated. See Jackson's section 3.4.) The field at the tip of a needle-shaped conductor will be very strong. If the needle is surrounded by air there will be ionization of the air, followed by an electric discharge. Lightning has struck.[39]

The intuition behind this result is that electric lines of force crowd together around edges and tips. Using the same pictorial argument you can see that electric field lines will crowd together around any tall object standing on a plain, if there is a thundercloud above. See Figure 11.

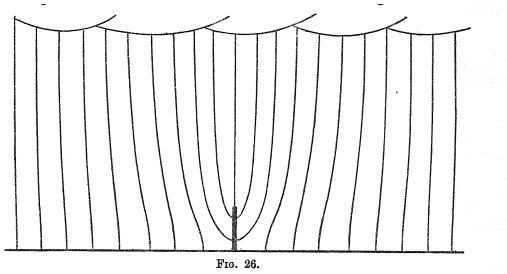

Figure 11: A tall object below a thundercloud. (This is Figure 26 in Jeans's book.)

Problem 4: Consider two conducting spheres, radii a and b, distance between centres c, the first sphere at $\Phi = 0$, the other at $\Phi = V$. Place

[39]Reference: For an interesting discussion of thunderstorms, see *The Feynman Lectures Vol II*, Chapter 9.

imaginary charges inside the spheres until the boundary conditions hold, in a stepwise procedure so that the first imaginary charge is inside the second sphere and ensures that $\Phi = V$ on its surface, the second charge is inside the first sphere and compensates for the first charge so that $\Phi = 0$ on the first sphere. Go on in this way. It is enough to do the first few steps, but after each step you should use a computer to plot the potential at the surface of the first sphere, for suitable values of the parameters. This will give you a feeling for how quickly the procedure converges.

In spherical polar coordinates the Laplace operator becomes

$$\nabla^2 = \partial_r^2 + \frac{2}{r}\partial_r + \frac{1}{r^2}\triangle_S \ , \tag{145}$$

where

$$\triangle_S = \partial_\theta^2 + \cot\theta\partial_\theta + \frac{1}{\sin^2\theta}\partial_\phi^2 \tag{146}$$

is known as the *Laplacian on the unit sphere.*[40] You can derive this in various ways. Let me make a little advertisement for the course in relativity at this point. There you will learn elementary tensor calculus. In particular you will learn that the Laplace equation can be written as

$$\nabla^2 f = \frac{1}{\sqrt{g}}\partial_i\left(\sqrt{g}g^{ij}\partial_j f\right) \ , \tag{147}$$

where g^{ij} is the inverse of a symmetric matrix g_{ij} and g is the determinant of g_{ij}. The matrix g_{ij} is known as the *metric tensor*. If you use Cartesian coordinates in flat space the metric tensor is equal to the Kronecker delta, $g_{ij} = \delta_{ij}$. If you use spherical polar coordinates it is equal to

$$g_{ij} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & r^2 & 0 \\ 0 & 0 & r^2\sin^2\theta \end{pmatrix} \ . \tag{148}$$

Assuming for the sake of the argument that you have taken a relativity course it is trivial to see this, and then you can derive the form of the Laplace equation in your head.[41]

Solutions of the Laplace equation and the rotation group

[40] Exercise: Apply the Laplace operator to r^n and r^{-n-1}, and notice that they behave in the same (simple) way.

[41] Exercise: Suppose you have not taken the relativity course. Just accept my claims, and show that Eq. (147) gives the Laplace equation using spherical polars.

Let us try to find all solutions of the Laplace equation that are regular at the origin. "Regular at the origin" means that the function can be expanded in a power series

$$f(\mathbf{x}) = c_0 + c_i x_i + c_{ij} x_i x_j + c_{ijk} x_i x_j x_k + \dots \ , \tag{149}$$

where Einstein's summation convention was used.[42] In the expansion the terms have been grouped into *homogeneous functions* in a natural way. A function is homogeneous *of order* ℓ if it has the property that

$$f(\lambda x, \lambda y, \lambda z) = \lambda^\ell f(x, y, z) \tag{150}$$

for every real number λ. For instance, the function $f^{(3)}(\mathbf{x}) = c_{ijk} x_i x_j z_k$ is homogeneous of order 3 and conversely every regular function homogeneous of order 3 can be written in this way.

Now suppose that we do a rotation around the origin. Then the function $f(\mathbf{x})$, with all its bumps and features, will turn into a different function $f'(\mathbf{x}) = f(R^{-1}\mathbf{x})$, where R is a *rotation matrix*. I hope it is obvious that if you perform a rotation a function homogeneous of order ℓ will turn into another function that is again homogeneous of order ℓ. It then follows from Eq. (149) that the action of a rotation on our general function f has been broken down so that it acts on small pieces, each of which transform into themselves under rotations.

Now we would like to know how many functions that exist at each 'level' ℓ, that is to say we want to count the number of free parameters at each level. For instance, the most general polynomial homogeneous of order 2 is

$$a_1 x^2 + a_2 y^2 + a_3 z^2 + a_4 xy + a_5 xz + a_6 yz \ . \tag{151}$$

So there are six free parameter in this case. This is the same number of parameters as in a symmetric three by three matrix, not coincidentally because in the expansion (149) the 'matrix' c_{ij} is symmetric.[43] It is actually easy to do the counting at arbitrary ℓ. Indeed, looking at Figure 12 you should be able to convince yourself that there will be

[42] Exercise: Prove that if $f(r, \theta, \phi)$ is a solution of the Laplace equation then so is $F(r, \theta, \phi) = \frac{a}{r} f(\frac{a^2}{r}, \theta, \phi)$, where a is a constant having the dimension of length. Can you see why, if f is regular at the origin, F is said to be regular at infinity?

[43] Exercise: Why is it symmetric? Next, identify the parameters a_i with specific components of c_{ij}. Give a similarly explicit treatment of level 3.

$$1 + 2 + \ldots + (\ell + 1) = \frac{(\ell + 2)(\ell + 1)}{2} \tag{152}$$

coefficients at level ℓ. But now suppose that we insist that the function f is a solution of the Laplace equation. Applying the Laplace operator to a function homogeneous of order ℓ will result in a polynomial of order $\ell - 2$, and by the same argument this polynomial will have $\ell(\ell - 1)/2$ terms.[44] Its coefficients must be set to zero. Hence we have to impose this many conditions on the $(\ell + 2)(\ell + 1)/2$ coefficients we started out with. It follows that the number of free parameters that describe a solution of the Laplace equation homogeneous of order ℓ and regular at the origin is

$$\frac{(\ell + 2)(\ell + 1)}{2} - \frac{\ell(\ell - 1)}{2} = 2\ell + 1 \ . \tag{153}$$

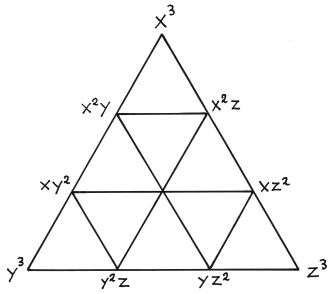

Figure 12: Counting the number of terms in a homogeneous polynomial in three variables.)

A useful way to summarize this discussion is as follows: The space of regular solutions of the Laplace equation is infinite dimensional, since an infinite number of parameters is needed to specify a particular solution. Because the solutions can be superposed this space is a vector space $\mathbf{V}$. Under the action of the rotation group it breaks down into an infinite sequence of vector spaces of dimension $2\ell + 1$, each of which transforms into itself under rotations. Hence we have

$$\mathbf{V} = \mathbf{R}^1 \oplus \mathbf{R}^3 \oplus \mathbf{R}^5 \oplus \mathbf{R}^7 \oplus \ldots \ . \tag{154}$$

[44]Exercise: There is a gap in the logic here. Can you see it? Can you fill it?

In the quantum mechanical theory of angular momentum we are led to consider also half integer values of ℓ, but this can be ignored here.

Spherical harmonics

Having divided the vector space of solutions of the Laplace equation into subspaces $\mathbf{R}^{2\ell+1}$ of dimension $2\ell+1$, we want to find a suitable basis in each subspace. We will make use of the spherical Laplacian $\triangle_S$ for this purpose. Each subspace will be provided with a basis consisting of its eigenfunctions. The eigenfunctions $Y_{\ell,m}$ are known as *spherical harmonics*. They obey

$$\triangle_S Y_{\ell,m} = -\ell(\ell+1)Y_{\ell,m} \ , \tag{155}$$

and they form a complete orthonormal set of functions in terms of which any regular function on the sphere can be expanded.[45] If you did the first exercise in this chapter you can see why such eigenfunctions will be useful when we look for solutions of the Laplace equation in space.

For given ℓ there will be $2\ell+1$ different orthogonal eigenfunctions labelled by m, and there are some choices to be made there. The standard choice uses complex valued functions. There is nothing wrong with that. Suppose you have the usual real and orthogonal vectors $\mathbf{e}_x, \mathbf{e}_y, \mathbf{e}_z$ spanning $\mathbf{R}^3$. You can, if you want, replace two of them with the complex conjugate vectors

$$\mathbf{e}_+ = \frac{1}{2}(\mathbf{e}_x + i\mathbf{e}_y) \ , \qquad \mathbf{e}_- = \frac{1}{2}(\mathbf{e}_x - i\mathbf{e}_y) \ . \tag{156}$$

They are orthonormal in the sense that $\langle e_+|e_-\rangle = \mathbf{e}_+^* \cdot \mathbf{e}_- = \mathbf{0}$, and you can expand any real vector $\mathbf{x}$ as

$$\mathbf{x} = x\mathbf{e}_x + y\mathbf{e}_y + z\mathbf{e}_z = (x-iy)\mathbf{e}_+ + (x+iy)\mathbf{e}_- + z\mathbf{e}_z \ . \tag{157}$$

You see that $\mathbf{x}$ is real because, in the complex basis, two of its components are complex conjugates of each other.

We proceed similarly with the spherical harmonics, which are defined by

[45] Exercise: Suppose that $\psi(\theta,\phi)$ is an eigenfunction of $\triangle_S$. Consider the integral $\int \nabla\psi \cdot \nabla\psi \ \mathrm{d}S$ over the unit sphere, and prove that the eigenvalue must be negative.

$$Y_{\ell,m}(\theta,\phi) = \sqrt{\frac{2\ell+1}{4\pi}\frac{(\ell-m)!}{(\ell+m)!}} P_\ell^m(\cos\theta) e^{im\phi} \ . \tag{158}$$

Some signs may differ from author to author. I use Jackson's conventions. The P_ℓ^m are *associated Legendre polynomials*. $P_\ell^0 = P_\ell$ are the *Legendre polynomials*. Considered as functions of x the latter obey the differential equation

$$\frac{\mathrm{d}}{\mathrm{d}x}\left[(1-x^2)\frac{\mathrm{d}P_\ell}{\mathrm{d}x}\right] + \ell(\ell+1)P_\ell = 0 \ . \tag{159}$$

We intend to set $x = \cos\theta$ and we therefore require the functions to be finite on the interval $[-1,1]$. Then we find (if we do the necessary work) that ℓ has to be a non-negative integer, and then the Legendre polynomials can be defined recursively by

$$P_0(x) = 1 \ , \qquad P_1(x) = x \ , \qquad (\ell+1)P_{\ell+1} = (2\ell+1)xP_\ell - \ell P_{\ell-1} \ . \tag{160}$$

You will have to consult Jackson for the properties of the associated Legendre polynomials. The spherical harmonics are orthonormal functions on the sphere in the sense that

$$\int_0^{2\pi}\int_0^{\pi} Y^*_{\ell',m'}(\theta,\phi) Y_{\ell,m}(\theta,\phi) \ \sin\theta \mathrm{d}\theta\mathrm{d}\phi = \delta_{\ell,\ell'}\delta_{m,m'} \ . \tag{161}$$

Moreover (unless $\ell = 0$) they come in complex conjugate pairs,

$$Y_{\ell,-m} = (-1)^m Y^*_{\ell,m} \ . \tag{162}$$

Here are the first few spherical harmonics explicitly, in spherical polars and in Cartesian coordinates:

$$\begin{array}{lcrcrl}
Y_{0,0} &=& \frac{1}{\sqrt{4\pi}} & & & \\
Y_{1,1} &=& -\sqrt{\frac{3}{8\pi}} & \sin\theta e^{i\phi} & = & -\sqrt{\frac{3}{8\pi}} \ \frac{1}{r}(x+iy) \\
Y_{1,0} &=& \sqrt{\frac{3}{4\pi}} & \cos\theta & = & \sqrt{\frac{3}{4\pi}} \ \frac{1}{r}z \\
Y_{2,2} &=& \sqrt{\frac{15}{32\pi}} & \sin^2\theta e^{2i\phi} & = & \sqrt{\frac{15}{32\pi}} \ \frac{1}{r^2}(x^2-y^2+2ixy) \\
Y_{2,1} &=& -\sqrt{\frac{15}{8\pi}} & \sin\theta\cos\theta e^{i\phi} & = & -\sqrt{\frac{15}{8\pi}} \ \frac{1}{r^2}(xz+iyz) \\
Y_{2,0} &=& \sqrt{\frac{5}{4\pi}} & (\frac{3}{2}\cos^2\theta - \frac{1}{2}) & = & \sqrt{\frac{5}{4\pi}} \ \frac{1}{r^2}(z^2 - \frac{1}{2}x^2 - \frac{1}{2}y^2)
\end{array} \tag{163}$$

The complex conjugate ones are not listed. To see the structure, close your eyes to normalisation factors and overall signs. Notice that

$$Y_{\ell,0}(\theta,\phi) = \sqrt{\frac{2\ell+1}{4\pi}} P_\ell(\cos\theta) \ . \tag{164}$$

The functions $Y_{\ell,0}$, or equivalently the Legendre polynomials, are enough to expand any *axially symmetric* function on the sphere, that is to say any function that is unchanged if you rotate the sphere around the (conventionally chosen) z-axis.[46]

The real functions $Y_{\ell,0}$ are sometimes called *zonal harmonics*, because their zeroes occur at special values of the latitude. When $m \neq 0$ you can form *tesseral harmonics* like $(Y_{1,1} - Y_{1,-1})$ and $i(Y_{1,1} + Y_{1,-1})$, real functions that have zeroes at constant longitude as well.

A generating function for the Legendre polynomials

An interesting way to define the Legendre polynomials is to define a *generating function* of two variables, and then to expand it in a Taylor series in one of them, as follows:

$$g(t,x) \equiv \frac{1}{\sqrt{1-2xt+t^2}} = \sum_{n=0}^{\infty} \frac{\partial_t^n g(0,x)}{n!} t^n \ . \tag{165}$$

Here there is the common abuse of notation that $\partial_t g(0,x)$ means the derivative of $g(t,x)$ with respect to t, evaluated at $t=0$. The series converges if $|t|<1$. Calculating the first few terms one becomes convinced that[47]

$$g(t,x) = \sum_{n=0}^{\infty} P_n(x) t^n \ , \tag{166}$$

where the $P_n(x)$ are the usual Legendre polynomials defined by the recursion relation (160). A proof can be obtained by noting that

$$(1-2xt+t^2)\partial_t g + (t-x)g = 0 \ . \tag{167}$$

[46] Exercise: Given equations (161) and (164), prove that $\int_{-1}^{1} P_\ell(x) P_{\ell'} \mathrm{d}x = 2/(2\ell+1)\delta_{\ell,\ell'}$.

[47] Exercise: Do it, at least up to fourth order.

If we insert the expansions of g and $\partial_t g$ and manipulate the resulting sums a little we find that the coefficients $P_n(x)$ obey the recurrence relation (160) that defines the Legendre polynomials.[48]

Now the generating function $g(t, x)$ should remind you of something.

Expanding the fundamental solution

Consider the fundamental solution of the Laplace equation—the key to electrostatics, if you like. Letting γ be the angle between the vectors $\mathbf{x}$ and $\mathbf{x}'$ it can be written as

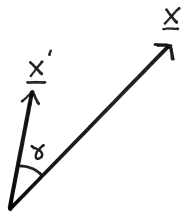

Figure 13: The angle γ. We illustrate the case $r = r_>$, $r' = r_<$.

$$\frac{1}{|\mathbf{x} - \mathbf{x}'|} = \frac{1}{\sqrt{r^2 + r'^2 - 2rr' \cos\gamma}} \,. \tag{168}$$

Suppose that $r > r'$. Then we can rewrite this as

$$\frac{1}{|\mathbf{x} - \mathbf{x}'|} = \frac{1}{r} \frac{1}{\sqrt{1 - 2\cos\gamma \frac{r'}{r} + \left(\frac{r'}{r}\right)^2}} \,. \tag{169}$$

But this is precisely our generating function $g(t, x)$ for the Legendre polynomials, with $x = \cos\gamma$ and $t = r'/r$. So we conclude that

$$\frac{1}{|\mathbf{x} - \mathbf{x}'|} = \frac{1}{r} \sum_{\ell=0}^{\infty} P_\ell(\cos\gamma) \left(\frac{r'}{r}\right)^\ell \,. \tag{170}$$

[48]Exercise: Do it! (Psychologically, the previous exercise is probably more convincing.)

The series converges because we assumed that $r'/r < 1$. It could be that we want to make the opposite assumption, so our result is often written on the form

$$\frac{1}{|\mathbf{x}-\mathbf{x}'|} = \frac{1}{r_>}\sum_{\ell=0}^{\infty} P_\ell(\cos\gamma)\left(\frac{r_<}{r_>}\right)^\ell \ . \tag{171}$$

The notation should be self-explanatory. However, we will mostly be interested in the case when the observation point $\mathbf{x}$ lies far outside the source point $\mathbf{x}'$, so the previous form is good enough for us.

The result (170) has only one drawback, which is that it is not expressed in a suitable coordinate system. What we need is an *addition theorem* for spherical harmonics. *Let the directions of the vectors* $\mathbf{x}$ *and* $\mathbf{x}'$ *be defined by the angles* (θ,ϕ) *and* (θ',ϕ')*, respectively. Then, if* γ *is the angle between the vectors, it holds that*

$$P_\ell(\cos\gamma) = \frac{4\pi}{2\ell+1}\sum_{m=-\ell}^{l} Y^*_{\ell,m}(\theta',\phi')Y_{\ell,m}(\theta,\phi) \ . \tag{172}$$

You will find the proof in Jackson's section 3.6. Let us check if it is reasonable by setting $\ell = 1$. To define γ we introduce two unit vectors,

$$\mathbf{n} = \sin\theta\cos\phi\mathbf{e}_x + \sin\theta\sin\phi\mathbf{e}_y + \cos\theta\mathbf{e}_z \tag{173}$$

and similarly for $\mathbf{n}'$. Then

$$\cos\gamma = \mathbf{n}\cdot\mathbf{n}' = \ldots = \cos\theta\cos\theta' + \sin\theta\sin\theta'\cos(\phi-\phi') \ , \tag{174}$$

where a trigonometric identity was employed in the intermediary step. Using the expressions for Y_{1m} we will be able to confirm that indeed[49]

$$\frac{4\pi}{3}\left(Y^*_{1,-1}(\theta',\phi')Y_{1,-1}(\theta,\phi) + Y^*_{1,1}(\theta',\phi')Y_{1,1}(\theta,\phi) + Y^*_{1,0}(\theta',\phi')Y_{1,0}(\theta,\phi)\right) = \tag{175}$$

$$= \cos\gamma = P_1(\cos\gamma) \ .$$

[49]Exercise: Carry out all these steps.

So the addition theorem seems plausible. For the full proof I refer to Jackson.

The multipole expansion

We are ready to analyze a charge distribution $\rho(\mathbf{x})$. We assume it vanishes outside some sphere of a fixed radius, and fix an origin at the centre of that sphere. We want to know the electric potential and the electric field outside the sphere, and especially at large distances. Hence we would like to expand $\Phi(\mathbf{x})$ as a power series in $1/r$.

We have all the tools. We simply apply the addition theorem to eq. (170) and stick the result into the general solution for the potential:

$$\Phi(\mathbf{x}) = \frac{1}{4\pi\epsilon_0}\int \frac{\rho(\mathbf{x}')}{|\mathbf{x}-\mathbf{x}'|}\ \mathrm{d}^3x' = \tag{176}$$

$$= \frac{1}{\epsilon_0}\sum_{\ell=0}^{\infty}\sum_{m=-\ell}^{\ell}\frac{1}{2\ell+1}\int r'^{\ell}Y^*_{\ell,m}(\theta',\phi')\rho(\mathbf{x}')\ \mathrm{d}^3x'\ \frac{Y_{\ell,m}(\theta,\phi)}{r^{\ell+1}}\ .$$

Whatever the charge distribution is, each integral will produce a number that we can call $q_{\ell m}$. Hence we have shown that the potential admits of an expansion in the form

$$\Phi(\mathbf{x}) = \frac{1}{\epsilon_0}\sum_{\ell=0}^{\infty}\left[\sum_{m=-\ell}^{\ell}\frac{q_{\ell m}}{2\ell+1}\frac{Y_{\ell,m}(\theta,\phi)}{r^{\ell+1}}\right]\ . \tag{177}$$

The series begins with a *monopole term* falling off like $1/r$, then comes a *dipole term* falling off like $1/r^2$, a *quadrupole term* falling off like $1/r^3$, and so on. Some knowledge of Greek will tell you what the higher order terms are called.

An important fact that you should not overlook is that each individual term (with ℓ fixed) is in itself a solution of the Laplace equation.[50]

For many purposes it is preferable to have the result expressed in Cartesian coordinates. We then get the multipole expansion in the form

[50]Exercise: If this is not completely obvious, then verify it!

$$\Phi(\mathbf{x}) = \frac{1}{4\pi\epsilon_0}\left(\frac{q}{r} + \frac{\mathbf{p}\cdot\mathbf{x}}{r^3} + \frac{1}{2}\sum_{i,j} Q_{ij}\frac{x_i x_j - \frac{1}{3}\delta_{ij}r^2}{r^5} + \dots\right) , \tag{178}$$

where the *monopole charge* and the *dipole vector* are

$$q = \int \rho(\mathbf{x})\ \mathrm{d}^3x \quad \text{and} \quad \mathbf{p} = \int \mathbf{x}\rho(\mathbf{x})\ \mathrm{d}^3x\ , \tag{179}$$

and the *quadrupole tensor* is

$$Q_{ij} = \int \left(3x_i x_j - r^2\delta_{ij}\right)\rho(\mathbf{x})\ \mathrm{d}^3x \quad \Rightarrow \quad Q_{ii} = 0\ . \tag{180}$$

We see that the $\ell = 2$ subspace of the space of functions on the sphere is not built from arbitrary second order polynomials in $x/r, y/r, z/r$. To give a solution of the Laplace equation they must be restricted in such a way that the tensor Q_{ij} is *traceless*. This is in fact ensured by the definition (180).[51]

We do not go to higher order in these notes, but many people do. Quantum chemists may need to consider (say) octupole–hexadecapole interactions betwen molecules when they study crystal structures. The Gravity Recovery and Climate Experiment used satellites to monitor the first 2159 multipole moments of the gravitational potential of the Earth, enough to see the time dependence caused by the seasonal melting of the polar caps.

More about the multipoles

It is important to realize that the multipole moments depend on the choice of origin for our coordinate system, so they are to some extent arbitrary. However, the lowest non-vanishing multipole does not depend on the coordinate system, and is an intrinsic property of the charge distribution. As an example, suppose that the monopole term vanishes, and shift the origin with a constant vector $\mathbf{x}_0$. (This is a *passive* coordinate transformation, in the language we will use later.) We then find

[51] Exercise: Given this understanding you can simplify eq. (178). How? Is it a good idea?

$$\mathbf{p} = \int \mathbf{x}\rho(\mathbf{x})\ \mathrm{d}^3x \quad \rightarrow \quad \mathbf{p}' = \int \mathbf{x}\rho(\mathbf{x}+\mathbf{x}_0)\ \mathrm{d}^3x = \int (\mathbf{x}-\mathbf{x}_0)\rho(\mathbf{x})\ \mathrm{d}^3x = \tag{181}$$

$$= \int \mathbf{x}\rho(\mathbf{x})\ \mathrm{d}^3x - \mathbf{x}_0 \int \rho(\mathbf{x})\ \mathrm{d}^3x = \mathbf{p} - q\mathbf{x}_0\ .$$

If the monopole charge q vanishes the dipole moment is unaffected by the change of origin.[52] When, in a book, you see that *the* electric quadrupole moments of atomic nuclei have been measured and tabulated, what is being meant is the quadrupole moment relative to the centre of the charge distribution evaluated in a specified angular momentum eigenstate.[53]

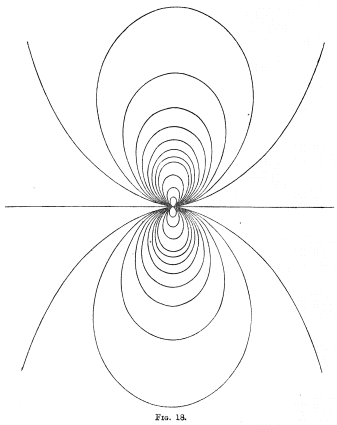

Figure 14: A pure dipole field. (This is Figure 18 in Jeans' book.)

If you know all the multipoles you can reconstruct the charge distribution. If the charge distribution is far away, your detectors will be able to detect only a certain number of terms—first the monopole, then the dipole, and so on. Of course the detector will be sensitive only to the electric field, not to the potential itself. So we have to take the gradient of Φ to obtain the multipole expansion of the electric fied. For the dipole term this results in

[52] Exercise: Prove that if $q = 0$ and $\mathbf{p} = 0$ then the quadrupole tensor is invariant under this change of origin.

[53] You are given the multipole moments relative to a certain origin. Shift the origin using a constant vector $\mathbf{x}_0$. Compute the quadrupole moment relative to the new origin.

$$\mathbf{E} = -\nabla\left(\frac{1}{4\pi\epsilon_0}\frac{\mathbf{p}\cdot\mathbf{x}}{r^3}\right) = \frac{1}{4\pi\epsilon_0}\left(\frac{3\mathbf{x}\ \mathbf{p}\cdot\mathbf{x}}{r^5} + \frac{1}{r^3}\mathbf{p}\right) = \ \Big/\mathbf{n} = \frac{\mathbf{x}}{r}\Big/\ = \\ = \frac{1}{4\pi\epsilon_0}\frac{3\mathbf{n}\ \mathbf{p}\cdot\mathbf{n} - \mathbf{p}}{r^3}\ . \tag{182}$$

It falls with distance as $1/r^3$. This is what a *polar* molecule (with zero total charge) looks like from far away. In itself this electric field is a solution of Maxwell's vacuum equations, and it is interesting to know what it looks like. See Figure 14.

Some intuition for how dipoles behave is useful. If you place one in a homogeneous external electric field it will not be subject to a force, but it will be subject to a torque. You can test your intuition of this behaviour against the expression for the energy of a charge distribution placed in a slowly varying electric field that Jackson derives in his section 4.2. The expression is

$$W = \int \rho(\mathbf{x})\Phi(\mathbf{x})\ \mathrm{d}^3x \approx q\Phi(0) - \mathbf{p}\cdot\mathbf{E}(0) - \frac{1}{6}\sum_{i,j} Q_{ij}\partial_i E_j(0)\ . \tag{183}$$

Here the charge distribution is represented by q, p_i, and Q_{ij}, and its aim in life is to minimize this energy.

For the exercise class:

- Compute the first five Legendre polynomials by applying Gram-Schmidt orthogonalization to the polynomials 1, x, x^2, x^3, x^4.
- Show that you recover the first five Legendre polynomials by expanding the generating function $g(t,x) = (1-2xt+t^2)^{-1/2}$ to fourth order in t.
- Find the electric field outside a spherical conductor placed in a constant electric field in two ways: using the method of images, and using an expansion in spherical harmonics.

- Compute the first non-vanishing multipole moments for i) two charges q at $(\pm a, 0, 0)$, charge $-2q$ at $(0, 0, b)$ ii) four charges q at $(\pm a, \pm a, 0)$, two charges $-2q$ at $(0, 0, \pm b)$. Check your results.

- You have a supply of point charges $\pm q$. Place such point charges at the corners of a regular octahedron. You can make the monopole moment vanish. Can you make the both the monopole and the dipole moment vanish? If so, can you make the quadrupole moment vanish too? Repeat the exercise for a regular cube.

- For a dipole field, locate those points in space where the field points in a direction orthogonal to the dipole vector.

Faraday performed a series of experiments in which he started with a condenser consisting of two conducting plates, and then inserted slabs of insulating materials between the plates. The geometry, charges and potentials are as in Figure 15, and the *capacitance* of the condenser is defined to be

$$C \equiv \frac{Q}{\Delta\Phi} = \frac{Q}{\Phi_2 - \Phi_1} \ . \tag{184}$$

If there is vacuum between the plates Maxwell's equations tell us that

$$C = \frac{Q}{Ed} = \frac{A\epsilon_0}{d} \ . \tag{185}$$

Faraday found that when a slab of insulator was inserted between the plates the capacitance changed to

$$C = \frac{A\epsilon}{d} \ , \tag{186}$$

where the *dielectric constant* ϵ depends on the material and the temperature.[54] We see from Table 1 that $\epsilon > \epsilon_0$.

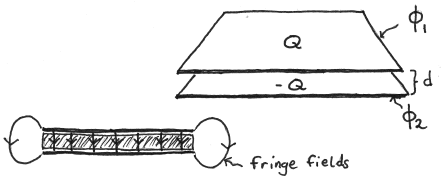

Figure 15: A condenser, with and without a slab of dielectric material inserted.

For a crystalline material such as quartz the dielectric constant depends on how the slab is cut out of the material, in other words it depends on direction. There are also non-linear crystals in which the capacitance, and with it the dielectric constant, depends on the charge Q on the plates and not only on the material itself. These will be ignored here.

[54] Exercise: How would you do the experiment, i.e. how would you control $\Delta\Phi$ and Q?

Table 1: Dielectric constants for some materials.

Substance		ϵ/ϵ_0
Air	gas, 0° C	1.00059
HCl	gas, 0° C	1.0046
Water	gas, 110° C	1.0126
	liquid, 20° C	80
Ammonia	liquid, −34° C	22
Paraffin	solid, 20° C	≈ 2.3
Pyrex glass	solid, 20° C	4.00

The task now is to develop a phenomenological theory for these phenomena, and then to explain why it works, in the sense that we would like to calculate the dielectric constant and its dependence on temperature in terms of the microscopic properties of the material.

The displacement field

An *ad hoc* modification of the theory that covers the facts is

$$\begin{cases} \nabla \cdot \mathbf{D} = \rho \\ \nabla \times \mathbf{E} = 0 \\ \mathbf{D} = \epsilon \mathbf{E} \ . \end{cases} \tag{187}$$

Here ρ is that part of the charge density that is under the control of the experimenter. There are also *bound* charges inside the dielectric material, but they are *not* included in ρ since the material is neutral on average. The field $\mathbf{D}$ is called the *displacement field*, and it is connected to the electric field by a *constitutive relation* $\mathbf{D} = \mathbf{D}(\mathbf{E})$. We have assumed the latter to be just a linear relation. This works well for many materials but fails completely for

a non-linear crystal.

It is still true that $\mathbf{E} = -\nabla\Phi$, and in the set up of the experiment the potential difference can be obtained by integrating along the fringe fields outside the dielectric. However, when $\epsilon > \epsilon_0$ we see that the electric field is actually not as strong as we would expect given the amount of charge Q that has been placed on the plates. If we believe in the molecular theory of the material (not so self evidently true in Faraday's time) it is easy to see why this could be so. The individual molecules are neutral, but they will be distorted to dipoles by the imposed electric field. Hence there is a *dipole density* $\mathbf{P}(\mathbf{x})$ inside the material. Accepting this, we find that the electric potential inside the dielectric must be

$$\Phi(\mathbf{x}) = \frac{1}{4\pi\epsilon_0}\int\left(\frac{\rho(\mathbf{x}')}{|\mathbf{x}-\mathbf{x}'|} + \frac{\mathbf{P}(\mathbf{x}')\cdot(\mathbf{x}-\mathbf{x}')}{|\mathbf{x}-\mathbf{x}'|^3}\right)\mathrm{d}^3x' \ . \tag{188}$$

Here, and in what follows, ρ refers to the charges under our control. There is no net contribution to the charge density from the molecules inside the dielectric anyway. Now we observe that

$$\frac{\mathbf{x}-\mathbf{x}'}{|\mathbf{x}-\mathbf{x}'|^3} = -\frac{\mathbf{x}'-\mathbf{x}}{|\mathbf{x}-\mathbf{x}'|^3} = \nabla'\left(\frac{1}{|\mathbf{x}-\mathbf{x}'|}\right) \ . \tag{189}$$

So we perform a partial integration and drop the surface term since $\mathbf{P}(\mathbf{x}) = 0$ far away. The result is

$$\begin{aligned} \Phi(\mathbf{x}) &= \frac{1}{4\pi\epsilon_0}\int\frac{\rho(\mathbf{x}') - \nabla'\cdot\mathbf{P}(\mathbf{x}')}{|\mathbf{x}-\mathbf{x}'|}\mathrm{d}^3x' \\ &\Rightarrow \\ \nabla\cdot\mathbf{E} &= -\nabla^2\Phi = \frac{1}{\epsilon_0}(\rho - \nabla\cdot\mathbf{P}) \ . \end{aligned} \tag{190}$$

We rewrite this result on the form

$$\nabla\cdot(\epsilon_0\mathbf{E} + \mathbf{P}) = \rho \ , \tag{191}$$

and we are almost done. We define the displacement field as

$$\mathbf{D} = \epsilon_0\mathbf{E} + \mathbf{P} \ , \tag{192}$$

and make the assumption that the polarisation density depends linearly on the electric field,

$$\mathbf{P}(\mathbf{x}) = \epsilon_0 \chi(\mathbf{x}) \mathbf{E}(\mathbf{x}) \tag{193}$$

(as seems reasonable if the material is isotropic and the electric field not too strong). We then arrive at the phenomenological equations (187), with the dielectric constant appearing in the constitutive relation defined by

$$\epsilon(\mathbf{x}) = \epsilon_0 (1 + \chi(\mathbf{x})) \ . \tag{194}$$

Note that the dielectric constant ϵ is allowed to depend on position, otherwise we would be unable to deal with inhomogeneous materials. In the time dependent case it will also depend on the frequency of the applied fields. We will have to return to this point.

We can now see how to explain the experiments with the condenser. Suppose that there is a small gap between the dielectric and the plates, and introduce a volume element that straddles a piece of the boundary of the dielectric. There are no visible charges ρ there, hence $\nabla \cdot \mathbf{D} = 0$. Given this, we find when we integrate over the volume element that

$$\int_V \nabla \cdot \mathbf{E} \ \mathrm{d}V = -\frac{1}{\epsilon_0} \int_V \nabla \cdot \mathbf{P} \ \mathrm{d}V = -\frac{1}{\epsilon_0} \int_{\text{inside}} \mathbf{P} \cdot \mathbf{da} \neq 0 \ . \tag{195}$$

(To picture the volume element, return to Figure 1.) This means that a surface charge is present, even though it is not included in ρ. We can now see why the capacitance went up: Viewed from the inside of the material the total charge on the plates is effectively reduced, and with it the electric field inside the material as well as the potential difference between the plates. Had we used an ideal conductor for the slab of material there would have been no electric field inside, and we would have reached the limit $\epsilon = \infty$.

Electrostatic boundary value problems in media

Provided the dielectric constant is independent of position electrostatics in a medium is no more difficult than electrostatics in a vacuum. In particular the usual uniqueness theorem applies.

A step in the direction of inhomogeneous media is to assume that we are dealing with two regions with different values of the constant ϵ. To deal with this we need to know what discontinuities in the fields that are allowed by the equations. One finds that the normal component of the displacement field is continuous (or has a discontinuity caused by a visible surface charge σ), and similarly the tangential component of the electric field is continuous. In equations, if ϵ_1, ϵ_2 are the dielectric constants of the two media and if the normal vector $\mathbf{n}_{21}$ is directed towards the second medium,

$$(\mathbf{D}_2 - \mathbf{D}_1) \cdot \mathbf{n}_{21} = \sigma \ , \qquad (\mathbf{E}_2 - \mathbf{E}_1) \times \mathbf{n}_{21} = 0 \ . \tag{196}$$

The equations do not demand more than this, which means that the normal component of $\mathbf{E}$ and the tangential component of $\mathbf{D}$ will typically be discontinuous.

It is instructive to reconsider the problem of a single point charge q placed in a medium filling the region $x < 0$ and having dielectric constant ϵ_1, while a medium with dielectric constant ϵ_2 fills the region $x > 0$. We use the method of images. First we write down a solution that is to be valid for $x < 0$. When doing so we are allowed to add an image charge q' in the second medium. Then we write down a solution valid for $x > 0$. There are no charges present in this region, but we are allowed to place an image charge q'' in the first medium, and this image charge does not have to be the same charge as the point charge that is actually there. Assuming for simplicity that all three charges sit at distance d from the 'mirror' we still have two free parameters q' and q'' to play with. Thus equipped we try to meet the continuity requirements (196). If we find a solution we know it is the correct one, because the solution is unique. And we do find a solution. In Figure 16 we see the expected discontinuity of $\mathbf{E}$ in the normal direction at the boundary.[55]

Another problem considered in Jackson's section 4.4 is that of a sphere whose interior has a different dielectric constant than that of its exterior. It is further assumed that the electric field tends to a constant E_0 at infinity. You can test your intuition for how things should behave by asking yourself if the electric field inside the sphere is weaker or stronger than E_0.

[55]Exercise: Go through the details! In Figure 16, on what side of the boundary is the dielectric constant the largest?

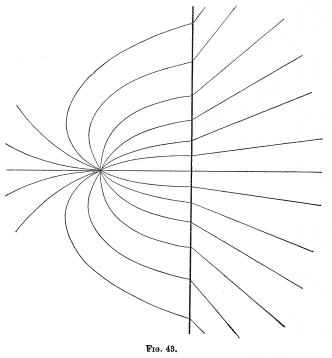

Figure 16: The method of images when there is a medium behind the mirror. (This is Figure 43 in Jeans' book.)

Energy density in dielectrics

We found formula (86) for the energy density in the electrostatic field by considering the energy needed to create a certain configuration of charges. We need to revisit this if we are inside a dielectric. We assume that the dielectric body is rigid, so that no mechanical work comes into the picture. Adding a small amount of charge $\delta\rho$ requires an amount of work

$$\delta W = \int \Phi \delta\rho \, \mathrm{d}V = \int \Phi \delta \nabla \cdot \mathbf{D} \, \mathrm{d}V = \int \Phi \nabla \cdot \delta\mathbf{D} \, \mathrm{d}V \ . \tag{197}$$

We perform a partial integration and drop the surface term (because the charge distribution is assumed to be localized). Then

$$\delta W = -\int \delta\mathbf{D} \cdot \nabla\Phi \, \mathrm{d}V = \int \mathbf{E} \cdot \delta\mathbf{D} \, \mathrm{d}V \ . \tag{198}$$

At this point things can get very complicated because we need a constitutive relation between $\mathbf{D}$ and $\mathbf{E}$. We stick to the simple assumption that $\mathbf{D} = \epsilon\mathbf{E}$, where the dielectric constant may depend on position but does not depend on $\mathbf{E}$. Then we find

$$\mathbf{E} \cdot \delta\mathbf{D} = \epsilon \mathbf{E} \cdot \delta\mathbf{E} = \frac{\epsilon}{2}\delta(\mathbf{E} \cdot \mathbf{E}) = \frac{1}{2}\delta(\mathbf{E} \cdot \mathbf{D}) \ . \tag{199}$$

The total amount of work needed to build up the charge density ρ inside the dielectric is then

$$W = \frac{1}{2}\int \mathbf{E}\cdot\mathbf{D}\ \mathrm{d}V\ . \tag{200}$$

However, we must be careful when we interpret this conclusion.

We have overlooked the fact that the dielectric constant depends on the temperature, which in its turn will be affected by the changing electric field. To make sure that ϵ does not depend on the field we must therefore allow heat to flow into or out of the system in order to keep the process isothermal. Thermodynamically, equation (200) does not refer to the total energy U stored in the field. It refers to the *free energy*

$$F = U - TS\ . \tag{201}$$

The free energy is defined in such a way that isothermal changes in the free energy are caused purely by work, in other words it represents the maximum amount of work that can be extracted from the system. In the usual notation of thermodynamics the work term is given by $\mathbf{E}\cdot\mathrm{d}\mathbf{D}$, and the free energy obeys

$$\mathrm{d}F = \mathrm{d}U - \mathrm{d}(TS) = T\mathrm{d}S + \mathbf{E}\cdot\mathrm{d}\mathbf{D} - \mathrm{d}(TS) = \mathbf{E}\cdot\mathrm{d}\mathbf{D} - S\mathrm{d}T\ . \tag{202}$$

For an isothermal process $\mathrm{d}T = 0$, and only the work term remains.

What is the electric field anyway?

Reflecting on what matter is you realize that it consists of electrically charged electrons and nuclei. This means that the 'true' electric field is subject to dramatic variations once you reach length scales of 10^{-14} metres or thereabouts. The macroscopic field $\mathbf{E}$ that we are dealing with in this chapter shows no such variations. Hence it must be obtained by some kind of averaging procedure. The averaging is spatial. Time averaging is not, or should not be, necessary because microscopic time variations are uncorrelated over the distances involved in the spatial averaging. There are three length scales to be kept in mind, the size d of the atoms, the scale L over which averaging is performed, and the size R of the dielectric body we are interested in. It is assumed that

$$d << L << R \ . \tag{203}$$

It is reasonable to assume that L is around 100 Å, so that a volume of size L^3 contains around a million atoms. This is nicely consistent with the fact that visible light is reflected and refracted by dielectrics as if they were continuous media, while X-rays reveal the microscopic structure. Notice that the size of the volume element is selected so that it is microscopically large—we can perform reliable averaging over it—and macroscopically small— it will be regarded as an infinitesimal volume element when we do calculus with the macroscopic field. Of course it is not *a priori* obvious that this makes sense, but it does, in many circumstances.

The averaging is performed using test functions $f(\mathbf{x})$ that are approximately equal to one in regions of a diameter L, and fall smoothly to zero soon after that. So for an arbitrary function F of space and time we define

$$\langle F(\mathbf{x},t)\rangle = \int \mathrm{d}^3x' f(\mathbf{x}')F(\mathbf{x}-\mathbf{x}',t) \ . \tag{204}$$

Fortunately this averaging is linear in the quantity to be averaged, and therefore it commutes with the taking of derivatives:

$$\partial_i\langle F(\mathbf{x},t)\rangle = \langle \partial_i F(\mathbf{x},t)\rangle \ , \qquad \partial_t\langle F(\mathbf{x},t)\rangle = \langle \partial_t F(\mathbf{x},t)\rangle \ . \tag{205}$$

Hence we can get Maxwell's equations for the averaged field $\mathbf{E}$ by averaging over Maxwell's equations for the microscopic fields. The details are in Jackson's section 6.6.

The averaged macroscopic field $\mathbf{E}$ is relevant for a charged particle passing quickly through the medium, but we will have to think carefully when we try to understand the polarisation density in terms of properties of the individual molecules in the medium. The molecules develop a dipole moment because they are in an electric field, but the electric field that distorts them is quite different from the averaged macroscopic field $\mathbf{E}$.

The field acting on a molecule in the medium

We now take up the challenge of calculating the field that acts on an individual molecule in the material between the condenser plates. The idea is to

surround the molecule with a sphere large enough so that we can treat the charges in the exterior of the sphere with the usual averaging. If $\mathbf{E}$ is the averaged field, then the electric field $\mathbf{E}_{\rm mol}$ felt by the molecule at its centre is

$$\mathbf{E}_{\rm mol} = \mathbf{E} + \left(\begin{array}{c} \text{exact field created by} \\ \text{charges in the ball} \end{array} \right) - \left(\begin{array}{c} \text{contribution to } \mathbf{E} \text{ by} \\ \text{charges in the ball} \end{array} \right) \ . \quad (206)$$

It would seem that the second term on the right hand side would be difficult to compute. However, if we regard the molecules within the ball as electric dipoles all we have to do is to take the average of the contribution from all of them. Notice that when we compute this average we will ignore all surface effects, so taking this average is rather different from calculating the total average field $\mathbf{E}$.

An electric dipole within the ball gives rise to a dipole field

$$\mathbf{E}_{\rm dipole} = -\frac{1}{4\pi\epsilon_0}\left(\frac{r^2\mathbf{p} - 3\mathbf{x}\ \mathbf{p}\cdot\mathbf{x}}{r^5}\right) \ . \quad (207)$$

Let us look at its x-component

$$E_x = -\frac{1}{4\pi\epsilon_0}\frac{1}{r^5}\left((y^2+z^2-2x^2)p_x - 3xyp_y - 3xzp_z\right) \ . \quad (208)$$

Now we take the average $\langle E_x \rangle$. If the medium is isotropic no particular direction is singled out by the average, and it must hold that

$$\langle x^2 \rangle = \langle y^2 \rangle = \langle z^2 \rangle \ , \qquad \langle xy \rangle = \langle xz \rangle = 0 \ . \quad (209)$$

Hence this averaged field is zero in an isotropic medium. Jackson also cites a calculation showing that it vanishes in a crystal with cubic symmetry. Hence we set the second term on the right hand side of equation (206) to zero.

The third term on the right hand side was computed in Jackson's section 4.1. Let the ball have radius L and let it enclose all the relevant charges. For the purposes of this calculation let us change the notation and let $\mathbf{E}$ be the electric field that we want to average over the ball. If there are no charges present the average equals the value taken by $\mathbf{E}$ at the centre of the ball, whatever that is.[56] If there is a non-vanishing charge density ρ within

[56] Exercise: Why?

the ball we must perform an integral that we can transform into a surface integral because the integrand is a gradient:

$$\int_{\text{ball}} \mathbf{E}\ \mathrm{d}V = -\int_{\text{ball}} \nabla\Phi\ \mathrm{d}V = -\int_{\text{sphere}} \Phi R^2 \mathbf{n} \mathrm{d}\Omega\ . \tag{210}$$

You may recall the statement from an exercise in connection with Gauss' theorem. Actually there are three integrals, one for each component. As usual

$$\Phi(\mathbf{x}) = \frac{1}{4\pi\epsilon_0} \int \mathrm{d}^3x'\ \frac{\rho(\mathbf{x}')}{|\mathbf{x}-\mathbf{x}'|}\ . \tag{211}$$

We obtain

$$\int_{\text{ball}} \mathbf{E}\ \mathrm{d}V = -\frac{R^2}{4\pi\epsilon_0} \int \mathrm{d}^3x'\ \rho(\mathbf{x}') \int \frac{\mathbf{n}\ \mathrm{d}\Omega}{|\mathbf{x}-\mathbf{x}'|}\ . \tag{212}$$

Now we observe that the components of the normal vector to the sphere, $\mathbf{n}$, lie in the $l = 1$ subspace of the space of functions on the sphere. So in the rightmost integral we can expand $1/|\mathbf{x}-\mathbf{x}'|$ in spherical harmonics and be confident that only the $l = 1$ term contributes to the integral. Thus

$$\int \frac{\mathbf{n}\ \mathrm{d}\Omega}{|\mathbf{x}-\mathbf{x}'|} = \frac{r'}{R^2} \int \mathbf{n} \cos\gamma\ \mathrm{d}\Omega\ . \tag{213}$$

If we remember the formulas for $\mathbf{n}$, $\cos\gamma$, and $\mathrm{d}\Omega$ a short calculation gives[57]

$$\int \mathbf{n} \cos\gamma\ \mathrm{d}\Omega = \frac{4\pi}{3} \frac{\mathbf{x}'}{r'}\ . \tag{214}$$

Putting it all together

$$\int_{\text{ball}} \mathbf{E}\ \mathrm{d}V = -\frac{1}{3\epsilon_0} \int \rho(\mathbf{x}')\mathbf{x}'\ \mathrm{d}^3x' = -\frac{1}{3\epsilon_0}\ \mathbf{p}\ , \tag{215}$$

where $\mathbf{p}$ is the dipole moment of the charges with respect to the centre of the ball.

We now return to eq. (206), and revert to the notation that $\mathbf{E}$ denotes the averaged field. When we average over the ball we must divide by its volume, so we obtain

[57] Exercise: Do it!

$$\mathbf{E}_{\rm mol} = \mathbf{E} + 0 - \frac{\frac{-1}{3\epsilon_0}\mathbf{P}}{\frac{4\pi L^3}{3}} \ , \tag{216}$$

or better, if we introduce the polarisation density $\mathbf{P}$,

$$\mathbf{E}_{\rm mol} = \mathbf{E} + \frac{1}{3\epsilon_0}\mathbf{P} \ . \tag{217}$$

This is the formula we will work with in order to build a microscopic theory of the dielectric constant.

Microscopic theory of the dielectric constant

We are now ready to start the development of a theory to explain the numbers that appear in Table 1. We assume that the medium is built out of molecules. There are two main cases to consider: *Non-polar* molecules like O_2 and N_2 that have zero dipole moments in the absence of an external field, and *polar* molecules like H_2O that have non-zero intrinsic dipole moments. If we have a large number of polar molecules in zero field their dipole moments will average to zero, but if an external field is present they will align themselves to produce a non-zero average. The alignment can be destroyed by collisions, which means that it will be a function of temperature. This case is briefly discussed in Jackson's section 4.6. Here we discuss only the case of non–polar molecules.

To get a first feeling for how much the molecules are likely to be deformed by the applied electric field, let us model a hydrogen atom as a homogeneous sphere of negative charge $-q$ and a moveable nucleus of charge $+q$. Suppose the radius of the sphere is $a = 1$ Ångström. If the nucleus is displaced to sit at distance b from the centre it will be attracted back to the centre by an electric field whose strength comes out to be, in SI units,[58]

$$E \approx 10^{11} \cdot \frac{b}{a} \ \ \frac{\rm V}{\rm m} \ . \tag{218}$$

This is huge. As it happens the strongest electric field so far generated in the lab was about 10^{11} Volts per metre. It was produced in a volume of one

[58]Exercise: Check this calculation.

cubic micrometre by focused laser beams, and lasted for around a picosecond. The comforting conclusion is that under ordinary circumstances the atoms and the molecules inside the medium are likely to be only slightly deformed. Thus the dipole approximation is likely to be very accurate and the effect likely to be linear in the electric field.

For an individual molecule we make the Ansatz that its dipole moment will be given by

$$\mathbf{p}_{\text{mol}} = \epsilon_0 \gamma \mathbf{E}_{\text{mol}} \ . \tag{219}$$

We can ask a quantum chemist to calculate γ for the molecule in question, but we will be able to make interesting predictions by simply assuming that such a constant exists. The polarisation density in the medium is then

$$\mathbf{P} = N \langle \mathbf{p}_{\text{mol}} \rangle \tag{220}$$

where N is the number of molecules per unit volume. We also know from eq. (217) how to relate $\mathbf{E}_{\text{mol}}$ to the averaged electric field $\mathbf{E}$. So we obtain the equation

$$\mathbf{P} = N\gamma \left(\epsilon_0 \mathbf{E} + \frac{1}{3} \mathbf{P} \right) \ . \tag{221}$$

In the macroscopic theory we related $\mathbf{E}$ and $\mathbf{P}$ by the equation $\mathbf{P} = \epsilon_0 \chi_e \mathbf{E}$, so we can now read off that

$$\chi_e = \frac{N\gamma}{1 - \frac{1}{3} N\gamma} \ . \tag{222}$$

Recalling that $\epsilon = \epsilon_0 (1 + \chi_e)$ this can be rewritten as

$$\gamma = \frac{3}{N} \frac{\epsilon - \epsilon_0}{\epsilon + 2\epsilon_0} \ . \tag{223}$$

This is known as the *Clausius–Mossotti equation*. Even if we do not know the value of the molecular constant γ we can now make predictions about how the dielectric constant varies with density.

You may be slightly worried about what happens at high densities, when $N\gamma$ reaches 3. The answer is that the simple assumptions we made when we derived the equation will fail well before we reach that point.

Problem 5: For pure nitrogen the following data were obtained (in 1934):[59]

Temperature (°C):	23.8	23.8	23.8	23.8
Pressure (atm)	1.02	57.5	221.6	1011.6
Density (kg/m^3)	1.18	66.04	236.1	578.0
Dielectric constant	1.00052	1.03109	1.11413	1.29633

Check how well this agrees with the Clausius-Mossotti equation. How well does the ideal gas law work at these pressures?

[59]Reference: For a slightly earlier review see A. Michels and C. Michels, Phil. Trans. Roy. Soc. **A231** 409.

MAGNETOSTATICS

The equations of magnetostatics are

$$\begin{cases} \nabla \times \mathbf{B} = \mu_o \mathbf{J} \\ \\ \nabla \cdot \mathbf{B} = 0 \ . \end{cases} \quad . \tag{224}$$

Note that $\nabla \cdot \mathbf{J} = 0$, so the currents have to form closed loops (or perhaps they disappear to infinity). To these equations we add the Lorentz force acting on an electrically charged test particle,

$$\mathbf{F} = q\mathbf{v} \times \mathbf{B} \ . \tag{225}$$

Notice the dramatic departure from old ideas about how forces act. The force is not central and it does no work on the test particle.

In electrostatics the first move is to solve one of the equations for the electric field in terms of the potential function. We proceed similarly here:

$$\nabla \cdot \mathbf{B} = 0 \quad \Rightarrow \quad \oint_S \mathbf{B} \cdot \mathbf{da} = 0 \quad \Rightarrow \quad \mathbf{B} = \nabla \times \mathbf{A} \tag{226}$$

for all closed surfaces S and for some vector field $\mathbf{A}$ called the *vector potential*. Like the electric potential it is not a directly observable object, but the ambiguities in its definition are more serious. In fact, given any function $\Psi(\mathbf{x})$, the vector field

$$\mathbf{A}'(\mathbf{x}) = \mathbf{A}(\mathbf{x}) + \nabla\Psi(\mathbf{x}) \tag{227}$$

gives rise to the same magnetic field as does $\mathbf{A}(\mathbf{x})$, and hence it gives rise to the same observable force on test particles. The transformation $\mathbf{A} \to \mathbf{A}'$ is known as a *gauge transformation*. It leaves the magnetic field invariant, or in other words the magnetic field is *gauge invariant*. The integral of the vector potential around any closed loop in space is gauge invariant as well, and by hypothesis so are all physical effects.

The option to change the vector potential with the gradient of an arbitrary function can be put to immediate use. Suppose that $\nabla \cdot \mathbf{A} \neq 0$. Performing a gauge transformation to a physically equivalent vector potential $\mathbf{A}'$ we find

$$\nabla \cdot \mathbf{A}' = \nabla \cdot \mathbf{A} + \nabla^2 \Psi \ . \tag{228}$$

Let us choose the arbitrary function Ψ to be

$$\Psi(\mathbf{x}) = \frac{1}{4\pi} \int \frac{\nabla' \cdot \mathbf{A}(\mathbf{x}')}{|\mathbf{x} - \mathbf{x}'|} \mathrm{d}^3 x' \quad \Rightarrow \quad \nabla^2 \Psi = -\nabla \cdot \mathbf{A} \ . \tag{229}$$

With this choice

$$\nabla \cdot \mathbf{A}'(\mathbf{x}) = 0 \ . \tag{230}$$

That is to say, in the equivalence class of vector potentials that give rise to the same magnetic field we have found one representative whose divergence vanishes. We do not have to go through the work of actually finding a function Ψ. It is enough to know that we *can* make make this choice, which is known as *imposing the Coulomb gauge*.

From now until further notice we assume that we have imposed the Coulomb gauge. Hence

$$\nabla \cdot \mathbf{A} = 0 \ . \tag{231}$$

Having disposed of the equation $\nabla \cdot \mathbf{B} = 0$, and having imposed the Coulomb gauge, the remaining magnetostatic equation is

$$\nabla \times \mathbf{B} = \nabla \times (\nabla \times \mathbf{A}) = \nabla(\nabla \cdot \mathbf{A}) - \nabla^2 \mathbf{A} = -\nabla^2 \mathbf{A} = \mu_o \mathbf{J} \ . \tag{232}$$

This is just three copies of Poisson's equation for the vector potential. Provided the current vanishes at infinity we know that the unique solution is

$$\mathbf{A}(\mathbf{x}) = \frac{\mu_0}{4\pi} \int \frac{\mathbf{J}(\mathbf{x}')}{|\mathbf{x} - \mathbf{x}'|} \mathrm{d}^3 x' \ . \tag{233}$$

The magnetic field is easily computed using the formula $\mathbf{B} = \nabla \times \mathbf{A}$. This is—in a way—the general and complete solution of the equations of magnetostatics. But as you know from the theory of electrostatics the story has barely begun.

Two simple examples

As the first example of how to compute the magnetic field from a fixed current, let us consider a current running along a straight line along the z-axis. Thus

$$J_z(\mathbf{x}) = I\delta(x)\delta(y) \ , \qquad J_x = J_y = 0 \ . \tag{234}$$

This example suffers from the problem that the current does not vanish at infinity. Hence we have to keep our eyes open when we proceed.

The magnetic field is most easily found by noting that the field lines have to form circles around the z-axis, and the strength of the magnetic field is determined by the current passing through a surface spanned by such a circle. Indeed, using Stokes' theorem,

$$I = \int_S \mathbf{J}\cdot\mathbf{da} = \frac{1}{\mu_0}\int_S (\nabla\times\mathbf{B})\cdot\mathbf{da} = \frac{1}{\mu_0}\int_{\partial S}\mathbf{B}\cdot\mathbf{dl} = \frac{2\pi r}{\mu_0}B_\phi \ . \tag{235}$$

So the solution is

$$B_\phi = \frac{I\mu_0}{2\pi}\frac{1}{r} \ , \qquad B_r = B_z = 0 \ . \tag{236}$$

The picture is clear.

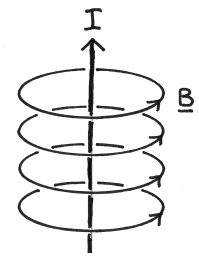

Figure 17: Magnetic field around an infinitely long straight current.

The general solution (233) was derived under the assumption that the current distribution is localised, so there may be a problem if we apply it to the present case. It turns out that the problem is mild, and it can be dealt with. With an arbitrary constant added eq. (233) gives for the only non-vanishing component of the vector potential that

$$A_z(\mathbf{x}) = \frac{\mu_0}{4\pi}\int \frac{I\delta(x')\delta(y')\mathrm{d}x'\mathrm{d}y'\mathrm{d}z'}{\sqrt{(x-x')^2+(y-y')^2+(z-z')^2}} + \text{constant} = \tag{237}$$

$$= \frac{\mu_0}{4\pi}\int_{-\infty}^{\infty}\frac{\mathrm{d}z'}{\sqrt{r^2+z'^2}} + \text{constant} = \frac{\mu_0}{2\pi}\lim_{L\to\infty}\int_0^L \frac{\mathrm{d}z'}{\sqrt{r^2+z'^2}} + \text{constant}\ .$$

The integral is divergent. We sidestep this difficulty by means of the constant. Let it depend on L. This is allowed because the Coulomb gauge is still preserved. Then

$$A_z(\mathbf{x}) = \lim_{L\to\infty}\left(\frac{I\mu_0}{2\pi}\int_0^L \frac{\mathrm{d}z'}{\sqrt{r^2+z'^2}} + k(L)\right) =$$
$$= \lim_{L\to\infty}\left(\frac{I\mu_0}{2\pi}\left[\ln\left(z'+\sqrt{r^2+z'^2}\right)\right]_0^L + k(L)\right)\ .$$

With a suitable choice of $k(L)$ this becomes

$$A_z = -\frac{I\mu_0}{2\pi}\ln r \quad \Rightarrow \quad B_\phi = (\nabla\times\mathbf{A})_\phi = \frac{I\mu_0}{2\pi}\frac{1}{r}\ , \tag{238}$$

in full agreement with what we had before.[60]

So much for the magnetic field from an infinite straight wire. The next example to be considered is the magnetic field from a circular current loop. This is conceptually straightforward since the current distribution is now localized, but it is technically somewhat tricky because there is no obvious way to choose a coordinate system for this problem. Choosing spherical polars the solution comes out in terms of elliptic integrals. Jackson provides the details in section 5.5. They confirm that Figure 3 is qualitatively correct.

Asymptotics of the static magnetic field

[60] Exercise: You may or may not feel that this is a convincing derivation. If not, calculate **B** from **A** before performing the integral. Then perform the modified integral that you obtain in this way.

We now ask what the magnetostatic field looks like far away from its sources. One part of the answer is obvious. In vacuum the magnetostatic field obeys the same equations as does the electrostatic field, so the terms in the multipole expansion will be the same as well. However, the coefficients in front of the various terms carry information about the sources of the field, and here there will be significant differences. Since there are no magnetic monopoles the expansion will start with a dipole term:

$$\mathbf{B} = \frac{\mu_0}{4\pi} \frac{3\mathbf{x}\ \mathbf{m}\cdot\mathbf{x} - r^2\mathbf{m}}{r^5} + \text{quadrupole} + \dots \ . \tag{239}$$

Our question is how the dipole vector $\mathbf{m}$ can be calculated from the currents that are sourcing the field. We assume that we can find a closed surface surrounding the system such that all currents vanish outside it.

It is convenient to do the calculation for the vector potential in the Coulomb gauge, so we go back to the general solution and do the expansion as

$$\mathbf{A} = \frac{\mu_0}{4\pi} \int \frac{\mathbf{J}(\mathbf{x}')}{|\mathbf{x}-\mathbf{x}'|} \mathrm{d}^3x' = \frac{\mu_0}{4\pi} \int \mathbf{J}(\mathbf{x}') \left(\frac{1}{|\mathbf{x}|} + \frac{\mathbf{x}\cdot\mathbf{x}'}{|\mathbf{x}|^3} + \dots \right) \mathrm{d}^3x' \ . \tag{240}$$

It turns out that the boldface vector notation leads to some obscurities, so we will gradually switch to the index notation that we briefly described in our Introduction. (Jackson does that too.)

We begin with an identity:

$$\nabla\cdot(x_i\mathbf{J}) = (\nabla x_i)\cdot\mathbf{J} + x_i\nabla\cdot\mathbf{J} = J_i \ . \tag{241}$$

In the second step we used $\nabla\cdot\mathbf{J} = 0$, as well as $\partial_j x_i = \delta_{ij}$. It follows that

$$\int \frac{J_i(\mathbf{x}')}{|\mathbf{x}|}\ \mathrm{d}^3x' = \frac{1}{|\mathbf{x}|}\int J_i(\mathbf{x}')\mathrm{d}^3x' = 0 \ , \tag{242}$$

since the current is localised and we are integrating a total divergence. This confirms that there is no monopole term present in the field.

For the dipole term we employ the identity

$$\nabla\cdot(x_ix_j\mathbf{J}) = x_jJ_i + x_iJ_j \ . \tag{243}$$

This means that when we are inside an integral over all space we can write $x_j J_i = -x_i J_j$ for any localised distribution of currents.

For a pure dipole field the ith component of the vector potential is

$$A_i(\mathbf{x}) = \frac{\mu_0}{4\pi} \frac{\mathbf{x}}{|\mathbf{x}|^3} \cdot \int \mathbf{x}' J_i' \, \mathrm{d}^3 x' \ . \tag{244}$$

Not only are we edging towards index notation, we also introduced the notational device that $J_i' \equiv J_i(\mathbf{x}')$. To proceed we go all the way. We use Einstein's summation convention to deal with scalar products and the ϵ–δ identity (18) to deal with repeated vector products. We calculate that

$$\begin{aligned} \mathbf{x} \cdot \int \mathbf{x}' J_i' &= x_j \int x_j' J_i' = /\text{using the identity (243) } / = \\ = -\frac{1}{2} x_j \int (x_i' J_j' - x_j' J_i') &= /\text{using the } \epsilon\text{–}\delta \text{ identity backwards}/ = \\ = -\frac{1}{2} x_j \int \epsilon_{ijk} \epsilon_{kmn} x_m' J_n' &= -\frac{1}{2} \epsilon_{ijk} x_j \int \epsilon_{kmn} x_m' J_n' \ . \end{aligned} \tag{245}$$

We need a name here. We define the *magnetic dipole moment* $\mathbf{m}$ by

$$m_k = \frac{1}{2} \int \epsilon_{kmn} x_m J_n(x) \, \mathrm{d}^3 x \quad \Leftrightarrow \quad \mathbf{m} = \frac{1}{2} \int \mathbf{x} \times \mathbf{J}(\mathbf{x}) \, \mathrm{d}^3 x \ . \tag{246}$$

This understood we have proved that, in the Coulomb gauge, a magnetic dipole field is given by

$$\mathbf{A}(\mathbf{x}) = \frac{\mu_0}{4\pi} \frac{\mathbf{m} \times \mathbf{x}}{|\mathbf{x}|^3} \ . \tag{247}$$

The currents influence the dipole field to exactly the extent that they influence the magnetic dipole moment $\mathbf{m}$.

With the solution for the vector potential in hand it remains to show that the magnetic dipole field is that given by eq. (239), with the precise definition of $\mathbf{m}$ that we came up with here. I leave that to you.[61] And I leave the magnetic quadrupole field to its fate.

[61] Exercise: Do it using index notation!

Three magnetic dipole moments

We have seen that, far away from their sources, the electric and magnetic fields look the same apart from the fact that the magnetic monopole field vanishes. The sources influence the dipole fields only to the extent that they influence the dipole moments, which are given respectively by

$$\mathbf{p} = \int \mathbf{x}\rho(\mathbf{x})\ \mathrm{d}^3x \quad \text{and} \quad \mathbf{m} = \frac{1}{2}\int \mathbf{x}\times\mathbf{J}(\mathbf{x})\ \mathrm{d}^3x\ . \tag{248}$$

There are three memorable cases of magnetic dipoles to discuss.

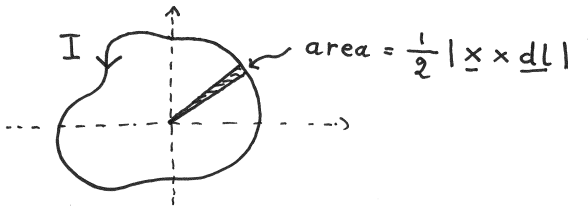

Figure 18: A current forming a loop in a plane.

First, consider a current in a wire confined to a two dimensional plane, with the geometry sketched in Figure 18. The magnetic moment will point along the z-axis (say). Regardless of the shape of the curve we find that

$$m_z = \frac{I}{2}\oint_{\text{wire}} (\mathbf{x}\times\mathbf{dl})_z = I \times (\text{area spanned by the loop})\ . \tag{249}$$

A nice result.

The next case is a swarm of point charges. We assume that each individual particle moves on some trajectory $\mathbf{x}_n = \mathbf{x}_n(t)$, where the index labels the particles rather than the components of the vectors. As a result there is an electric current distribution

$$\mathbf{J}(\mathbf{x}) = \sum_{\text{particles}} q_n\mathbf{v}_n\delta(\mathbf{x}-\mathbf{x}_n)\ . \tag{250}$$

Because of the delta functions the integral that we need to perform is trivial, and we find

$$\mathbf{m} = \frac{1}{2} \sum_{\text{particles}} q_n \mathbf{x}_n \times \mathbf{v}_n = \sum_{\text{particles}} \frac{q_n}{2m_n} \mathbf{L}_n \ , \tag{251}$$

where $\mathbf{L}_n$ is the angular momentum of the nth particle. If the particles have the same mass $m_n = m$ and charge $q_n = q$ we get the simple formula

$$\mathbf{m} = \frac{q}{2m} \mathbf{L} \ , \tag{252}$$

where $\mathbf{L}$ is the total angular momentum of the swarm.

We have to be careful if we stick this result into equation (233). The particle positions in equation (250) are time dependent, and therefore so is the current. Hence this is not really a problem in magnetostatics. However, since we are considering a "swarm" of particles the result can nevertheless be correct on the average. Jackson brings up this difficulty in his section 5.2, but defers the solution to the exercises in his chapter 14.

The third case is the observed magnetic moment of an electron with charge e and spin $\mathbf{S}$. It is

$$\mathbf{m} \approx 2 \frac{e}{2m} \mathbf{S} \approx 2.00231930436 \frac{e}{2m} \mathbf{S} \ . \tag{253}$$

The first approximation (to the truth) can be derived (as an exact result) from the Dirac equation, and the second from QED. The factor of 2 is somewhat curious, and is known as the *gyromagnetic ratio* of the electron. It may amuse you to know that, according to relativity theory, an electrically charged spinning black hole also has a gyromagnetic ratio of 2 even though it resembles an electron in no other respect.[62]

The force acting on a magnet

So far we have not even mentioned *magnets*. It is a bit like giving 'Hamlet' with the main character excluded. As a preliminary step to remedy this omission, let us compute the force acting on a magnet when placed in an external magnetic field. In the calculation we will regard the magnet as a localised current distribution leaving a dipole moment $\mathbf{m}$ as a fingerprint

[62] Reference: D. Garfinkle and J. Traschen, *Gyromagnetic ratio of a black hole*, Phys. Rev. **D42** (1990) 419.

when viewed at a distance. We will see that under reasonable assumptions it is enough to know that fingerprint.

We assume that the field is slowly varying, so that we can use the first order Taylor expansion

$$B_i(\mathbf{x}) = B_i(0) + \mathbf{x} \cdot (\nabla B_i)_{|0} \ , \tag{254}$$

where the derivative is evaluated at $\mathbf{x} = 0$. The force acting on a current distribution is

$$\mathbf{F}(0) = \int \mathbf{J}(\mathbf{x}) \times \mathbf{B}(\mathbf{x}) \ \mathrm{d}^3x \ . \tag{255}$$

Plugging our expression for $\mathbf{B}(\mathbf{x})$ into this we see that the term $B_i(0)$ does not contribute, because we have already shown that the integral over any component of the conserved current $\mathbf{J}$ vanishes. So we obtain

$$F_i = \int \epsilon_{ijk} J_j(x) x_m (\partial_m B_k)_{|0} \ \mathrm{d}^3x = \epsilon_{ijk} (\partial_m B_k)_{|0} \int x_m J_j(x) \ \mathrm{d}^3x \ . \tag{256}$$

But in the course of the calculation (245) we proved that

$$V_n \int x_n J_j \mathrm{d}^3x = -\epsilon_{jrs} V_r m_s \tag{257}$$

for an arbitrary vector V_n. (We gave the argument for the vector x_n, but it applies to arbitrary vectors.) Now we have two epsilon symbols in the formula for the force, and it is only a question of applying the ϵ–δ identity to arrive at[63]

$$F_i = \partial_i (B_k m_k) \quad \Leftrightarrow \quad \mathbf{F} = \nabla(\mathbf{m} \cdot \mathbf{B}) \ . \tag{258}$$

In words, in a constant magnetic field there is no force acting on the magnet. In a slowly varying magnetic field only the dipole moment of the magnet matters.

We see that the motion of a permanent magnet in a magnetic field varying slowly on the length scale defined by the magnet is determined by the potential energy

[63] Exercise: Do it!

$$U = -\mathbf{m} \cdot \mathbf{B} \ . \tag{259}$$

Experimentalists have devised very many ways to create magnetic fields with all sorts of properties, but you can now easily prove that it is impossible to create a static magnetic field that will make a permanent magnet hang freely in the air.[64] Of course, you should not believe everything you prove.[65]

In agreement with intuition we find that there is no net force acting on magnetic dipole in a homogeneous magnetic field. On the other hand it is well known that magnetic dipoles such as compass needles are subject to a torque when placed in a static magnetic field. The details are in Jackson's section 5.7.

Looking for magnetic monopoles

At the end of section 5.7 Jackson briefly alludes to an interesting argument concerning the magnetic moment μ_p possessed by the nucleus of the hydrogen atom. It was suggested by no less a luminary than Schwinger that they are caused by magnetic monopole charges located within the proton. To refute this suggestion Jackson compares the evaluation of the integral (210) with the corresponding integral in magnetostatics.

To understand the argument, notice that the electron has an intrinsic magnetic moment μ_e, and it interacts with the magnetic field of the proton through the hyperfine structure term

$$H_{\text{hfs}} = -\mu_e \cdot \mathbf{B}(\mathbf{x}) \tag{260}$$

in the Hamiltonian. The contribution averages to zero for a dipole field, so we consider the field $\mathbf{B}$ only in a region near the nucleus, where the dipole approximation *fails*. This region is so small that the s-wave wave function ψ is about constant there, so the contribution to the energy is

[64] Exercise: Consider permanent magnets ($\mathbf{m}$ constant), paramagnets and diamagnets ($\mathbf{m} = k\mathbf{B}$ with $k > 0$ for paramagnets, $k < 0$ for diamagnets). For each case investigate whether you can choose $\mathbf{B}(\mathbf{x})$ such that a stable equilibrium position for the dipole exists.

[65] Reference: The simulation shown on 3dhouse.se/ingemar/exjobb/exjobb.html agrees accurately with what Axel Erbing saw in the lab. See his Master's thesis, available on the same page.

$$\Delta E = \int_{r<R} \psi^*(\mathbf{x}) H_{\rm hfs} \psi(\mathbf{x}) \, \mathrm{d}^3x \approx -|\psi(0)|^2 \, \mu_e \cdot \int_{r<R} \mathbf{B}(\mathbf{x}) \, \mathrm{d}^3x \; . \qquad (261)$$

But if this magnetic field arises from magnetic monopole charges confined inside the nucleus then our result (215) applies to the integral that remains. We then obtain

$$\Delta E = \frac{\mu_0}{3} |\psi(0)|^2 \mu_e \cdot \mu_p \; . \qquad (262)$$

Putting in the numbers one finds that the resulting energy splitting of the hydrogen ground state corresponds to a wavelength of 42 cm. But no such line is present in the spectrum, so Schwinger must be wrong.

To get a correct result we have to do the integral under the assumption that the magnetic field is generated by electric currents confined inside the nucleus. As a first step

$$\int_{r<R} \mathbf{B}(\mathbf{x}) \, \mathrm{d}^3x = \int_{r<R} \nabla \times \mathbf{A} \, \mathrm{d}^3x = R^2 \int \mathbf{n} \times \mathbf{A} \, \mathrm{d}\Omega \; . \qquad (263)$$

This we can related to the currents using our solution for $\mathbf{A}$, eq. (233). The calculation is an exercise.[66] The result is

$$\int_{\rm ball} \mathbf{B} \, \mathrm{d}V = \frac{2\mu_0}{3} \mathbf{m} \; . \qquad (264)$$

This is what we have to insert into eq. (261) if we assume that the magnetic field is created by electric currents rather than by magnetic charges. The beauty of the argument is that no detailed information about these currents is needed. Comparing to eq. (262) we see that the energy splitting changes sign. More importantly there is an extra factor of 2, so the calculation predicts—correctly—that there is a 21 cm line in the hydrogen spectrum.

Magnetism in media

When we consider the magnetic field in a medium there is a twist right at the beginning. The magnetostatic equations must be modified in order to

[66] Exercise: Do it! (It is OK to look in Jackson.)

account for a magnetic dipole density $\mathbf{M}$ inside the medium. When doing this one arrives at

$$\begin{cases} \nabla \times \mathbf{H} = \mathbf{J} \\ \\ \nabla \cdot \mathbf{B} = 0 \end{cases} \tag{265}$$

where the macroscopic field $\mathbf{H}$ is related to the magnetisation $\mathbf{M}$ by

$$\mathbf{H} = \frac{1}{\mu_0}\mathbf{B} - \mathbf{M} \quad \Leftrightarrow \quad \mathbf{B} = \mu_0\mathbf{H} + \mu_0\mathbf{M} \; . \tag{266}$$

We have to convert the last equation into a constitutive relation $\mathbf{B} = \mathbf{B}(\mathbf{H})$ before we can use this system.

Now this looks completely backwards compared to the electrostatic equations

$$\nabla \cdot \mathbf{D} = \rho \; , \quad \nabla \times \mathbf{E} = 0 \; , \quad \mathbf{D} = \epsilon_0\mathbf{E} + \mathbf{P} \; , \quad \mathbf{D} = \mathbf{D}(\mathbf{E}) \; . \tag{267}$$

So does the terminology that you may come across:

$\mathbf{E}$:	electric field	$\mathbf{H}$:	magnetic field
$\mathbf{D}$:	displacement field	$\mathbf{B}$:	magnetic induction .

To see why things are presented this way, ask how you would create the fields to some specification. You would probably create the electric field by choosing the electric potential. Hence the emphasis on $\mathbf{E} = -\nabla\Phi$. But you would create the magnetic field by choosing some external currents, hence the emphasis on $\mathbf{H}$. That is to say, $\mathbf{E}$ and $\mathbf{H}$ are the fields you can easily control. You will be able to detect a changing $\mathbf{B}$ by the electromotive force it induces in a coil through Faraday's law. This explains the terminology (that I prefer to avoid).

Concerning the constitutive relation that we seek, there are many cases where the simple assumption

$$\mathbf{B} = \mu\mathbf{H} \tag{268}$$

works, with a magnetic permeability μ that depends on position but not on $\mathbf{H}$. However, this constant can now take either sign. In vacuum, $\mu = \mu_0$. In *diamagnets* $\mu < \mu_0$. Then the molecules do not have permanent magnetic moments, but when $\mathbf{H}$ is switched on various atomic currents react by opposing the field. In *paramagnets* $\mu > \mu_0$. Then there are permanent molecular magnetic moments, typically coming from the electrons in incompletely filled atomic shells. When $\mathbf{H}$ is switched on these tend to line up with with $\mathbf{H}$, and support it.[67]

Ferromagnets and hysteresis phenomena

Then we come to the difficult case of ferromagnets, including iron, nickel, and cobalt. Their molecules have permanent magnetic moments that line up with each other also in zero field due to involved quantum mechanical effects. However, the domains in which this happens may be microscopically large but macroscopically small. In a piece of naturally occurring iron the magnetic moments from these domains tend to average out, and there is no net magnetisation of the piece of iron. It is however possible to line up the domains with each other by means of a magnetic field, and they will then tend to stay aligned also when the field is removed.

Parenthetically we remark that there exist materials where molecular electric dipole moments do line up in zero field in an analogous way. They are called ferroelectrics by analogy, but do not include iron. They are not so easy to find because electric dipoles in air are quickly neutralized by free charges. The first ferroelectric material (the Rochelle salt) was discovered in 1920.

Coming back to the magnetisation of iron, we imagine an experiment in which a piece of iron is subjected to a changing external field $\mathbf{H}$, controlled by a current going through a coil. There is another coil coupled to a galvanometer, allowing us to detect any changes in the total field $\mathbf{B}$ through the induced emf. See Figure 19.

What one typically finds is something like the curves shown in Figure 20.

[67] Exercise: Solve the Lorentz equation for a charged particle in a homogeneous magnetic field. Based on this solution, would you expect a plasma to be diamagnetic or paramagnetic?

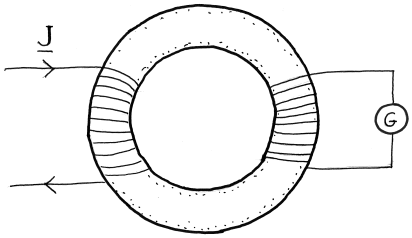

Figure 19: **H** is changed by sending a controlled current **J** through a coil, and the resulting changes in **B** are measured using a second coil.

We start with no net magnetic field. When **H** is first applied the observed magnetic flux increases slowly until we reach point *a* on the curve. It increases sharply after that, and it is easy to guess that this happens because entire domains of aligned dipoles switch directions inside the material. Eventually all the domains have flipped, and saturation is reached, at *b*. If we try to reverse the process by decreasing the strength of the applied **H** then the domains stay aligned all the way down to zero field, and start to reverse only when the direction of **H** has been switched, at *c*. So the process is in fact not reversed, instead the material has been trapped in a hysteresis loop *b*–*c*–*d*–*e*–*b*. It does not return to the origin of the diagram.

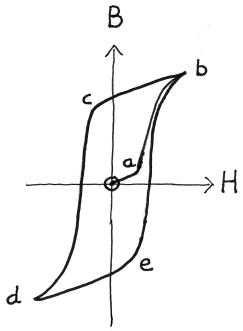

Figure 20: A hysteresis loop is created.

The domain flips are known as *Barkhausen jumps*, after the person who first heard them. Namely, in the experiment he connected the second coil to an amplifier and a pair of earphones, and literally heard the flips happen.

Reflecting on this phenomenon you can see why we have referred to $\mathbf{B} = \mathbf{B}(\mathbf{H})$ as a constitutive *relation* rather than as a function. It may very well not be a single valued function, and it can depend on the previous history of the material. To emphasise that this is a possibility to be reckoned with the constitutive relations are often written using square brackets as $\mathbf{B} = \mathbf{B}[\mathbf{H}]$ and $\mathbf{D} = \mathbf{D}[\mathbf{E}]$. We are, however, now infringing on solid state physics, and stop here.

For the exercise class:

- Jackson, Problem 5.1.
- Consider two curves in space, $C : x_i(s)$ and $C' : x'_i(s')$. They are circles topologically. Define the *linking number* m as follows: Deform one of the curves to a circle, and count the number of times the second curve passes through the disk spanned by that circle, counting $+1$ if it passes in the direction of the normal of the disk and -1 if it passes in the other direction. Use your knowledge of magnetostatics to prove that

$$m = \frac{1}{4\pi}\epsilon_{ijk}\int_C\int_{C'}\frac{(x_i - x'_i)\mathrm{d}l_j\mathrm{d}l_k}{|x - x'|^3} \ . \tag{269}$$

- Define a vector potential on a region of space strictly outside the z-axis, such that $\mathbf{A}(\mathbf{x})$ is independent of z, gives a vanishing magnetic field outside the z-axis, and cannot be gauge transformed to zero. Discuss the last point in some detail, and give a physical interpretation.[68]

[68]Reference: When you are done, consult T. T. Wu and C. N. Yang, *Concept of non-integrable phase factors and global formulation of gauge fields*, Phys. Rev. **D12** (1975) 3845.

SPECIAL RELATIVITY

From now on I will use Gaussian units, just as Jackson does in his chapter 11. Using what we will soon call Cartesian tensors Maxwell's equations then read

$$\partial_i B_i = 0 \qquad \text{Gilbert's law} \tag{270}$$

$$\epsilon_{ijk}\partial_j E_k + \frac{1}{c}\partial_t B_i = 0 \qquad \text{Faraday's law} \tag{271}$$

$$\partial_i E_i = 4\pi\rho \qquad \text{Gauss' law} \tag{272}$$

$$\epsilon_{ijk}\partial_j B_k - \frac{1}{c}\partial_t E_i = \frac{4\pi}{c} j_i \qquad \text{Ampère-Maxwell's law .} \tag{273}$$

If you look in a quantum field theory book you will likely find that it uses a 'rationalized' version of Gaussian units, in which the factor of 4π is treated differently. The constant c takes the value 300 000 km/s.

Some questions at issue

As you probably know special relativity was discovered through a careful analysis of the transformations leaving Maxwell's equations invariant.[69] The question answered, or rather side-stepped, by Einstein was: 300 000 km/s relative to what? But there was another issue that other people had already spent hard work on. To see what it was, consider a charged particle moving with constant velocity $\mathbf{v}$ through a magnetic field. Because it is moving it feels a Lorentz force. But we have learned from Newton that only relative velocities can be observed. So let us consider a charged particle at rest, and let the magnet move with velocity $-\mathbf{v}$. The relative velocity is unchanged, so again the particle should feel a force. But since it is now at rest this

[69]Reference: A. Einstein: *Zur Elektrodynamik bewegter Körper*, Annalen der Physik **17** (1905) 891.

means that a moving magnet must give rise to an electric field, not just a magnetic field. Newton's ideas further suggest that if we know the magnetic field from a magnet at rest then we should be able to find the fields from a magnet moving at constant velocity by applying a simple transformation to the static magnetic field. Thus the question answered by Lorentz and others was: What is that transformation?

At the time all of this required clear thinking and hard work. If you invest in some mathematics first, it takes no effort. Einstein is supposed to have complained that he did not understand relativity theory anymore, once the mathematicians took it in hand. But it did not take him long to realize that tensor calculus is a good thing.

What is a transformation, anyway?

Our first task is to understand what is being meant when we say that a set of equations is invariant under a transformation. Suppose we fix Cartesian coordinates in space, so that points can be identified by their coordinates. Given a point $\mathbf{x}$ we define a new point $\mathbf{x}'$ by

$$\mathbf{x}' = \mathbf{x} + \mathbf{a} \ , \tag{274}$$

where $\mathbf{a}$ is a constant vector. Equation (274) then describes a *translation* in space. We call it an *active* transformation. Actually there is another way to read the same equation. We can agree that all points in space are fixed, but that we change the coordinate system so that the point that had coordinates $\mathbf{x}$ now gets the new coordinates $\mathbf{x}'$. This is a *passive* coordinate transformation. Special relativity is often presented as being concerned with how a phenomenon looks like when viewed from different 'reference frames'. But here we are more interested in the fact that it also claims that if there exists a solution of Maxwell's equations describing (say) a static charge distribution, then there must exist a solution describing a charge distribution moving with constant velocity. Hence we take the active point of view throughout the chapter, and the coordinate system is fixed once and for all.

Now consider a function of space and time, such as $\rho(\mathbf{x}, t)$. Given a transformation like the one above we can then introduce a *new* function ρ' by means of the equation

$$\rho'(\mathbf{x}', t) = \rho(\mathbf{x}, t) \ . \tag{275}$$

In words, the new function takes the same value at the point $\mathbf{x}'$ as the old function did at the point $\mathbf{x}$. To be explicit about it,

$$\rho'(\mathbf{x} + \mathbf{a}, t) = \rho(\mathbf{x}, t) \quad \Leftrightarrow \quad \rho'(\mathbf{x}, t) = \rho(\mathbf{x} - \mathbf{a}, t) \ . \tag{276}$$

We have moved the charge density to a new position. Let us apply the same transformation to the electric field,

$$\mathbf{E}'(\mathbf{x}, t) = \mathbf{E}(\mathbf{x} - \mathbf{a}, t) \ , \tag{277}$$

and similarly for the magnetic field $\mathbf{B}$ and for the current density $\mathbf{j}$. You can now check that Maxwell's equations are *invariant* under translations in space, which means that if $(\mathbf{E}, \mathbf{B}, \rho, \mathbf{j})$ is a solution of Maxwell's equations, so is $(\mathbf{E}', \mathbf{B}', \rho', \mathbf{j}')$.

Next, consider the *Galilei transformation*

$$\mathbf{x}' = \mathbf{x} + \mathbf{v}t \ , \qquad t' = t \ , \tag{278}$$

where $\mathbf{v}$ is a constant velocity vector. Suppose that the function $\rho(\mathbf{x})$ describes a static charge distribution. Then the new function $\rho'(\mathbf{x}, t) = \rho(\mathbf{x} - \mathbf{v}t)$ is a charge distribution moving with constant velocity. Perform the same transformation of the electric and magnetic fields. It will then be found that Maxwell's equations are *not* invariant under this transformation.[70] Special relativity is a response to this quandary. It deals with transformations of a four dimensional *spacetime*, with a fourth coordinate $x^0 = ct$.

Tensors

The tool we need now is called *tensors*. For definiteness, suppose that space—or rather, spacetime—has four dimensions, so that each point can be uniquely described by four coordinates x^0, x^2, x^2, and x^3. We collect the coordinates into a vector x^α, where the index α ranges from 0 to 3. We are interested

[70] Exercise: Check that Maxwell's equations are invariant under translations, and that Faraday's law is invariant under the Galilei transformation proposed in Jackson's section 5.13.

in linear transformations, possibly including a constant shift, so we assume that

$$x^\alpha \to x'^\alpha = x'^\alpha(x) = \sum_{\beta=0}^{3} \Lambda^\alpha{}_\beta x^\beta + a^\alpha \ . \tag{279}$$

This is a matrix equation. You may wonder why I write one index 'upstairs' and one index 'downstairs'. The reason begins to emerge when we compute how derivatives transform. Using the chain rule we get

$$\partial'_\alpha \equiv \frac{\partial}{\partial x'^\alpha} = \sum_{\beta=0}^{3} \frac{\partial x^\beta}{\partial x'^\alpha} \frac{\partial}{\partial x^\beta} = \sum_{\beta=0}^{3} (\Lambda^{-1})^\beta{}_\alpha \frac{\partial}{\partial x^\beta} \ . \tag{280}$$

We see that a gradient vector transforms in a different way than does the coordinate vector.

We will use Einstein's summation convention from now on, with one important extra rule. It will be true by convention that

$$(\Lambda^{-1})^\beta{}_\alpha \partial_\beta \equiv \sum_{\beta=0}^{3} (\Lambda^{-1})^\beta{}_\alpha \frac{\partial}{\partial x^\beta} \ . \tag{281}$$

The new rule is that repeated indices (implying a sum) are only allowed if one of them is upstairs and the other downstairs. It is *forbidden* to write expressions like $(\Lambda^{-1})^\beta{}_\alpha \partial_\alpha$. The reason why will soon become clear.

Now we want a rule to tell us how various objects defined on the space transform when the active transformation is carried through. First, a *scalar* function is by definition a function that transforms according to

$$\phi'(x') = \phi(x) \ . \tag{282}$$

The new function takes the same value at the new point as the old function did at the old point. Second, a *contravariant vector field* is a set of four functions that transform according to

$$V'^\alpha(x') = \frac{\partial x'^\alpha}{\partial x^\beta} V^\beta(x) = \Lambda^\alpha{}_\beta V^\beta(x) \ . \tag{283}$$

A component of the new vector takes a value at the new point which is a linear combination of the values taken by the old vector components at the old point.

Third, a *covariant vector field* is a set of four functions that transform according to

$$U'_\alpha(x') = \frac{\partial x^\beta}{\partial x'^\alpha} U_\beta(x) = (\Lambda^{-1})^\beta{}_\alpha U_\beta(x) \ . \tag{284}$$

We can take scalar products between contravariant and covariant vectors,

$$U'_\alpha V'^\alpha = U_\beta (\Lambda^{-1})^\beta{}_\alpha \Lambda^\alpha{}_\gamma V^\gamma = U_\beta (\Lambda^{-1}\Lambda)^\beta{}_\gamma V^\gamma = U_\beta \delta^\beta_\gamma V^\gamma = U_\alpha V^\alpha \ . \tag{285}$$

The result is indeed a scalar function. Contravariant and covariant vectors are analogous to Dirac's bras and kets in quantum mechanics. It is not possible to define a scalar product between (say) two covariant vectors.[71]

In quantum mechanics we often define vectors that belong to tensor products of vector spaces, such as $|\psi\rangle \otimes |\phi\rangle$. In tensor calculus we can define tensors with an arbitrary number of indices. Sticking to two indices to begin with, we find three different kinds of such tensors. Contravariant tensors transforming according to

$$T'^{\alpha\beta}(x') = \Lambda^\alpha{}_\mu \Lambda^\beta{}_\nu T^{\mu\nu}(x) \ , \tag{286}$$

covariant tensors transforming according to

$$T'_{\alpha\beta}(x') = (\Lambda^{-1})^\mu{}_\alpha (\Lambda^{-1})^\nu{}_\beta T_{\mu\nu}(x) \ , \tag{287}$$

and mixed tensors transforming according to

$$T'^\alpha{}_\beta(x') = \Lambda^\alpha{}_\mu (\Lambda^{-1})^\nu{}_\beta T^\mu{}_\nu(x) \ . \tag{288}$$

The generalisation to an arbitrary number of indices should be clear.

Two tensors of the same kind can be added together, but it is forbidden to add tensors of different kinds because the index structure of each term in a tensor equation must be the same. There are two different ways to get a new tensor from a pair of old ones. We can use outer multiplication. As an example,

$$T_{\alpha\beta\gamma} = U_\alpha V_{\beta\gamma} = V_{\beta\gamma} U_\alpha \tag{289}$$

[71] Exercise: What would go wrong if we applied the Einstein convention to the (forbidden) expression $U^\alpha V^\alpha$? Equivalently, why did Einstein forbid us to write such things?

is a covariant tensor with three indices. We can also use *contraction*, that is to say we can sum over a pair of (suitably placed) indices. Thus

$$T_\alpha{}^\gamma = S_{\alpha\beta}U^{\beta\gamma} \tag{290}$$

is a mixed tensor with two indices. Since we already concluded that a gradient transforms like a vector we can also obtain tensors with an extra covariant index by taking a derivative. Thus

$$T_{\alpha\beta\gamma}(x) = \partial_\alpha S_{\beta\gamma}(x) \tag{291}$$

is a covariant tensor field with three indices. Here, however, a word of warning is necessary: This last construction works only because the matrix $\Lambda^\alpha_{\ \beta}$ is independent of the coordinates. More work is needed if this is not true.[72]

In general the order of two downstairs or upstairs indices matters. That is to say, $T_{\alpha\beta}$ is not in general equal to $T_{\alpha\beta}$. If it is true that

$$T_{\alpha\beta} = T_{\beta\alpha} \tag{292}$$

then the tensor is said to be *symmetric*. If

$$T_{\alpha\beta} = -T_{\beta\alpha} \tag{293}$$

it is *antisymmetric*. Any tensor can be divided into a symmetric and an anti-symmetric part,

$$T_{\alpha\beta} = T_{(\alpha,\beta)} + T_{[\alpha,\beta]}\ , \qquad T_{(\alpha\beta)} = T_{(\beta\alpha)}\ , \quad T_{[\alpha\beta]} + T_{[\beta\alpha]} = 0\ . \tag{294}$$

For tensors with two and three indices, respectively, we have

$$T_{[\alpha\beta]} = \frac{1}{2}(T_{\alpha\beta} - T_{\beta\alpha}) \tag{295}$$

$$T_{[\alpha\beta\gamma]} = \frac{1}{6}(T_{\alpha\beta\gamma} + T_{\beta\gamma\alpha} + T_{\gamma\alpha\beta} - T_{\beta\alpha\gamma} - T_{\gamma\beta\alpha} - T_{\alpha\gamma\beta})\ . \tag{296}$$

[72]Exercise: Check carefully that $V_\beta T^{\beta\alpha}$ is a contravariant vector. What can go wrong if you apply this argument to $\partial_\beta T^{\beta\alpha}$?

The number of components in a tensor with two indices is 4^2 (or d^2 if the indices go from 1 to d). If the tensor is symmetric the number of independent components drops to ten, if it is anti-symmetric it drops to six.[73]

The actual values taken by the components of a tensor at some given point can be changed by an active (or passive) transformation. But there is an exception to this rule, namely the Kronecker delta that can be defined by

$$\delta^\alpha_\beta = \begin{cases} 1 & \text{if} \quad \alpha = \beta \\ 0 & \text{if} \quad \alpha \neq \beta \, . \end{cases} \tag{297}$$

These values cannot be changed by any transformation because

$$\delta'^\alpha_\beta = \Lambda^\alpha{}_\mu (\Lambda^{-1})^\nu{}_\beta \delta^\mu_\nu = \Lambda^\alpha{}_\mu (\Lambda^{-1})^\mu{}_\beta = \delta^\alpha_\beta \, . \tag{298}$$

For this reason the Kronecker delta is said to be an *invariant tensor*.[74]

This concludes our discussion of tensors in general. We can easily include spacetime translations in the discussion, but had we allowed more general transformations than those given by a constant matrix in equation (279) we would faced a lengthy discussion about how spatial derivatives can be included in the tensor formalism. Since we restricted the allowed transformations in the way we did we can omit this. In fact, we will go the other way and restrict the allowed transformations even further. As a result we will obtain additional invariant tensors to play with, and we will need these in order to formulate Maxwell's equations in tensor language.

Cartesian tensors

We now turn to Maxwell's equations as given in equations (270)–(273). The Latin indices there run from 1 to 3, but that is a minor difference. More striking is that the equations seem to violate the rule that you cannot repeat two downstairs indices in the same term of an equation. Indeed, all the indices are downstairs. What is the magical ingredient that allows us to do this?

[73] Exercise: Count the number of independent components in a totally anti-symmetric tensor with either three or four indices (given that the dimension is four).

[74] Exercise: Prove that the Kronecker delta is an invariant tensor also if do not assume that the transformation $x^\alpha \to x'^\alpha = x'^\alpha(x)$ is linear, as we did in equation (279).

The answer is the tensor

$$g^{ij} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} . \tag{299}$$

We require that this tensor be invariant. This means that we *restrict* the allowed transformations so that

$$g'^{ij} = \Lambda^i{}_k \Lambda^j{}_l g^{kl} . \tag{300}$$

In matrix notation this becomes

$$\mathbf{1} = \Lambda \mathbf{1} \Lambda^{\mathrm{T}} \quad \Leftrightarrow \quad \Lambda^{\mathrm{T}} = \Lambda^{-1} . \tag{301}$$

So we have restricted ourselves to transformations given by *orthogonal* matrices Λ. As you know (?) these are all rotations and all reflections, and nothing else.

There is an additional invariant tensor g_{ij} that is the inverse of g^{ij}, in the sense that

$$g_{ik} g^{kj} = \delta_i^j . \tag{302}$$

It is known as the *metric tensor*. The idea is that the length squared of a contravariant vector can now be defined as

$$||v||^2 = g_{ij} v^i v^j . \tag{303}$$

This is a scalar function, just as the scalar product between a covariant and a contravariant vector is a scalar function.

The metric tensor blurs the distinction between covariant and contravariant vectors. Given a contravariant vector v^i we can define a contravariant vector v_i, and conversely, through

$$v_i = g_{ij} v^j \quad \Leftrightarrow \quad v^i = g^{ij} v_i . \tag{304}$$

Because of our choice of the metric tensor the components of v^i and v_i are identical. At this point then we can decide to disregard the distinction between upstairs and downstairs, place all indices downstairs, and agree that Einstein's summation convention now applies to expressions like $v_i v_i = v_i g^{ij} v_j$. In fact it would be snobbish not to do this.

This is the justification for writing Maxwell's equations in the form (270)–(273). Note carefully however that this works only because we work in a Cartesian coordinate system and because we restrict ourselves to transformations that are either rotations or reflections (or translations).

Our metric tensor $g_{ij} = \delta_{ij}$ is an invariant tensor under rotations and reflections. But it is worth noticing that the ϵ-tensor works a little differently. We find (and we now brutally place all indices downstairs)

$$\epsilon'_{ijk} = \Lambda_{im}\Lambda_{jn}\Lambda_{kr}\epsilon_{mnr} = \det\Lambda\ \epsilon_{ijk}\ . \tag{305}$$

If you think about it this is essentially the definition of the determinant of a matrix.[75] But for an orthogonal matrix Λ we have $\det\Lambda = \pm 1$, where the plus sign applies to rotations and the minus sign to reflections. Hence the ϵ-tensor is an invariant tensor under rotations, but not under reflections. To ensure that Maxwell's equations are invariant under reflections we must have

$$E'_i(x') = \Lambda_{ij}E_j(x) \qquad B'_i(x') = \det\Lambda\ \Lambda_{ij}B_j(x)\ . \tag{306}$$

This means that the electric field changes sign under the transformation $\mathbf{x} \to -\mathbf{x}$, but the magnetic field does not. This is often expressed by saying that the magnetic field is a *pseudo-vector*, not a vector.

If you think pseudo-vectors are awkward you can work with proper vectors and tensors throughout by defining

$$F_{ij} = \epsilon_{ijk}B_k \quad \Leftrightarrow \quad B_i = \frac{1}{2}\epsilon_{ijk}F_{jk}\ . \tag{307}$$

The anti-symmetric tensor F_{ij} has three components, just like the magnetic field, but it is a genuine tensor transforming as

$$F'_{ij}(x') = \Lambda_{im}\Lambda_{jn}F_{mn}(x)\ , \tag{308}$$

regardless of whether Λ gives a rotation or a reflection. We can then rewrite Maxwell's equations as[76]

$$\begin{cases} \partial_i B_i = 0 & \partial_i E_j - \partial_j E_i + \frac{1}{c}\partial_t F_{ij} = 0 \\ \partial_i E_i = 4\pi\rho & \partial_j F_{ij} - \frac{1}{c}\partial_t E_i = \frac{4\pi}{c} j_i\ . \end{cases} \tag{309}$$

[75] Exercise: If you do not see it immediately, prove it!

[76] Exercise: Again, if it is not obvious to you, prove it!

This way of writing them is actually the first step towards writing the equations in a form that makes their behaviour under Lorentz transformations obvious.

Invariance of Maxwell's equations

Now we are coming to the point of this apparent digression from the subject. We want to write Maxwell's equations in a way that makes their behaviour under Lorentz transformation transparent, and we will use tensors on a four dimensional spacetime for this purpose. We have already collected the coordinates into a 4-vector,

$$x^\alpha = \begin{pmatrix} x^0 \\ \mathbf{x} \end{pmatrix} = \begin{pmatrix} ct \\ \mathbf{x} \end{pmatrix} . \tag{310}$$

It seems reasonable to do the same for the charge density and the current,

$$J^\alpha = \begin{pmatrix} \rho \\ \mathbf{j}/c \end{pmatrix} . \tag{311}$$

But what do we do about the six components in the electric and the magnetic field? They will be joined together to form a tensor with six components, and an anti-symmetric tensor $F_{\alpha\beta} = -F_{\beta\alpha}$ fits the bill. So we define the *Maxwell tensor*

$$F_{\alpha\beta} = \begin{pmatrix} 0 & F_{0i} \\ F_{i0} & F_{ij} \end{pmatrix} = \begin{pmatrix} 0 & -E_i \\ E_i & \epsilon_{ijk}B_k \end{pmatrix} . \tag{312}$$

You can now check that the first pair of Maxwell's equations in (309), that is Gilbert's and Faraday's laws, take the elegant form[77]

$$\partial_\alpha F_{\beta\gamma} + \partial_\gamma F_{\alpha\beta} + \partial_\beta F_{\gamma\alpha} = 0 . \tag{313}$$

The remaining pair is not so easily handled though.

[77] Exercise: Prove that the left hand side is a totally anti-symmetric tensor with three indices, hence that it has four components. Then prove that these components are precisely the two Maxwell equations that I claim they are.

We have to follow the rules, which means that if we want to contract a tensor with a gradient then that tensor must have an index upstairs. The only choice in sight is to set

$$\partial_\beta F^{\alpha\beta} = 4\pi J^\alpha \ . \tag{314}$$

But we have not defined $F^{\alpha\beta}$. We must find a way to raise the indices on $F_{\alpha\beta}$. We introduce a metric tensor $g_{\alpha\beta}$ to do this, and define

$$F^{\alpha\beta} = g^{\alpha\gamma} g^{\beta\delta} F_{\gamma\delta} \ . \tag{315}$$

The obvious option, to set $g_{\alpha\beta}$ equal to a Kronecker delta, does not work. To recover the remaining Maxwell equations from equations (312), (315), and (314) requires the definition[78]

$$g_{\alpha\beta} = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} = g^{\alpha\beta} \ . \tag{316}$$

With this metric tensor we have succeeded in rewriting Maxwell's equations in tensor form.

To repeat, Maxwell's equations are

$$\begin{cases} \partial_\alpha F_{\beta\gamma} + \partial_\gamma F_{\alpha\beta} + \partial_\beta F_{\gamma\alpha} = 0 \\ \\ \partial_\beta F^{\alpha\beta} = 4\pi J^\alpha \end{cases} \ . \tag{317}$$

Some things are now obvious. For instance, here is the proof that the equations imply charge conservation:

$$4\pi \partial_\alpha J^\alpha = \partial_\alpha \partial_\beta F^{\alpha\beta} = 0 \ . \tag{318}$$

If it escapes you why this is obvious, here is the complete argument:

$$\partial_\alpha \partial_\beta F^{\alpha\beta} = \partial_\beta \partial_\alpha F^{\beta\alpha} = -\partial_\alpha \partial_\beta F^{\alpha\beta} \quad \Rightarrow \quad \partial_\alpha \partial_\beta F^{\alpha\beta} = 0 \ . \tag{319}$$

[78]Exercise: Check this. This is one of the most interesting signs in physics. A less interesting sign is an overall change of sign of $g_{\alpha\beta}$. Convince yourself that this is allowed.

First the names of the summation indices were changed, then we used the fact that $F^{\alpha\beta}$ is anti-symmetric while $\partial_\beta\partial_\alpha = \partial_\alpha\partial_\beta$. This is a standard trick in many tensor calculations. It is similarly easy to show that the components of $F_{\alpha\beta}$ obey the wave equation whenever $J^\alpha = 0$.[79]

Lorentz transformations

Now we are in position to see what kind of transformations that leave Maxwell's equations invariant. Since we brought the metric tensor (316) into the equations the transformations that take solutions of Maxwell's equations to other solutions of Maxwell's equations must have the property that

$$\Lambda^\alpha{}_\gamma \Lambda^\beta{}_\delta g^{\gamma\delta} = g^{\alpha\beta} \ . \tag{320}$$

A little reflection will tell you that this includes all rotations in space. They are given by matrices of the form

$$\Lambda = \left(\begin{array}{c|c} 1 & 0 \\ \hline 0 & R \end{array}\right) \ , \tag{321}$$

where R is an orthogonal matrix. The more interesting ones mix space and time together. You can check that one possibility is

$$\Lambda^\alpha{}_\beta = \begin{pmatrix} \cosh\alpha & \sinh\alpha & 0 & 0 \\ \sinh\alpha & \cosh\alpha & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \ , \tag{322}$$

where α is a real number. This is called a *Lorentz boost*. It transforms the spacetime point x^α to a new point x'^α according to

$$x'^\alpha = \Lambda^\alpha{}_\beta x^\beta \ , \tag{323}$$

or more explicitly (remembering that $x^0 = ct$)

$$ct' = \cosh\alpha \ ct + \sinh\alpha x \qquad x' = \sinh\alpha \ ct + \cosh\alpha \ x \tag{324}$$

[79] Exercise: Do it, in at most two lines.

together with $y' = y$, $z' = z$.[80]

The four dimensional space whose coordinates are x^α and whose metric tensor is $g_{\alpha\beta}$ is known as *Minkowski space*. I will assume that you know the rest of the story, and stop here.[81] The point is that the story is forced on us if we insist on writing Maxwell's equations in tensor form.

The stress-energy tensor of the electromagnetic field

We now take up the question how our proposed formula for the energy in an electromagnetic field is to be transcribed into tensor notation. In Gaussian units

$$E = \frac{1}{8\pi} \int (\mathbf{E} \cdot \mathbf{E} + \mathbf{B} \cdot \mathbf{B}) \, \mathrm{d}^3 V \ . \tag{325}$$

In special relativity energy is to be regarded as the fourth component of a four-vector

$$P^\alpha = \begin{pmatrix} E \\ p_i \end{pmatrix} \ . \tag{326}$$

The space over which we integrate in eq. (325) is a three dimensional *hypersurface* in a four dimensional spacetime. As such it has a (future directed) normal vector n_α, see Figure 21. We choose our space to be defined by $t = 0$, and then the normal vector has a single non-vanishing component. When we integrate over the volume we are really computing the flux of a four-vector field through a hypersurface, so the volume element must 'point' in some direction. Putting these ingredients together we expect to obtain our four-vector as

$$P^\alpha = \int T^{\alpha\beta} n_\beta \, \mathrm{d}V = \int T^{\alpha 0} \, \mathrm{d}V \ . \tag{327}$$

Here $T^{\alpha\beta}$ is a tensor that we must choose so that

[80] Exercise: Draw a diagram of the (ct, x)–plane showing how points are transformed as you vary α. Compare with a similar drawing of how the (x, y)–plane behaves under rotations.

[81] Exercise: If you need a reminder, replace the parameter α with the parameter $v = c \tanh \alpha$ in equation (324). What is the interpretation of v?

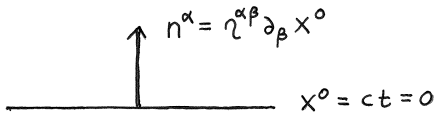

Figure 21: Space at time $t = 0$ as a three dimensional hypersurface, and its normal vector.

$$T^{0\beta}n_\beta = T^{00} = \frac{1}{8\pi}(\mathbf{E}\cdot\mathbf{E} + \mathbf{B}\cdot\mathbf{B}) = \frac{1}{8\pi}\left(F_{0i}F_{0i} + F_{ij}F_{ij}\right) \ . \tag{328}$$

Moreover local energy conservation must be built in. That is to say, in vacuum we must have

$$\partial_\gamma T^{\gamma 0} = \partial_t T^{00} + \partial_i T^{i0} = 0 \ . \tag{329}$$

Hence the components $T^{i\beta}n_\beta = T^{i0}$ must yield our expression for the energy density flow of the field, otherwise known as the Poynting vector. At the end the components T^{0i} and T^{ij} must be given some interpretation too.

The tensor $T^{\alpha\beta}$ is known as the *energy-momentum tensor*, or sometimes as the *stress-energy tensor* for reasons having to do with the components T^{ij}. To understand the latter we need a small excursion into fluid mechanics. Consider an arbitrary volume V immersed in the fluid. It is subject to two kinds of forces, volume forces F_i that act on every particle in the volume, and surface forces that are the stresses exerted on the volume by the surrounding fluid. By a clever argument of Cauchy's—he considered the forces acting on a volume element small enough so that volume forces can be ignored—it can be shown that this force is linear in the normal vector n_i of surface.[82] That is to say, the surface forces are given locally by

$$f_i = T_{ij}n_j \ , \tag{330}$$

where T_{ij} is known as the *stress tensor*. In equilibrium the forces have to balance, so we obtain the two equations

$$\int_V F_i \ \mathrm{d}V + \int_{\partial V} T_{ij}n_j \ \mathrm{d}S = 0 \tag{331}$$

[82]Reference: See *The Feynman Lectures Vol II*, Chapter 31.

$$\int_V \epsilon_{ijk} x_j F_k \, \mathrm{d}V + \int_{\partial V} \epsilon_{ijk} x_j T_{kr} n_r \, \mathrm{d}S = 0 \; . \tag{332}$$

The first tells us that (in equilibrium) the volume element does not accelerate, the second that it does not start to rotate. Begin with the first equation and apply Gauss' theorem:

$$\int_V (F_i + \partial_j T_{ij}) \, \mathrm{d}V = 0 \; . \tag{333}$$

But the volume V can be chosen arbitrarily, so we conclude that

$$F_i + \partial_j T_{ij} = 0 \; . \tag{334}$$

This is known as the *equation of momentum balance*. If the electromagnetic field carries momentum, we expect a similar equation to turn up in Maxwell's theory. Now let us consider the second equilibrium equation. Treating it in the same way we obtain

$$\int_V (\epsilon_{ijk} x_j F_k + \partial_r (\epsilon_{ijk} x_j T_{kr}) \, \mathrm{d}V = 0 \;\; \Rightarrow \;\; \epsilon_{ijk} (x_j F_k + x_j \partial_r T_{kr} + T_{kj}) = 0 \; . \tag{335}$$

Using the equation of momentum balance this reduces to

$$\epsilon_{ijk} T_{jk} = 0 \quad \Leftrightarrow \quad T_{ij} = T_{ji} \; . \tag{336}$$

The stress tensor must be a *symmetric tensor*, otherwise all the fluid elements would start to rotate on their own initiative. See Figure 22.

Coming back to the stress-energy tensor $T^{\alpha\beta}$ of the electromagnetic field, if its spatial components are to give us a symmetric stress tensor then the full stress-energy tensor must be symmetric too. And it is not so hard to guess what it must be. It has to be quadratic in the field strength, and it needs two indices. There are only two obvious terms with these properties that we can form, and we expect the stress-energy tensor to be a linear combination of the two. Thus

$$T^{\alpha\beta} = a_1 F^{\alpha}_{\;\gamma} F^{\gamma\beta} + a_2 \eta^{\alpha\beta} F_{\gamma\delta} F^{\alpha\beta} \; . \tag{337}$$

Now we calculate

$$F_{\gamma\delta} F^{\gamma\delta} = F_{ij} F^{ij} + 2 F_{0i} F^{0i} = \; \ldots \; = -2(E_i E_i - B_i B_i) \tag{338}$$

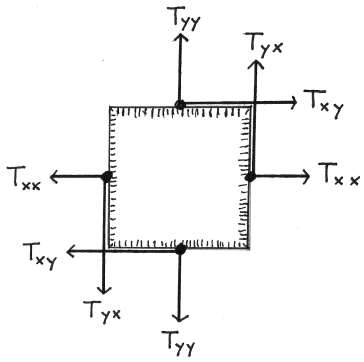

Figure 22: The surface forces acting on a small fluid element. If $T_{xy} \neq T_{yx}$ it would start to rotate.

and similarly

$$F^0{}_\gamma F^{0\gamma} = E_i E_i. \tag{339}$$

We must ensure that the energy density T^{00} comes out correctly. Recalling that $\eta^{00} = -1$ we are led to propose that *the stress-energy tensor of the electromagnetic field is*

$$T^{\alpha\beta} = \frac{1}{4\pi}\left(F^{\alpha\gamma}F^{\beta}{}_{\gamma} - \frac{1}{4}\eta^{\alpha\beta}F_{\gamma\delta}F^{\gamma\delta}\right) \ . \tag{340}$$

To make local conservation of energy come out correctly, in accordance with equation (329), the components T^{i0} must equal the Poynting vector. And indeed[83]

$$T^{i0} = -\frac{1}{4\pi}F_{ij}F_{0j} = \frac{1}{4\pi}\epsilon_{ijk}E_j B_k \ . \tag{341}$$

A calculation using the tensor formulation of Maxwell's equations shows that

$$\partial_\gamma T^{\gamma\beta} = J_\gamma F^{\gamma\beta} \ . \tag{342}$$

I leave that to you as an exercise.[84]

[83]Exercise: Check all the statements in this paragraph.

[84]Exercise: Using the tensor formulation, prove that the energy-momentum tensor of the electromagnetic field obeys $\partial_\gamma T^{\gamma\beta} = J_\gamma F^{\gamma\beta}$. When you have done this, you know how to handle indices.

But there is more to the story than energy conservation. In components we obtain

$$\partial_\gamma T^{\gamma 0} = J_\gamma F^{\gamma 0} \quad \Rightarrow \quad \partial_t u(x) + \nabla \cdot \mathbf{S} = -\mathbf{J} \cdot \mathbf{E} \tag{343}$$

$$\partial_\gamma T^{\gamma i} = J_\gamma F^{\gamma i} \quad \Rightarrow \quad \partial_t S_i + \partial_j T_{ji} = -\rho E_i - \epsilon_{ijk} J_j B_k \ , \tag{344}$$

where the *Maxwell stress tensor* is

$$T_{ij} = -\frac{1}{4\pi}\left(E_i E_j - \frac{1}{2}\delta_{ij}E^2 + B_i B_j - \frac{1}{2}\delta_{ij}B^2\right) \ . \tag{345}$$

The first conservation equation is our familiar formulation of local conservation of energy. The second equation reveals that the Poynting vector plays a dual role, because $S^i = T^{i0} = T^{0i}$ can be interpreted either as the energy density flow or as the *momentum density* of the electromagnetic field. In relativity theory these two are the same thing (up to factors of c). The stress tensor plays a role in the local balance of momentum, and it provided Maxwell with the means to banish action-at-a-distance from physics.[85]

We need to add two remarks. First, in relativity theory every kind of matter has a stress-energy tensor associated to it. A very special property of the electromagnetic stress-energy tensor is that

$$\eta_{\alpha\beta} T^{\alpha\beta} = T_\alpha{}^\alpha = 0 \ . \tag{346}$$

This is actually a good way to remember the factor "$-1/4$" in the definition.

The second remark is that our identification of the local energy density and the local energy flow in electrodynamics can now be set on a firm footing, if we bring in the ideas of General Relativity. To do so we rewrite Maxwell's equations in a way that allows spacetime to be curved, so that $g_{\alpha\beta}$ is no longer equal to the Minkowski space metric tensor. This gives

$$\frac{1}{\sqrt{-g}}\partial_\beta(\sqrt{-g}F^{\alpha\beta}) = 0 \ , \qquad F_{\alpha\beta} = \partial_\alpha A_\beta - \partial_\beta A_\alpha \ . \tag{347}$$

[85]Exercise: Calculate T_{ij} for a purely electric field. Choose coordinates so that the x-axis lies along the electric field. Can you interpret the result?

Here g is the determinant of the spacetime metric, which is negative, hence the sign under the square root. To complete the theory we write down Einstein's field equations for the metric tensor. It takes the form

$$G_{\alpha\beta}(g, \partial g, \partial\partial G) = \frac{8\pi G}{c^4} T_{\alpha\beta} \ . \tag{348}$$

Here $G_{\alpha\beta}$ is a tensor formed from the metric tensor and its first and second derivatives. It would take us too far afield to explain the details here. The point is that on the right hand side we find the stress-energy tensor of the electromagnetic field, and it can be shown that equation (340) is the only possibility that is consistent with Maxwell's equations (347). This is to say that coupling to gravity resolves the ambiguities we encountered when we tried to define the energy density of the electromagnetic field.

A derivation of Lorentz' equation

In the earlier parts of these notes we concluded that the electromagnetic field carries energy and momentum. The arguments we used were based on Lorentz' equation (7), which tells us how the field does work on charged particles. But now there is a new claim: When coupling electromagnetism to gravity Einstein's general relativity theory demands that there is a stress-energy tensor associated to the electromagnetic field. Moreover it has to take the specific form (340). The momentum density appears as part of this stress-energy tensor, and turns out to equal the Poynting vector field that we had arrived at already. Moreover the Poynting vector obeys the balance equation (344). But the right hand side of this equation tells us that the momentum density is changing because of a force exerted on the matter, as described by ρ and J_i. This is to say that the electromagnetic field acts on matter with a force per unit volume equal to

$$f_i = \rho E_i + \epsilon_{ijk} J_j B_k \ . \tag{349}$$

In effect, this is Lorentz' equation. We can therefore claim that the latter is a consequence of the way in which electromagnetism couples to gravity.

For the exercise class:

- Consider a scalar field having the form

$$\phi(x) = \frac{1}{r^2} \ , \qquad r^2 = x^2 + y^2 + z^2 \ .$$

Perform a Lorentz boost in the $t-x$-plane, and express the new function ϕ' that you obtain in this way as a function of the coordinates (t, x, y, z). What does the new function look like?

- Consider the electromagnetic field from a point charge at rest at the origin,

$$E_i(x) = \frac{1}{r^3} \begin{pmatrix} x \\ y \\ z \end{pmatrix} \qquad B_i(x) = 0 \ .$$

Perform a Lorentz boost in the $t - x$-plane. Compute the electromagnetic field you obtain, and express it as a function of the coordinates (t, x, y, z).

- Repeat the previous exercise for the electromagnetic field

$$E_2 = \cos(t - x) \qquad B_3 = \cos(t - x) \ ,$$

all other components vanishing.

THE WAVE EQUATION

We have a satisfactory theory of the Laplace equation and of the Laplace equation including sources, also known as Poisson's equation. To tackle electrodynamics we need an equally satisfactory theory of the *wave equation*

$$\Box\Psi = \left(\partial_x^2 + \partial_y^2 + \partial_z^2 - \frac{1}{c^2}\partial_t^2\right)\Psi(x,y,z,t) = 0 \tag{350}$$

and of the wave equation driven by sources,

$$\Box\Psi(x,y,z,t) = -4\pi f(x,y,z,t)\ . \tag{351}$$

The operator on the left hand side is known as the *d'Alembert operator*. It plays a starring role throughout the second half of Jackson's book, but all I can do here is to collect a few highlights.

The wave equation in $1+1$ *dimensions*

We begin in $1+1$ dimensions, where the wave equation is

$$\left(\partial_x^2 - \frac{1}{c^2}\partial_t^2\right)\Psi(x,t) = 0\ . \tag{352}$$

This is just about the simplest of all second order partial differential equations, and was first studied by d'Alembert in connection with the vibrating string. It is interesting to compare and contrast its general solution with the general solution of the two dimensional Laplace equation. We solve the wave equation by transforming to the new variables

$$u = x - ct\ , \qquad v = x + ct\ . \tag{353}$$

The equation becomes

$$\partial_u\partial_v\Psi(u,v) = 0\ . \tag{354}$$

The general solution is therefore

$$\Psi(x,t) = f(v) + g(u) = f(x+ct) + g(x-ct)\ . \tag{355}$$

Two free functions appear in it. These functions can be quite general. We can regard the formula as defining a generalized or *weak* solution of the wave equation even if f and g are not differentiable.

To get a completely specified problem we supplement the equation with *initial data* specified at $t = 0$. Let

$$\Psi(x,0) = u_0(x) \ , \qquad \partial_t \Psi(x,0) = u_1(x) \ . \tag{356}$$

Here u_0, u_1 are functions that we can choose as we please. Then we consider the system of equations

$$\begin{cases} u_0(x) &= f(x) + g(x) \\ \\ u_1(x) &= cf'(x) - cg'(x) \ . \end{cases} \tag{357}$$

The solution is

$$\begin{cases} f(x) &= \frac{1}{2}u_0(x) + \frac{1}{2c}\int^x u_1(s)\mathrm{d}s \\ \\ g(x) &= \frac{1}{2}u_0(x) - \frac{1}{2c}\int^x u_1(s)\mathrm{d}s \ . \end{cases} \tag{358}$$

The free functions appearing in the general solution are then completely fixed, and the solution to our *Cauchy problem* is

$$\Psi(x,t) = \frac{u_0(x+ct) + u_0(x-ct)}{2} + \frac{1}{2c}\int_{x-ct}^{x+ct} u_1(s)ds \ . \tag{359}$$

This is known as *d'Alembert's formula*. Once the initial data are specified, so is the solution, for all times.

Let us compare this with the general solution of the Laplace equation

$$\left(\partial_x^2 + \partial_y^2\right)\Phi(x,y) = 0 \ . \tag{360}$$

We know that the general solution can be characterized by fixing the values of the function on a boundary completely surrounding the region in which the solution is being sought, or alternatively by fixing the normal derivative on the boundary. For the wave equation we specify both, but not on a boundary surrounding the region. See Figure 23. The nature of the solutions is also very different. Solutions of the wave equation need not even be continuous.

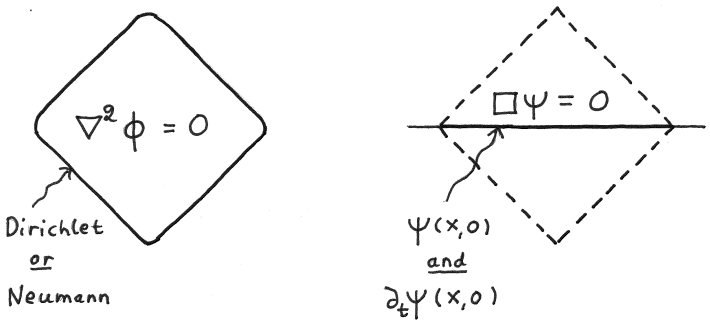

Figure 23: The boundary value problem for the Laplace equation, and the Cauchy problem for the wave equation. In the latter case the region where the initial data determine the solution is known as the *Cauchy development*.

For the Laplace equation the solution in the interior is the real part of an analytic function, the most regular kind of function known to mathematics.[86]

Plane waves and spherical waves

We can put the $1+1$ dimensional solution to work right away. First of all it describes *plane waves* in any higher dimension—just assume that the solution is to depend on only one of the spatial coordinates. This is important because any propagating wave in three dimensions will look like a plane wave within a small region far away from the source of the radiation.

But we can also apply d'Alembert's solution to spherically symmetric waves in three dimensions. Suppose that we are looking for solutions of the wave equation that do not depend on the angular coordinates at all. So we assume $\Psi = \Psi(r,t)$. If the dimension of space equals d the wave equation reduces to

$$\left(\partial_r^2 + \frac{d-1}{r}\partial_r\right)\Psi = \frac{1}{c^2}\partial_t^2\Psi \ . \tag{361}$$

[86]Exercise: You could imagine trying to solve the Laplace equation too as an initial value problem. Find a solution of the wave equation obeying $\Psi(x,0) = A\cos nx/n$ and $\partial_t\Psi = 0$ at $t = 0$. Then find a solution of the Laplace equation obeying $\Phi(x,0) = A\cos nx/n$ and $\partial_y\Phi = 0$ at $y = 0$. You can choose n to be any integer. Why is this solution of the Laplace equation useless in practice?

We compare this with the identity

$$\partial_r^2(r^a\Psi) = r^a\left(\partial_r^2\Psi + \frac{2a}{r}\partial_r\Psi + \frac{a(a-1)}{r^2}\Psi\right) \ . \tag{362}$$

If we set $a = 1$ and $d = 3$ we see that the equation for spherical waves becomes

$$\partial_r^2\,(r\Psi) = \frac{1}{c^2}\partial_t^2\,(r\Psi) \ . \tag{363}$$

The general solution of this equation is easily obtained from d'Alembert's formula. In fact it is

$$\Psi = \frac{1}{r}f(r+ct) + \frac{1}{r}g(r-ct) \ , \tag{364}$$

where f and g are arbitrary functions. There is still a significant difference between one and three dimensions though, because when r is large the function Ψ itself falls off like $1/r$. See Figure 24, where f has been set to zero in order to get a purely outgoing wave.

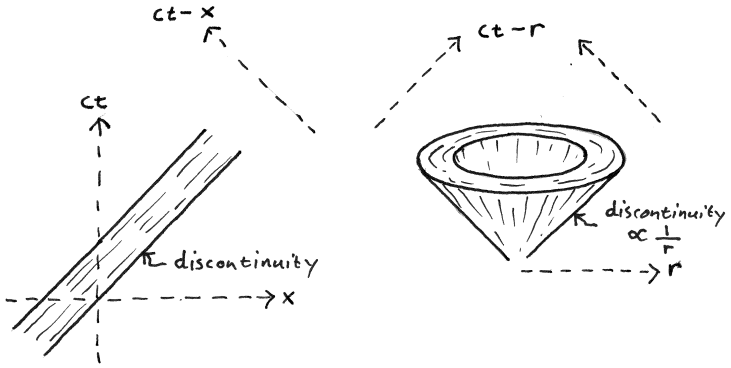

Figure 24: Plane waves and spherical waves depend on initial data that can be chosen to be discontinuous, and then the discontinuities propagate through spacetime. But the discontinuities in the spherical wave fall off with distance.

It is clearly a pleasant surprise that the physical dimension $d = 3$ turns out to be especially simple to treat.[87] As in the one dimensional case, the

[87]Exercise: Try the more general Ansatz $\Phi = c_1 r^a\Psi + c_2 r^{a+1}\partial_r\Psi$ where $\Psi = \Psi(r,t)$

spherical wave can be turned on and turned off at will. One consequence, as will become clear when we study the Green function, is that when someone plays the piano you hear each note distinctly and only once. Given that sound waves obey the wave equation this simply would not be true in a space of only two dimensions, where the effect of each individual note would 'ring on' in a way that would disturb the pleasure. We say that the three dimensional wave equation obeys *Huygens's principle in its strong form*, while the two dimensional wave equation does not. Exactly what we mean by this phrase will become clear soon (although for two dimensions you will have to consult Jackson's problem 6.1.a). A condition that can be imposed in any dimension is that there are no waves propagating into the past, a seemingly obvious statement that will require a condition on the Green function.

Green functions for the wave equation

To solve the initial value problems that we will encounter we want to have the 'inverse' of the d'Alembert operator at hand. That is to say, we want to find a Green function $G(x, x')$ such that

$$\Box G(x, x') = \delta(x, x') \ , \tag{365}$$

where the argument consists of Minkowski space coordinates, that is to say it includes both space and time. We have to carefully specify the conditions under which the inverse exists and is unique because, like the Laplace equation, the wave equation admits homogeneous solutions.

In his section 6.4 Jackson's opening move is a Fourier transformation with respect to time. He defines

$$\tilde{\Psi}(\mathbf{x}, \omega) = \int_{-\infty}^{\infty} \Psi(\mathbf{x}, t) e^{i\omega t} \, \mathrm{d}t \ . \tag{366}$$

This turns the wave equation into the *Helmholtz equation*,

$$\left(\nabla^2 - \frac{1}{c^2} \partial_t^2 \right) \Psi(\mathbf{x}, t) = 0 \quad \Leftrightarrow \quad \left(\nabla^2 + \frac{\omega^2}{c^2} \right) \tilde{\Psi}(\mathbf{x}, \omega) \ . \tag{367}$$

solves the d dimensional wave equation. In what dimensions can you arrange that Φ solves the $1 + 1$ dimensional wave equation?

Then Jackson reaches the answer rather quickly. We will however proceed in a more direct manner here.

Recall that the aim is to be able to write the solution of the inhomogeneous equation

$$\Box\Psi = f \tag{368}$$

on the form

$$\Psi(\mathbf{x},t) = \int G(\mathbf{x},t;\mathbf{x}',t')f(\mathbf{x}',t')\ \mathrm{d}^3x'\mathrm{d}t'\ . \tag{369}$$

Now the source $f(\mathbf{x},t)$ may have the property that $f=0$ for $t<0$. Then we are interested in the situation that there are no waves present when $t<0$, and that the disturbance propagates outwards from the source with the speed c. Hence we insist that

$$G(\mathbf{x},t;\mathbf{x}',t') = 0 \quad \text{if} \quad t < t' + \frac{|\mathbf{x}-\mathbf{x}'|}{c}\ . \tag{370}$$

The particular Green function that obeys this condition is called the *retarded Green function*, and denoted G_{ret}. It has the property that we see the effects of a change in the source only *after* it has happened. There is an *advanced* Green function as well, in which we see them only before it has happened. Mathematically, both are legitimate, but the retarded Green function is better adapted to the physics. Either way, we know from our discussion of the Laplace operator that, because homogeneous solutions exist, it is necessary to impose some conditions before we can invert the operator.

We now recall Huygens' principle in its strong form. The signal from the source should be heard only once. To build this in we also impose

$$G(\mathbf{x},t;\mathbf{x}',t') = 0 \quad \text{if} \quad t > t' + \frac{|\mathbf{x}-\mathbf{x}'|}{c}\ . \tag{371}$$

If this can be arranged then the retarded Green function must take the form

$$G_{\mathrm{ret}}(\mathbf{x},t;\mathbf{x}',t') = g(\mathbf{x},\mathbf{x}')\delta\left(t-t'-\frac{|\mathbf{x}-\mathbf{x}'|}{c}\right)\ . \tag{372}$$

For dimensional reasons the function in front has to have the dimension of inverse length, and because of translational symmetry it can only depend on $|\mathbf{x}-\mathbf{x}'|$. So we have reached the Ansatz

$$G_{\rm ret}(\mathbf{x},t;\mathbf{x}',t') = \frac{\delta\left(t-t'-\frac{|\mathbf{x}-\mathbf{x}'|}{c}\right)}{|\mathbf{x}-\mathbf{x}'|} \ . \tag{373}$$

It only remains to check that this function (or distribution) obeys $\Box G = \delta$.

This is a straightforward calculation, especially since we can use spherical polar coordinates for the calculation. Nothing depends on the angles, so what we do is to observe that[88]

$$\Box G_{\rm ret}(r,t) = \left(\nabla^2 - \frac{1}{c^2}\partial_t^2\right)\frac{\delta\left(\frac{r}{c}-t\right)}{r} = \nabla\cdot\left(\frac{1}{r}\nabla\delta + \delta\nabla\left(\frac{1}{r}\right)\right) - \frac{1}{c^2}\frac{\delta''}{r} = \tag{374}$$

$$= \frac{1}{r}\nabla^2\delta + 2\nabla\left(\frac{1}{r}\right)\cdot\nabla\delta + \delta\nabla^2\left(\frac{1}{r}\right) - \frac{1}{c^2}\frac{\delta''}{r} = \ldots = \delta\nabla^2\left(\frac{1}{r}\right) \ .$$

But we know that $\nabla^2(1/r) = -4\pi\delta^{(3)}(\mathbf{x})$, so—after adjusting a constant—we have verified that

$$\Box G_{\rm ret}(r,t) = -4\pi\delta^{(4)}(\mathbf{x},t) \ . \tag{375}$$

The conclusion is that, once we have imposed the condition that causes come before effects, the Green function (373) has the property that

$$\Box\Psi = f \quad \Leftrightarrow \quad \Psi(\mathbf{x},t) = -\frac{1}{4\pi}\int G_{\rm ret}(\mathbf{x},t;\mathbf{x}',t')f(\mathbf{x}',t')\ \mathrm{d}^3x'\mathrm{d}t' \ . \tag{376}$$

When we apply this to radiation problems in electrodynamics we will come to appreciate the meaning of the 'retarded' condition much better.

Monopole and dipole radiation

Let us consider the simplest possible source of radiation: We choose it to be a function of the form $f(\mathbf{x},t) = F(t)\delta^{(3)}(\mathbf{x})$. Using the expression for the retarded Green function we immediately obtain the signal

[88] Exercise: Fill in the dotted part of the calculation.

$$\Psi(\mathbf{x},t) = -\frac{1}{4\pi}\frac{F(t-r/c)}{r} \ . \tag{377}$$

Wherever we are we can read off what happened at the origin, but only with a time delay, and the strength of the signal decreases linearly with the distance r. It travels with velocity c, which in this application may be thought of as the velocity of sound.

With a view to what happens in electrodynamics, let us now consider a dipole source. We place one monopole source at the origin and one of opposite signature at a point displaced with ϵ along the z-axis, so that the source is

$$s(\mathbf{x},t) = F(t)\delta^{(3)}(\mathbf{x}) - F(t)\delta^{(3)}(\mathbf{x}') \ , \qquad \mathbf{x}' = \mathbf{x} - \epsilon \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \ . \tag{378}$$

We take the limit $\epsilon \to 0$ in such a way that $f(t) = \epsilon F(t)$ remains finite. Again applying the retarded Green function we obtain, in this limit, the signal[89]

$$\Psi(\mathbf{x},t) = \frac{\cos\theta}{4\pi}\partial_r\left(\frac{f(t-r/c)}{r}\right) = -\frac{\cos\theta}{4\pi}\left(\frac{f'(t-r/c)}{cr} + \frac{f(t-r/c}{r^2}\right) \ . \tag{379}$$

The point to notice here is that there are two quite different regions of space to pay attention to, the near zone where the second term dominates and in which we observe the field from an oscillating dipole, and a far zone at large r in which only the first term matters and we see radiation falling off with distance in the way we expect it to do. As we will soon see, in electrodynamics there are no monopole sources, and hence there is always a near zone and a far zone to consider.

Electromagnetic radiation

Now for a first look at the electromagnetic case. In vacuum the components of the electric and magnetic fields obey the wave equation, with c = the speed of light. In the presence of sources Maxwell's equations imply

[89] Exercise: Show this, using some care.

$$\Box \mathbf{E} = -\frac{1}{\epsilon_0}\left(-\nabla\rho - \frac{1}{c^2}\partial_t \mathbf{J}\right) \tag{380}$$

$$\Box \mathbf{B} = -\mu_0 \nabla \times \mathbf{J} \ . \tag{381}$$

If we know how the sources ρ and $\mathbf{J}$ behave we can calculate the electric and magnetic fields. For example

$$\mathbf{B}(\mathbf{x},t) = \mathbf{B}_{\rm in}(\mathbf{x},t) - \frac{1}{4\pi}\int G_{\rm ret}(\mathbf{x},t;\mathbf{x}',t')(-\mu_0 \nabla' \times \mathbf{J}(\mathbf{x}',t'))\ {\rm d}t'{\rm d}^3x' \ . \tag{382}$$

Here the in-field $\mathbf{B}_{\rm in}$ is any solution of the homogeneous equation $\Box\mathbf{B} = 0$. It is not 'caused' by the sources. We set it to zero for simplicity, that is $\mathbf{B}_{\rm in} = 0$ from now on, and turn to the interesting second term. Because of the delta function in the definition of $G_{\rm ret}$, eq. (373), we can do the integral over t' directly, but once this is done the quantities in the integrand have to be evaluated at the retarded time

$$t_{\rm ret} = t' = t - \frac{|\mathbf{x}-\mathbf{x}'|}{c} \ . \tag{383}$$

With this understanding our solution for the magnetic field is

$$\mathbf{B}(\mathbf{x},t) = \frac{\mu_0}{4\pi}\int \frac{[\nabla' \times \mathbf{J}']_{\rm ret}}{|\mathbf{x}-\mathbf{x}'|}\ {\rm d}^3x' \ . \tag{384}$$

For the electric field we obtain in a similar way that

$$\mathbf{E}(\mathbf{x},t) = \frac{1}{4\pi\epsilon_0}\int \frac{\left[-\nabla'\rho' - \frac{1}{c^2}\partial'_t \mathbf{J}'\right]_{\rm ret}}{|\mathbf{x}-\mathbf{x}'|}\ {\rm d}^3x' \ . \tag{385}$$

Because the integrands are to be evaluated at the retarded time, it takes a little effort to get used to these formulas.

Physically it is clear what is going on. The fields evaluated at some observation point $\mathbf{x}$ at time t is affected by the behaviour of the sources in the past, and more precisely by their behaviour on the past lightcone of the observation point. See Figure 25.

We still have to be careful when manipulating the formulas. In particular it should be kept in mind that

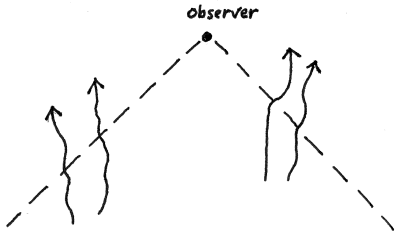

Figure 25: At the point $\mathbf{x}$ we detect radiation from accelerated charges crossing the backwards lightcone at some retarded time.

$$[\nabla' \rho' \,]_{\rm ret} \neq \nabla' \, [\rho' \,]_{\rm ret} \ . \tag{386}$$

To see why, we write things out. It will be convenient to introduce

$$\mathbf{R} = |\mathbf{x} - \mathbf{x}'| \ , \qquad \mathbf{e}_R = \frac{\mathbf{x} - \mathbf{x}'}{|\mathbf{x} - \mathbf{x}'|} \ . \tag{387}$$

The point is that $[\rho'(\mathbf{x}', t')]_{\rm ret}$ depends on $\mathbf{x}'$ in two different ways. Indeed

$$\begin{aligned} \nabla' \, [\rho'(\mathbf{x}', t') \,]_{\rm ret} &= \nabla' \rho' \left(\mathbf{x}', t - \frac{\mathbf{R}}{c} \right) = \\ &= [\nabla' \rho']_{\rm ret} + [\partial_{t'} \rho']_{\rm ret} \, \nabla' \left(t - \frac{\mathbf{R}}{c} \right) = [\nabla' \rho']_{\rm ret} + \frac{\mathbf{e}_R}{c} \, [\partial_{t'} \rho']_{\rm ret} \ . \end{aligned} \tag{388}$$

After performing a similar calculation for the current we can conclude that

$$[\nabla' \rho']_{\rm ret} = \nabla' \, [\rho']_{\rm ret} - \frac{1}{c} \, [\partial_{t'} \rho']_{\rm ret} \, \mathbf{e}_R \tag{389}$$

$$[\nabla' \times \mathbf{J}']_{\rm ret} = \nabla' \times [\mathbf{J}']_{\rm ret} + \frac{1}{c} \, [\partial_{t'} \mathbf{J}']_{\rm ret} \times \mathbf{e}_R \ . \tag{390}$$

These are the expressions that we must use in the integrands of equations (384 - 385).

You will observe that the solution contains a factor falling off like $1/r$ in the integrand. Nevertheless, in the static case the solution must reduce to our

previous solutions for the electric and magnetic fields from static sources,[90] even though the latter fall off like $1/r^2$. However, you will also recall that it is easier to obtain the static solutions if we first solve for the electric potential and for the vector potential. Radiation problems can be handled in an analogous way, and we turn to this task next.

For the exercise class:

- Consider a three dimensional surface in spacetime, given by some function $f(\mathbf{x}, t) = 0$. What conditions do you have to impose on the hypersurface in order to guarantee that there exists a solution of the wave equation which is zero on one side of this 'hypersurface', and non-zero on the other? Give two interesting examples.

- Consider the driven harmonic oscillator

$$\ddot{x} + \omega_0^2 x = f(t) \ .$$

  Using Fourier transformation find a Green function $G_{\rm R}(t)$ obeying

$$\ddot{G}_{\rm R} + \omega_0^2 G_{\rm R} = \delta(t) \ , \qquad G_{\rm R}(t) = 0 \text{ if } t < 0 \ .$$

  Use it to write down the general solution for the driven harmonic oscillator. Repeat the exercise for the damped and driven harmonic oscillator

$$\ddot{x} + \gamma\dot{x} + \omega_0^2 x = f(t) \ , \qquad \gamma > 0 \ .$$

  How much of this can you do if you replace the condition on the Green function with the condition $G_{\rm A} = 0$ for $t > 0$?

- In a d-dimensional space ($d > 2$) the Green function for the Laplace equation is (up to a constant) $G(r) = 1/r^{d-2}$. Prove that $\delta(r-t)/r^{d-2}$ is a Green function for the wave equation if and only if $d = 3$.

- Jackson, problem 6.1.

[90] Exercise: Verify that it does.

RADIATION IN THE LORENZ GAUGE[91]

Let us now come to grips with electromagnetic radiation making full use of the vector potential, the tensor formalism, and (of course) Gaussian units.

The vector potential, tensorially

Maxwell's equations are the two tensor equations

$$\begin{cases} \partial_\beta F^{\alpha\beta} = 4\pi J^\alpha \\ \\ \partial_\alpha F_{\beta\gamma} + \partial_\gamma F_{\alpha\beta} + \partial_\beta F_{\gamma\alpha} = 0 \qquad . \end{cases} \tag{391}$$

Next we bring in a kind of generalized Stokes' theorem, valid in any dimension. It says that

$$\left.\begin{array}{l} F_{\alpha\beta} = -F_{\beta\alpha} \\ \\ \partial_\alpha F_{\beta\gamma} + \partial_\gamma F_{\alpha\beta} + \partial_\beta F_{\gamma\alpha} = 0 \end{array}\right\} \quad \Leftrightarrow \quad F_{\alpha\beta} = \partial_\alpha A_\beta - \partial_\beta A_\alpha \ . \tag{392}$$

This is to say that such a vector field A_α exists if and only if there exists a tensor $F_{\alpha\beta}$ obeying the stated conditions. That vector field is determined only up to a gradient because

$$\partial_\alpha A_\beta - \partial_\beta A_\alpha = 0 \quad \Leftrightarrow \quad A_\alpha = \partial_\alpha \Lambda \tag{393}$$

for an arbitrary function Λ. The theorem holds in any dimension provided only that every loop can be contracted to a point, and it is easy to prove in one direction. We take the other direction on trust.[92]

This solves one out of the two Maxwell equations. The other becomes

$$\partial_\beta F^{\alpha\beta} = \partial_\beta(\partial^\alpha A^\beta - \partial^\beta A^\alpha) = -\Box A_\alpha + \partial^\alpha(\partial_\beta A^\beta) = 4\pi J^\alpha \ . \tag{394}$$

[91]Reference: This is not a misprint. See H. Kragh, *Ludvig Lorenz on Light and Electricity*, arXiv:1803.06371.

[92]Exercise: We know that the existence of the vector potential $\mathbf{A}$ follows from the condition $\nabla \cdot \mathbf{B} = 0$. Show that this is a consequence of the theorem in the text.

If the ambiguity in A_α can be utilized to ensure that

$$\partial \cdot A = \partial_\alpha A^\alpha = 0 \tag{395}$$

then the equation left to solve is

$$\Box A^\alpha = -4\pi J^\alpha \ . \tag{396}$$

If the sources are known the solution is

$$A^\alpha(x) = \int G_{\rm ret}(x;x') J^\alpha(x') \ {\rm d}^4x' \ . \tag{397}$$

It is over. But we have to go over the argument again, first to see how it is connected to the familiar vector formulation of Maxwell's equations, and then to make sure that we can indeed assume that $\partial \cdot A = 0$.

Translation

The components of J^α are $J^0 = \rho$ and $J^i = \mathbf{J}^i/c$. So it is immediately clear that

$$\partial_\alpha J^\alpha = 0 \quad \Leftrightarrow \quad \partial_t \rho + \nabla \cdot \mathbf{J} = 0 \ . \tag{398}$$

A little more care is needed to see how the electric potential enters the tensor equations. Set

$$A_\alpha = (A_0, \mathbf{A}) \ . \tag{399}$$

Recall that the claim that the electric field is the gradient of a function rested on the assumption that $\nabla \times \mathbf{E} = 0$, which is true only in the static case. According to how the 4-vector potential was introduced we now have

$$E_i = F_{i0} = \partial_i A_0 - \partial_0 A_i \tag{400}$$

$$B_i = \frac{1}{2}\epsilon_{ijk} F_{jk} = \epsilon_{ijk} \partial_j A_k \ . \tag{401}$$

So the magnetic field is indeed given as the curl of the spatial part of A_α. Remembering that $\partial_0 = 1/c \ \partial_t$ we define its time component as

$$A_0 = -\Phi \ , \tag{402}$$

and obtain for the electric field that

$$\mathbf{E} = -\nabla\Phi - \frac{1}{c}\partial_t \mathbf{A} \ . \tag{403}$$

The last term was missing in the static case.[93] [94]

The Lorenz gauge

When doing magnetostatics we found it convenient to impose the Coulomb gauge $\nabla \cdot \mathbf{A} = 0$. We reached this gauge by solving Poisson's equation for a suitable function Λ to be added to any vector potential that does not obey this condition. We could repeat this maneouvre here, but we would prefer to impose the *Lorenz gauge*

$$\partial \cdot A = \partial_\alpha A^\alpha = \nabla \cdot \mathbf{A} - \frac{1}{c}\partial_t A_0 = \nabla \cdot \mathbf{A} + \frac{1}{c}\partial_t \Phi = 0 \ . \tag{404}$$

Can this always be done? To answer this question we argue as follows: First we use Maxwell's equations to show that

$$\Box \partial \cdot A = -4\pi \partial_\alpha J^\alpha = 0 \ . \tag{405}$$

Hence, in every solution of Maxwell's equations, the field $\partial \cdot A$ obeys the homogeneous wave equation also in the presence of sources. But this means that we can impose the condition $\partial \cdot A = 0$ as part of the initial conditions for Maxwell's equations, and then the wave equation guarantees that $\partial \cdot A = 0$ for all times.

It is worth noting though that imposing the Lorenz gauge is a somewhat different matter from that of imposing the Coulomb gauge. With suitable boundary conditions the condition $\nabla \cdot \mathbf{A} = 0$ determines the vector potential uniquely. But consider the 4-vector potentials A_α and A'_α, where

[93] Exercise: In Gaussian units Faraday's law reads $\nabla \times \mathbf{E} + 1/c \ \partial_t \mathbf{B} = 0$. Derive eq. (403) directly from this.

[94] In vacuum a plane electromagnetic wave is defined by $\mathbf{E} = (f(z-ct), g(z-ct, 0)$, where f and g are arbitrary functions. Show that you obtain a solution of Maxwell's equations with a suitable choice of $\mathbf{B}$. Also find the four vector potential A_α.

$$A'_\alpha = A_\alpha + \partial_\alpha \Lambda \ . \tag{406}$$

Suppose that $\partial \cdot A = 0$. Provided that the function Λ is a solution of the homogeneous wave equation, $\Box\Lambda = 0$, the gauge transformed vector potential will also obey the Lorenz condition, $\partial \cdot A' = 0$. This subtlety has quite a few consequences in quantum electrodynamics, but we do not have to worry about it here.

We then have only the wave equation left to solve. In the Lorenz gauge the general solution to Maxwell's equations driven by known sources is derived from the 4-vector potential

$$\begin{aligned} A^\alpha(x) = A^\alpha_{\rm hom}(x) + \int G_{\rm ret}(x,x') J^\alpha(x') \ {\rm d}^4x' = \\ = A^\alpha_{\rm hom}(x) + \int \frac{[J^\alpha(x')]_{\rm ret}}{|\mathbf{x}-\mathbf{x}'|} \ {\rm d}^3x' \ , \end{aligned} \tag{407}$$

where $A^\alpha_{\rm hom}$ is any solution of the vacuum equations (with no sources).

Unfortunately we do not have the time needed to do much with this formula. It would take us into the second half of Jackson.

Radiation from accelerated charges

Figure 25 remains relevant: At a given point the solution is affected by the sources only by their behaviour on the past light cone of that point. But this raises many questions. In space, the electric field lines emanating from a static charge point straight out from (or into) the charge. This is natural. But we found, by Lorentz transforming that solution, that the electric field lines continue to point out from the momentary position of the charge also when it is moving with constant velocity. See Jackson's Figure 11.9 and our Figure 26. The direction of the field lines seems to have nothing to do with the position of the charge at the retarded time. There is actually no problem with this. If the particle moves with constant velocity its position at any time is predictable at the retarded time.

Continuing this train of thought, we may consider the electric field lines emanating from an electric charge that was moving with constant velocity

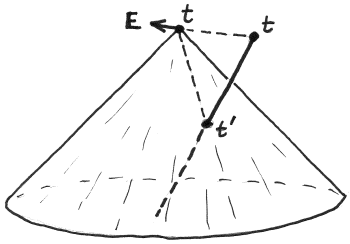

Figure 26: At any given time the field lines from a charged particle moving with constant speed point straight out from its present position.

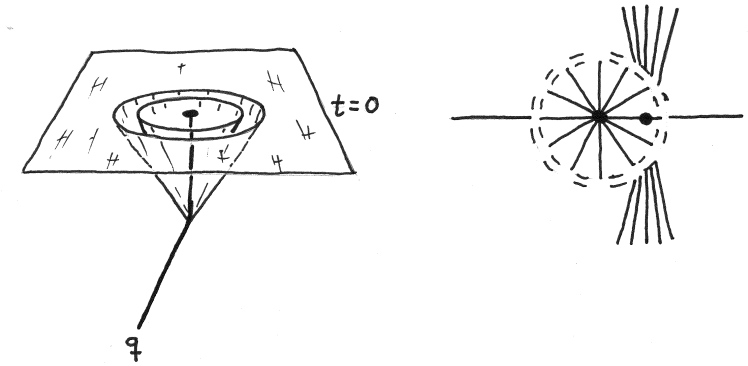

Figure 27: If a particle suddenly changes its speed it will be necessary to connect the field lines (at $t = 0$) in a way that is left for you as an exercise.

until some time $t_0 < 0$ in the past (taking $t = 0$ to be the present), then decelerated sharply, and became static shortly after. This is depicted in Figure 27. On reflection one sees that, at $t = 0$, there will be a sphere around the position of the charge where the electric field lines point radially outwards from the charge the way they do in electrostatics. There will also be a slightly bigger sphere such that none of the points outside it have been reached by the information that the charge stopped moving at $t = t_0$. In this region the field lines must be pointing out from the position that the charge would have reached, had it continued to move with constant velocity. Jackson's Figure 11.9 also shows that they crowd together along the directions that are orthogonal to the velocity vector. But the field lines must exist also in between the spheres. Thinking this through one realises that there must be a shell of transverse electric fields moving out with the speed of light from

the final position of the electric charge.[95] [96] This explains why an electric charge must emit radiation when it changes its state of motion.

It is important to see how the amplitude of the fields fall off with distance. For a static charge the inverse square law holds, so the electric field E from a charge e is $E \sim e/r^2$. If we assume (as turns out to be correct) that the radiation fields coming from an accelerated particle depend linearly on its acceleration a, dimensional analysis shows that

$$E \sim \frac{ea}{rc^2} \ . \tag{408}$$

Since the amplitude falls off only like $1/r$ this means that electromagnetic radiation can travel long distances before it becomes too weak to detect.[97]

Radiation in the Coulomb gauge

While the Lorenz gauge is very useful in radiation problem, it is still true that we can fix the gauge freedom by imposing the Coulomb gauge $\partial_i A_i = 0$ also in the time dependent case. But then there is an apparent paradox when we look at Maxwell's equations, because the time derivatives drop out of one of them:

$$\Box A_0 - \partial_0(\partial_i A_i - \partial_0 A_0) = \nabla^2 A_0 = -4\pi J_0 \tag{409}$$

$$\Box A_i - \partial_i(\partial_j A_j - \partial_0 A_0) = \Box A_i + \partial_i \partial_0 A_0 = -4\pi J_i \ . \tag{410}$$

This means that the time component A_0 is now determined by Poisson's equation also in the time dependent case. Thus, at a given time it is determined by what the source is doing at the *same* instant of time. Does this contradict relativistic causality?

[95] Exercise: Make a sketch showing how the electric field lines in Figure 27 must be completed, and verify that there is a shell of transverse radiation moving outwards with the speed of light.

[96] Reference: If you failed to do the exercise, take a look at R. Y. Tsien, *Pictures of dynamic electric fields*, Am. J. Phys. **40** (1972) 46. You will find this and many other illuminating illustrations there.

[97] Exercise: Apply local conservation of energy to a spherical electromagnetic wave. Show that the amplitudes of the fields in the wave must fall off like $1/r$, and hence that radiation fields do not follow the inverse square law.

The answer is no, because the physics is determined by the electric and magnetic fields, not by the vector potential directly. If we work out the electric field in the Coulomb gauge we find (in Gaussian units and in short hand notation)

$$E_i = F_{i0} = \partial_i A_0 - \partial_0 A_i = \partial_i A_0 + \partial_0 \frac{1}{\Box}(4\pi J_i + \partial_i \partial_0 A_0) =$$

$$= \frac{4\pi}{\Box}\partial_0 J_i + \partial_i \frac{1}{\Box}(\Box A_0 + \partial_0 \partial_0 A_0) = \tag{411}$$

$$= \frac{4\pi}{\Box}\partial_0 J_i + \partial_i \frac{1}{\Box}\nabla^2 A_0 = \frac{4\pi}{\Box}(\partial_0 J_i - \partial_i J_0) \ .$$

We made use of both of the Coulomb gauge equations, eq. (410) in the third step and (409) in the last. The result is exactly the same as the one we had in equation (385), which was derived without introducing the vector potential at all.[98] The non-locality present in Poisson's equation affects only gauge dependent quantities, not the gauge invariant electric field.

The Hertz dipole

It would be a pity not to include the radiation from an oscillating dipole here, even though Jackson has it in his chapter 9. We use the simple strategy of writing down an Ansatz for the field, and we check at the end that it is indeed the field we want. First we introduce a function

$$f = f\,(t - r/c) \tag{412}$$

where r is the distance from the origin. In this connection it is known as the *Hertz potential*. For the vector potential $\mathbf{A}$ we try

$$A_x = A_y = 0 \ , \qquad A_z = \frac{f'}{cr} \ . \tag{413}$$

By construction this solves the wave equation. In the Lorenz gauge the electric potential is then given by

[98] Exercise: To see that they are the same you have to decipher the notation. Do that.

$$\frac{1}{c}\partial_t \Phi = -\partial_z A_z = -\frac{\partial r}{\partial z}\partial_r A_z \ . \tag{414}$$

This is solved by

$$\Phi = \left(\frac{f'}{cr} + \frac{f}{r^2}\right)\frac{z}{r} \ . \tag{415}$$

Hence we have a solution of Maxwell's equations, with the Lorenz gauge built in. Switching to spherical polars, also for the vector field $\mathbf{A}$, we obtain

$$A_r = \frac{f'}{cr}\cos\theta \qquad A_\theta = -\frac{f'}{cr}\sin\theta \qquad \Phi = \left(\frac{f'}{cr} + \frac{f}{r^2}\right)\cos\theta \ , \tag{416}$$

all other components zero. From this we compute the electric and magnetic fields using the familiar formulæ

$$\mathbf{B} = \nabla \times \mathbf{A} \ , \qquad \mathbf{E} = -\frac{1}{c}\partial_t \mathbf{A} - \nabla\Phi \ . \tag{417}$$

The result is

$$E_r = \left(\frac{2f'}{cr^2} + \frac{2f}{r^3}\right)\cos\theta \qquad E_\theta = \frac{f''}{c^2 r}\sin\theta + \left(\frac{f'}{cr^2} + \frac{f}{r^3}\right)\sin\theta \tag{418}$$

$$B_\phi = \left(\frac{f''}{rc^2} + \frac{f'}{r^2 c}\right)\sin\theta \ , \tag{419}$$

all other components vanishing.

We can now investigate this field in the *near zone* and in the *far zone*. In the near zone the terms with the highest power of r in the denominator will dominate, so we have

$$E_r \approx \frac{2f}{r^3}\cos\theta \qquad E_\theta \approx \frac{f}{r^3}\sin\theta \qquad B_\phi \approx \frac{f'}{r^2 c}\sin\theta \ . \tag{420}$$

This is the field created by an electric dipole at the origin with dipole moment f along the z-axis, together with a magnetic field created by a current directed along the z-axis. If we choose the function f to be periodic we conclude that the source of our electromagnetic field is an oscillating dipole. To

be able to delineate the near zone with some precision, let us assume that the time dependence is sinusoidal, $f = A\sin(\omega t - kr)$. For the orders of magnitude we then find that

$$f' \sim \omega f = \frac{c}{\lambda} f \ , \qquad f'' \sim \omega^2 f = \frac{c^2}{\lambda^2} f \ . \tag{421}$$

By inspection we see that we are in the near zone if $r << \lambda$.

If $r >> \lambda$ we are in the far zone. Then the terms with the lowest power of r in the denominator dominate, and we have

$$E_\theta \approx \frac{f''}{rc^2}\sin\theta \qquad B_\phi \approx \frac{f''}{rc^2}\sin\theta \ , \tag{422}$$

all other components approximately vanishing. This is a transverse electromagnetic field whose amplitude is proportional to the acceleration of the oscillating dipole and inversely proportional to the distance from the source.[99]

Problem 6: Show that a possible choice of the vector potential is

$$A_\alpha = \int_0^1 s x^\beta F_{\beta\alpha}(sx)\ \mathrm{d}s \ . \tag{423}$$

This is known as the Poincaré gauge, for which $x^\alpha A_\alpha = 0$.

[99]Exercise: Check that this field does obey Maxwell's equations. Where is the source? Also calculate the Poynting vector in the far zone, and hence derive an expression for the energy flow there.

PLANE ELECTROMAGNETIC WAVES

We turn our attention to plane electromagnetic wave propagating in linear media, $\mathbf{D} = \epsilon\mathbf{E}$ and $\mathbf{B} = \mu\mathbf{H}$. Locally any electromagnetic wave looks plane, so this is an interesting case. When the wave meets a boundary where the dielectric constant changes a part of the wave is reflected while another part enters the other medium with a changed direction. The latter phenomenon is called *refraction*. We want a quantitative theory of reflection and refraction.

I will use Gaussian units throughout.

Monochromatic plane waves

Let us assume that ϵ and μ are constant as functions of space. It is still important to remember how they arise. In the medium ϵ (and perhaps also μ) differs from 1 due to interaction with molecules. The latter can be thought of as harmonic oscillators, with various resonance frequencies. This means that $\epsilon = \epsilon(\omega)$ will depend on the frequency ω of the wave. We therefore begin by isolating the part of the wave that has a definite frequency, known as a *monochromatic* wave.

We do this by means of a Fourier transformation in time, so that

$$\tilde{\mathbf{E}}(\mathbf{x},\omega) = \int_{-\infty}^{\infty} \mathbf{E}(\mathbf{x},t)e^{i\omega t}\,\mathrm{d}t \quad \Leftrightarrow \quad \mathbf{E}(\mathbf{x},t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \tilde{\mathbf{E}}(\mathbf{x},t)e^{-i\omega t}\,\mathrm{d}t \quad (424)$$

and similarly for $\mathbf{B}$. We assume that we are in a linear medium with ϵ and μ being constants, drop the tildes, and obtain Maxwell's equations for the Fourier transformed fields in the form

$$\begin{array}{ll} \nabla\cdot\mathbf{E} = 0 & \nabla\times\mathbf{E} - \frac{i}{c}\omega\mathbf{B} = 0 \\ \nabla\cdot\mathbf{B} = 0 & \nabla\times\mathbf{B} + \frac{i}{c}\omega\mu\epsilon\mathbf{B} = 0 \end{array} \quad . \quad (425)$$

From this we obtain Helmholtz' wave equation for $\mathbf{E}$,

$$\nabla^2\mathbf{E}(\mathbf{x},\omega) + \frac{1}{c^2}\omega^2\mu\epsilon\mathbf{E}(\mathbf{x},\omega) = 0\ , \quad (426)$$

and similarly for $\mathbf{B}$.

To describe the plane wave solutions we introduce three constant unit vectors $\underline{\epsilon}_1$, $\underline{\epsilon}_2$, and $\mathbf{n}$. Using $\mathbf{k} = k\mathbf{n}$ we find the solutions of the wave equation to be

$$\mathbf{E} = \underline{\epsilon}_1 E \cos\left(\mathbf{k}\cdot\mathbf{x} - \omega t + \text{phase}\right) \tag{427}$$

$$\mathbf{B} = \underline{\epsilon}_2 B \cos\left(\mathbf{k}\cdot\mathbf{x} - \omega t + \text{phase}'\right) , \tag{428}$$

where

$$k^2 = \frac{\omega^2 \mu\epsilon}{c^2} . \tag{429}$$

The full set of Maxwell's equations relate the electric and magnetic fields to each other. One finds that the phases that occur in the solutions are the same,

$$\text{phase}' = \text{phase} , \tag{430}$$

the amplitudes are related by

$$B = \sqrt{\mu\epsilon}E , \tag{431}$$

and the unit vectors that we have introduced must form an orthonormal set,

$$\mathbf{n}\cdot\underline{\epsilon}_1 = \mathbf{n}\cdot\underline{\epsilon}_2 = \underline{\epsilon}_1\cdot\underline{\epsilon}_2 = 0 . \tag{432}$$

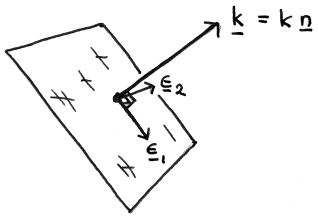

Figure 28: The three orthogonal vectors associated to a linearly polarized plane wave.

Thus the wave oscillates (along the $\underline{\epsilon}_i$) in a direction that is orthogonal to the direction of propagation (along $\mathbf{n}$), and the electric and magnetic fields are orthogonal to and in phase with each other.[100]

The *phase velocity* of the wave is

$$v_{\mathrm{p}} = \frac{c}{\sqrt{\mu\epsilon}} = \frac{c}{n} \ , \tag{433}$$

where we defined the *index of refraction*

$$n = \sqrt{\mu\epsilon} \ . \tag{434}$$

The velocity v_{p} is called the phase velocity for a reason. The point is that if we consider a wave propagating through a dielectric medium then the dielectric constant, and hence the index of refraction, will be a function of ω. When this happens the phase velocity depends on the frequency of the wave, and is not necessarily equal to what we would like to call the velocity of the wave packet that we obtain by superposing waves of different frequencies.

Polarisation

There is more structure to be noticed. We begin by writing down two linearly independent solutions for the electric field,

$$\begin{aligned} \mathbf{E}_1 &= E_1 \underline{\epsilon}_1 \cos\left(\mathbf{k}\cdot\mathbf{x} - \omega t\right) \\ \mathbf{E}_2 &= E_2 \underline{\epsilon}_2 \cos\left(\mathbf{k}\cdot\mathbf{x} - \omega t + \text{phase}\right) \ . \end{aligned} \tag{435}$$

The corresponding magnetic fields are easily written down when one wants them. These solutions are said to be *linearly polarized* because they oscillate in some constant direction. The general solution is a linear combination of the two. And unless the phase is set to zero, the direction of the electric field will rotate in the $(\underline{\epsilon}_1, \underline{\epsilon}_2)$–plane. This gives rise to *elliptic* polarizations, as well as two circularly polarized fields. See Figure 29.

[100] Exercise: Prove that Maxwell's equations do what I just claimed.

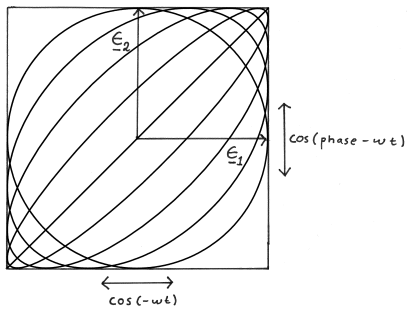

Figure 29: Linear, elliptical, and circular polarisation of an electromagnetic wave. For simplicity I have set $E_1 = E_2$ in the solution. In the picture the relative phase varies between 0 and $\pi/2$.

For serious calculations it is advisable to switch to a complex notation. We set

$$\mathbf{E} = \mathrm{Re}\left[\underline{\epsilon_1}\alpha e^{i(\mathbf{k}\cdot\mathbf{x}-\omega t)}\right] \tag{436}$$

where α is a complex number. Because the equations are linear in $\mathbf{E}$ and $\mathbf{B}$ we get, for instance,

$$\nabla \times (\mathrm{Re}[\mathbf{E}]) = \mathrm{Re}[\nabla \times \mathbf{E}] \ . \tag{437}$$

That is to say, we can do all the calculations using complex fields, and extract real parts at the end. But here we stick to real fields throughout.

Reflection and refraction

It is time to let a plane wave hit the (plane) boundary between two different media, say air and water. The light rays are directed orthogonally to the surfaces of constant phase. When they hit the boundary at an angle θ_i some of them will be reflected and some of them will cross the boundary at an angle θ_r called the angle of refraction. Snell's law for how θ_i and θ_r are related in terms of the index of refraction for the two different media is an immediate consequence of the wave description of light. See Figure 30.

But there is more to this. We want to derive the relative amplitudes of the incoming, reflected, and refracted waves. Moreover we want to investigate

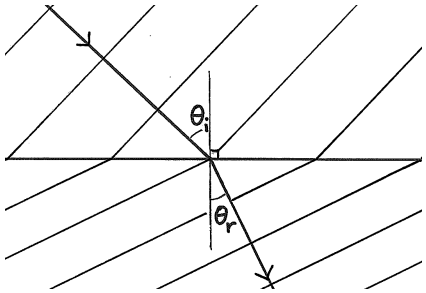

Figure 30: How to derive Snell's law. The wavelength depends on the index of refraction as $\lambda = 2\pi c/n\omega$.

the effects of the polarisation of the incoming light. Polarisation will be important because the calculation is an exercise in using what we know: that the normal components of **D** and **B**, and the tangential components of **E** and **H**, are continuous across the boundary. This gives two cases to consider, when **E** is tangential to the boundary and when **B** is tangential to the boundary. Afterwards a superposition can be made. For the details of the calculation, see Jackson's sections 7.1-4.

The results can be stated as follows: Let E_0, $E_0^p rime$, E_0'' be the amplitudes of respectively the incoming, refracted, and reflected field. When **E** is tangential to the boundary

$$\frac{E_0'}{E_0} = \frac{2n\cos\theta_i}{n\cos\theta_i + \frac{\mu}{\mu'}\sqrt{n'^{\,2} - n^2\sin^2\theta_i}}$$
$$\frac{E_0''}{E_0} = \frac{n\cos\theta_i - \frac{\mu}{\mu'}\sqrt{n'^{\,2} - n^2\sin^2\theta_i}}{n\cos\theta_i + \frac{\mu}{\mu'}\sqrt{n'^{\,2} - n^2\sin^2\theta_i}}\,. \tag{438}$$

When **B** is tangential to the boundary

$$\frac{E_0'}{E_0} = \frac{2nn'\cos\theta_i}{\frac{\mu}{\mu'}n'^{\,2}\cos\theta_i + n\sqrt{n'^{\,2} - n^2\sin^2\theta_i}} \tag{439}$$

$$\frac{E_0''}{E_0} = \frac{\frac{\mu}{\mu'} n'^{\,2} \cos\theta_i - n\sqrt{n'^{\,2} - n^2 \sin^2\theta_i}}{\frac{\mu}{\mu'} n'^2 \cos\theta_i + n\sqrt{n'^{\,2} - n^2 \sin^2\theta_i}} \ .$$

There is interesting physics in these formulas.[101] Let us just mention that the case $n > n'$ clearly deserves further thought. It has interesting applications when you try to transmit light through a pipe made of some suitable dielectric material. Again I have to refer to Jackson for the details.

A Gaussian beam

The plane waves are clearly of physical interest, and locally every wave looks like a plane wave. But it would be interesting to see a wave transporting a finite amount of energy, looking perhaps like the beam from a laser. This actually forces us to make the wave non-monochromatic, but we can stay with monochromatic waves if we make the *paraxial approximation*. Let us consider the scalar wave equation, and make the Ansatz

$$\Psi(t, x, y, z) = \psi(x, y, z) e^{i(kz - \omega t)} \ . \tag{440}$$

This should be a beam propagating close to the z-axis. We now make the approximation that

$$|\partial_z^2 \psi| << k |\partial_z \psi| \ . \tag{441}$$

Dropping the $\partial_z^2\psi$–terms from the wave equation for Ψ we find that it turns into

$$i\partial_z \psi = -\frac{1}{2k}(\partial_x^2 + \partial_y^2)\psi \ . \tag{442}$$

This is the paraxial approximation to the wave equation. It also happens to be the Schrödinger equation for a free particle in the plane. A solution is easily seen to be

$$\psi = \frac{1}{w_0^2 + 2iz/k} \exp\left(-\frac{x^2 + y^2}{w_0^2 + 2iz/2}\right) \ . \tag{443}$$

[101] For the second case, find an angle such that there is no reflection. What are the broader implications of this for the fabrication of sunglasses?

The parameter w_0 is called the waist of the beam, and it has to be significantly larger than the wavelength in order for the paraxial approximation to hold good.[102] And this is as far as we will go into laser physics, and as close as we will get to the interesting subject of diffraction (in Jackson's chapter 10).

Problem 7: Derive equations (439) in full detail, starting from Jackson's Figure 7.6b.

[102]Exercise: Check the statements made here, and plot the intensity of the wave for a some selected values of w_0.

ELECTRODYNAMICS IN MEDIA

We are now in position to ask what happens inside a medium when we no longer assume that everything is static. In particular we can consider electromagnetic waves propagating in media.

A model for the medium

In the spirit of our treatment of electrostatics in media we try to get by with a very simple model of how the electrons inside the medium respond to an imposed electric field. What we do want the model to catch is the fact that the response depends on the frequency of the time-varying electric field. So we will consider a single Fourier mode $\mathbf{E}(\omega, \mathbf{x})$. We assume that the electrons are bound to their positions with harmonic forces, which means that they respond as damped harmonic oscillators. They oscillate around their equilibrium positions according to

$$m(\ddot{\mathbf{x}} + \gamma_0 \dot{\mathbf{x}} + \omega_0^2 \mathbf{x}) = -e\mathbf{E}(\mathbf{x}, t) \ . \tag{444}$$

For simplicity we assume that the amplitude is small enough so that we can regard the electric field as being constant in space. For later purposes it is also important to record that it is reasonable to assume that $\gamma_0 << \omega_0$.

If we Fourier transform over time and consider a single mode we can solve the harmonic oscillator equation to obtain the resulting dipole moment

$$\mathbf{p}(\omega) = -e\mathbf{x}(\omega) = \frac{e^2}{m} \frac{\mathbf{E}(\omega)}{\omega_0^2 - \omega^2 - i\omega\gamma_0} \ . \tag{445}$$

In the static limit, $\omega = 0$, this reproduces eq. (219). We will assume that the medium is sufficiently dilute so that we can ignore the distinction between the macroscopic field $\mathbf{E}$ and the field $\mathbf{E}_{\text{mol}}$ that actually acts on the molecules. Our focus will be on the fact that the response of the medium depends very much on how close the frequency of the wave is to the eigenfrequency ω_0 of the oscillator.

Now there will be Z electrons per molecule, each contributing to the dipole moment according to the above. We take it that f_i of those electrons

can be described as oscillators of eigenfrequency ω_i. Evidently

$$\sum_i f_i = Z \ . \tag{446}$$

Finally there will be N molecules per unit volume. Following the discussion in the static case we calculate the total contribution to the dielectric constant, and end up with

$$\epsilon(\omega) = \epsilon_0 + \frac{Ne^2}{m} \sum_i \frac{f_i}{\omega_i^2 - \omega^2 - i\omega\gamma_i} \ . \tag{447}$$

Because we assumed $\gamma_i << \omega_i$ this is close to being real for most frequencies, but when the applied frequency ω is close to a resonance frequency ω_i damping does become important. When we come to wave propagation we will be especially interested in how the real part varies with ω. Regions where its derivative is negative are called regions of *anomalous dispersion*, for a reason that will become clear when we discuss the group velocity of the wave.[103]

Low and high frequency behaviour

What happens in the limit when the frequency of the applied field goes to zero depends on whether there is a resonant frequency with $\omega_0 = 0$, or not. If there is, then there are free electrons present and the medium is a conductor. If there is not, then the medium is an insulator and our formula for $\epsilon(\omega)$ becomes a formula for the molecular polarizability that appears in the Clausius–Mosotti equation,

$$\gamma_{\rm mol} = \frac{e^2}{\epsilon_0 m} \sum_i \frac{f_i}{\omega_i^2} \ . \tag{448}$$

To get a feeling for its magnitude we assume that the frequencies appearing here are of the order of the frequencies of light emitted by atoms, say 10^{16} $\rm s^{-1}$. If so the electronic contribution to $\gamma_{\rm mol}$ should be about 10^{-29} $\rm m^3$, or about one Ångström cubed, which is reasonable.[104]

[103] Exercise: Suppose that there is only one term in the sum. When exactly is the derivative of $\mathrm{Re}[\epsilon(\omega)]$ equal to zero? When is it negative?

[104] Exercise: Check this, and compare to Problem 5.

If there are free electrons present the term with eigenfrequency zero is singled out for special attention. We get

$$\epsilon(\omega) = \epsilon_0 + \frac{Ne^2}{m}\frac{if_0}{\omega(\gamma_0 - i\omega)} + \frac{Ne^2}{m}\sum_{i>0}\frac{f_i}{\omega_i^2 - \omega^2 - i\omega\gamma_i} = \\ = i\frac{Ne^2 f_0}{m\omega(\gamma_0 - i\omega)} + \epsilon_b \ , \tag{449}$$

where ϵ_b is defined to include all the contributions from the bound electrons. Let us focus on the first term, which is singular in the static limit. To understand its significance we look at the Ampère-Maxwell equation, which is the one where the electric current makes its appearance. Inside a conductor we can use Ohm's law $\mathbf{J} = \sigma\mathbf{E}$, where σ is the *conductivity*. If the displacement field $\mathbf{D}$ varies like $e^{-i\omega t}$ we obtain

$$\nabla \times \mathbf{H} = \mathbf{J} + \partial_t \mathbf{D} = -i\omega\left(\epsilon + i\frac{\sigma}{\omega}\right)\mathbf{E} \ . \tag{450}$$

The fit with equation (449) is perfect provided we regard the dielectric constant ϵ as coming from the bound electrons. Once we have set $\epsilon = \epsilon_b$ the conductivity must be

$$\sigma = \frac{Ne^2 f_0}{m(\gamma_0 - i\omega)} \ . \tag{451}$$

This is the *Drude model* for the electrical conductivity. To do better one has to bring quantum mechanics in.

At the other end of the spectrum, when the applied frequency is clearly higher than all the vibration frequencies in the medium, we obtain

$$\epsilon(\omega) = \epsilon_0\left(1 - \frac{\omega_p^2}{\omega^2}\right) \ , \tag{452}$$

where

$$\omega_p^2 = \frac{NZe^2}{\epsilon_0 m} \ . \tag{453}$$

The frequency ω_p depends on the total number NZ of electrons per volume, and is known as the *plasma frequency* of the medium. In a plasma all the electrons are free and we are always in the high frequency limit as defined here, whereas we have to have $\omega >> \omega_p$ if a dielectric medium is to be described by this limit.

Our all too brief account of the mechanism behind the interaction of light and matter is at its end.[105]

Dispersion and group velocity

We now return to equation (433), which gives the phase velocity of a single frequency or *monochromatic* plane wave propagating through a medium. For many kinds of media we can assume that $\mu \approx \mu_0$, but as we have just seen we expect ϵ to be a non-trivial function of ω. To begin with we assume that it has no imaginary part, that is to say that the damping can be neglected. According to our assumptions about the medium this will be true except close to the resonance frequencies where anomalous dispersion occurs. But we still have a non-trivial relation between the frequency ω and the *wave number* k to deal with, namely

$$\omega = \frac{ck}{\sqrt{\mu\epsilon(\omega)}} = \frac{ck}{n(\omega)} \ . \tag{454}$$

Given that the index of refraction n depends on ω there is a relation between the frequency and the wave number, which we write in the form

$$\omega = \omega(k) \ . \tag{455}$$

As you can deduce from Jackson's Figures 7.8 and 7.9 this can be a highly non-trivial function. (Solving for $k = k(\omega)$ would be less useful for now.)

We are now in a position to superpose solutions with different frequencies, and form wave packets

$$\Psi(x,t) = \int_{-\infty}^{\infty} A(k) e^{i(kx-\omega(k)t)} \ \mathrm{d}k \ . \tag{456}$$

[105]Reference: A nice survey is V. Weisskopf, *How light interacts with matter*, Scientific American 1968 no 3 p. 61.

At a given time this is a wave packet having some shape as a function of x. But as the wave propagates the speed of propagation depends on k, which means that some Fourier components move faster than others, and so the shape of the wave changes with time. However, as long as the wave is a recognizable peculiarity travelling through the medium, we want to know the speed it is travelling at, and the speed with which energy is transported.

So we take a close look at the wave packet (456). Its initial shape can be anything, and the Fourier components are moving with different speeds. Let us however look at the interesting case when the function $A(k)$ is strongly peaked around some special wave number $k = k_0$. Then we can expand

$$\omega(k) = \omega(k_0) + \frac{d\omega}{dk}_{|k_0} (k - k_0) + \dots \ , \tag{457}$$

and we can ignore the higher order terms. Let us denote

$$v_{\rm g} = \frac{d\omega}{dk}_{|k_0} \ . \tag{458}$$

We can now approximate

$$\begin{aligned} \Psi(x,t) \approx \int_{-\infty}^{\infty} A(k) e^{i(kx - \omega_0 t - v_{\rm g}(k-k_0)t)} \ {\rm d}k = \\ = e^{i(k_0 x - \omega_0 t)} \int_{-\infty}^{\infty} A(k) e^{ik(x - v_{\rm g} t)} \ {\rm d}k = e^{i(k_0 x - \omega_0 t)} \Psi(x - v_{\rm g} t, 0) \ . \end{aligned} \tag{459}$$

We see that the initial shape of the pulse is moving with the velocity $v_{\rm g}$, which is known as the *group velocity*.[106]

In the first successful experiment to determine the speed of light Fizeau used a rotating toothed wheel. Light pulses passing through an opening between the teeth would be reflected by a mirror 8.6 km away. Depending on the speed of light and the angular velocity of the wheel the light would then either pass through or be blocked on its way back. In this way the velocity of light in air was found to be close to $3.15 \cdot 10^8$ m/s, which was an improvement on the value obtained by astronomers. The modern value for

[106] Exercise: Superpose the two waves $a\cos(\omega_1 t - k_1 x)$ and $a\cos(\omega_2 t - k_2 x)$ and suppose that the difference $\omega_1 - \omega_2$ is small. Can you recognize a version of the group velocity? If the energy density is proportional to the amplitude squared, how fast does it travel?

the velocity of light in vacuum is $c = 299792458$ m/s (exactly, by definition of the metre). The point here is that Fizeau actually determined the group velocity of light in air.

A localized wave packet that is slowly deformed as it travels is only a special case of a *dispersive wave*, that is to say a wave where the crests move with a speed that depends on the frequency. An example worth watching is provided by surface waves propagating on water that is deep enough compared to the wavelength so that the waves are unaffected by the bottom of the lake. You can create such waves by throwing a pebble into the lake. The initial disturbance is complicated, but as it moves outwards the low frequency components travel the fastest. What you will then see is a system of rings in which new waves appear at the rear, travel through the group, and disappear at the front. The waves are roughly sinusoidal but the wavelength grows in the outwards direction. In this situation the group velocity determines the speed with which regions of constant wave number spread, as well as the speed with which such waves propagate energy.[107] Since this is not equal to the speed with which waves of constant wave number travel you have to watch the phenomenon closely in order to see what is going on.[108]

Dispersion of electromagnetic waves

We now focus on plane electromagnetic waves in a medium with index of refraction equal to $n = n(\omega)$. First we will compare the phase velocity to the group velocity. Recall that the phase velocity is

$$v_{\mathrm{p}} = \frac{\omega(k)}{k} = \frac{c}{n(\omega)} \, . \tag{460}$$

We expect the index of refraction to be larger than one, so the phase velocity should be smaller than c. Phase velocities larger than c would not necessarily cause any raised eyebrows, because we cannot send signals using

[107] Exercise: These waves obey the equation $\omega^2 = gk$ where $g = 9.82$ m/s^2. Compare the group velocity to the phase velocity.

[108] Reference: For the theory you can consult J. Lighthill: *Waves in Fluids*, Cambridge UP 1978.

an infinitely long wave train with a fixed frequency.[109] The group velocity is more interesting. We find

$$v_{\rm g} = \frac{1}{\frac{dk}{d\omega}|k_0} = \frac{1}{\frac{d}{d\omega}\left(\frac{n\omega}{c}\right)_{|k_0}} = \frac{c}{n + \omega\frac{dn}{d\omega}} \ . \tag{461}$$

Typically this is smaller than the phase velocity. However, we recall that when the frequency of the wave is close to a resonance frequency in the medium then we are in a region of 'anomalous dispersion' where $dn/d\omega$ is negative. (See Jackson's Figures 7.8 and 7.9.) When this happens the group velocity is larger than the phase velocity. In fact, it may even be larger than c, and this does cause raised eyebrows. It seems appropriate to end these notes with an attempt to travel faster than light, so let us look into this a little.

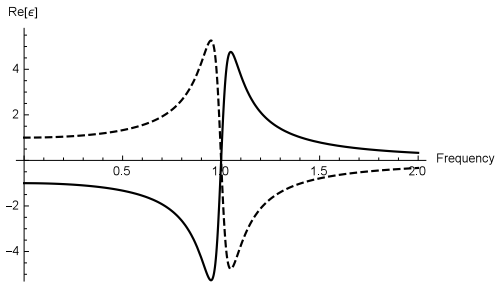

Figure 31: In an ordinary material the real part of the dielectric constant behaves according to the dashed curve. If the population of the energy levels has been inverted it will behave according to the black curve.

First of all the modern physicist is not confined to use materials that are easily found in nature. It is possible to create materials that behave quite differently from the illustrations shown by Jackson. One way to achieve this is by means of *optical pumping*, which requires us to think in quantum mechanical terms. Each atom in the medium has a discrete set of energy levels, and under normal circumstances one expects the number of filled levels to be proportional to the Boltzmann factor $e^{-E_n/kT}$. For moderate temperatures the ground state is the most populated one. Optical pumping

[109] Exercise: Post a large number of people equipped with semaphors along a straight line. How can they be used to create a wave travelling faster than light?

is a way to invert the population so that most of the atoms are in some excited state. A passing light wave will cause them to revert to the ground state through stimulated emission. Then our idea of damping is stood on its head, since the medium actually delivers energy to the wave. A useful model of what is going on is obtained by sticking to equation (447), but letting f be negative. The result (for one oscillator) is illustrated in Figure 31. Where the derivative is strongly positive we can create light pulses whose group velocity is very low. When we are well below the resonance frequency we find ourselves in a transparent band where the group velocity is larger than the phase velocity, and the phase velocity is larger than c because $n < 1$.

Playing tricks like that, it is quite possible to obtain well defined wave packets that in a sense move faster through the medium than they would move in vacuum. Figure 32 is supposed to give the story away.[110] There have been some quite dramatic claims in this direction. The winner is the claim by Heitmann and Nimtz, stating that Mozart's 40th symphony was sent through a tunnel of length 114 mm at a speed of 4.7 c.[111]

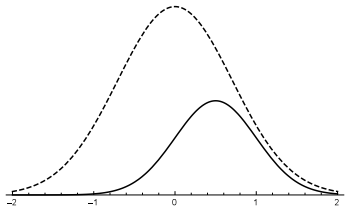

Figure 32: The dashed profile has propagated through vacuum. The smaller profile has propagated through the medium, and on the face of it it looks as if it has travelled faster than c. If the group velocity is negative it may even be said to exit the medium (a thin slab) before it has entered it.

These claims fade a little under scrutiny. The time advance achieved for Mozart's symphony was less than a nanosecond, far too short to be noticeable. Worse, the signal transmitted was for all practical purposes analytic. Predicting it a nanosecond in advance from a low order Taylor series expansion poses no problem at all. The pride of wave physics is that it admits

[110] Exercise: A locomotive moves at constant speed 100 km/s from Stockholm to Alvesta, pulling six wagons when it starts off. How can you make the velocity of the centre of mass of the train exceed 100 km/s?

[111] Reference: W. Heitmann and G. Nimtz, Phys. Lett. **A196** (1994) 154.

propagating discontinuities and the question at issue—if you want to challenge the supremacy of the magical velocity c—is whether such discontinuities can propagate faster than c. The answer is no, as you can see in Jackson's section 7.11. Hence the consensus is that relativity is not under threat.[112]

I should have had more to say, but I end here.[113]

Problem 8: In equation (456) set $A(k) = Ae^{-\sigma(k-k_0)^2}$. Expand $\omega(k)$ to first order in k and calculate $\Psi(x,t)$. Then expand to second order in k and see what happens.

[112]Reference: For a review of these things, see R. Y. Chiao and A. M. Steinberg, *Tunneling times and superluminality*, Progress in Optics **37** (1997) 347.

[113]There is a beautiful book containing much more: P. W. Milonni: *Fast Light, Slow Light, and Left–Handed Light*, IOP Publishing, Bristol 2005.