603: Electromagnetic Theory I

CONTENTS

- Maxwell's Equations: Introduction; units; boundary conditions.
- Electrostatics: Uniqueness theorem; Green's theorem; gauge potentials; energy
- Boundary value problems in electrostatics: Method of images; separation of variables in Cartesian, spherical polar and cylindrical polar coordinates.
- Dielectric media
- Multipole expansion
- Magnetostatics
- Time-dependent fields

Contents

1 Introduction

1.1 Maxwell's equations

The equations now known as Maxwell's equations were obtained over an extended period, principally during the early nineteenth century. Here, we shall take as our starting point the set of four differential equations as they were presented by Maxwell in about 1861. It was Maxwell who completed the process of constructing the equations, thereby achieving the first *unification* of fundamental theories in physics. Prior to Maxwell, there were two essentially independent theories, one describing electricity and the other describing magnetism, and it was he who brought about the synthesis that unified them into a single theory of *electromagnetism*. It was only later, after Einstein developed the theory of *Special Relativity* in 1905, that the magnitude of Maxwell's achievement really became clear. Especially, a quite remarkable feature of Maxwell's 1861 equations is that they are already completely compatible with special relativity, with no need for modification of any kind.[1] Aside from changes in notation and units, Maxwell's equations have remained otherwise unaltered since 1861.

Let us begin by considering Maxwell's equations in free space, by which is meant that the space outside of any conducting surfaces is assumed to be a vacuum. Using the SI system of units, Maxwell's equations are:

$$\begin{aligned} \vec{\nabla}\cdot\vec{E}' &= \frac{\rho'}{\epsilon_0}\,, \qquad \vec{\nabla}\times\vec{B}' - \mu_0\epsilon_0\,\frac{\partial\vec{E}'}{\partial t} = \mu_0\vec{J}'\,, \\ \vec{\nabla}\cdot\vec{B}' &= 0\,, \qquad \vec{\nabla}\times\vec{E}' + \frac{\partial\vec{B}'}{\partial t} = 0\,. \end{aligned} \tag{1.1}$$

Observe that I have written these equations with a "prime" on the electric field $\vec{E}$ the magnetic field $\vec{B}$, the electric charge density ρ and the electric current density $\vec{J}$. This is to indicate that these quantities are all expressed in the SI system of units. (The SI system typically maximises the number of "redundant" dimensionful constants, and so one might say that it is **S**uper **I**nconvenient.) The remaining quantities appearing in (1.1) are the constants ϵ_0 and μ_0, which are, respectively, the permitivity of free space and the permeability of free space. They have the values

$$\epsilon_0 \approx 8.85419\times 10^{-12} \quad \text{Farads/metre}\,, \qquad \mu_0 = 4\pi\times 10^{-7} \quad \text{Henries/metre} \tag{1.2}$$

[1] This contrasts with the case of Newtonian mechanics, which is not compatible with special relativity and therefore did not survive as a "fundamental" theory after 1905.

1.2 Gaussian units

SI units have their virtues for some purposes, but they can also be quite inconvenient in practice. This seems to be especially true in electromagnetism, and for this reason it is often more convenient to stick with an earlier system, known as *Gaussian* units. In this system, Maxwell's equations in free space take the form

$$\begin{aligned} \vec{\nabla}\cdot\vec{E} &= 4\pi\rho\,, \qquad \vec{\nabla}\times\vec{B} - \frac{1}{c}\,\frac{\partial\vec{E}}{\partial t} = \frac{4\pi}{c}\,\vec{J}\,, \\ \vec{\nabla}\cdot\vec{B} &= 0\,, \qquad \vec{\nabla}\times\vec{E} + \frac{1}{c}\,\frac{\partial\vec{B}}{\partial t} = 0\,, \end{aligned} \tag{1.3}$$

where c is the speed of light. Observe that here, we are writing the equations using the unprimed quantities $\vec{E}$, $\vec{B}$, ρ and $\vec{J}$, and it will probably therefore come as no surprise that it is this Gaussian system of units that I prefer to use. It should already be evident upon comparing (1.1) and (1.3) that Gaussian system is somewhat simpler, in that one needs only one "fundamental constant" (i.e. the speed of light) rather than two (the permittivity and the permeability of free space).[2] The introduction of the 4π factors in (1.3) may perhaps seem tiresome, but the advantage of doing so will become apparent in due course. (There can be no escape from having 4π factors somewhere, because of the fact that a unit sphere has area 4π.)

In order to ensure that we can, whenever desired, revert to SI units, it is useful to work out explicitly the relation between the Gaussian quantities (denoted without primes, as in (1.3)) and the SI quantities (denoted with primes, as in (1.1)). In order to do this, we first need to understand a very important property of the Maxwell equations, namely that they imply the existence of *electromagnetic waves* that propagate at the speed of light.

Consider the Maxwell equations (1.1) in a completely empty region of space, where there is no charge density ρ' and no current density $\vec{J}'$. Taking the curl of the last equation, and using the vector identity (we shall assume here we are using Cartesian coordinates) $\vec{\nabla}\times(\vec{\nabla}\times\vec{V}) = \vec{\nabla}(\vec{\nabla}\cdot\vec{V}) - \nabla^2\vec{V}$, we obtain

$$-\nabla^2\vec{E}' + \vec{\nabla}(\vec{\nabla}\cdot\vec{E}') + \vec{\nabla}\times\frac{\partial\vec{B}'}{\partial t} = 0\,, \tag{1.4}$$

and hence, since the partial derivative $\partial/\partial t$ commutes with the spatial partial derivatives in $\vec{\nabla}$,

$$-\nabla^2\vec{E}' + \vec{\nabla}(\vec{\nabla}\cdot\vec{E}') + \frac{\partial}{\partial t}\,(\vec{\nabla}\times\vec{B}') = 0\,. \tag{1.5}$$

[2] Actually, one can do even better by changing the units in which one measures length from the *metre* to the *light second*, or alternatively, changing the unit of time to the *light metre* (the time light takes to travel 1 metre). In either of these systems of units, the speed of light becomes equal to 1.

Using the first equation in (1.1) (with $\rho' = 0$) and the second equation (with $\vec{J}' = 0$) then gives

$$\nabla^2 \vec{E}' - \mu_0 \epsilon_0 \frac{\partial^2}{\partial t^2} \vec{E}' = 0 \,. \tag{1.6}$$

Analogous manipulations show that $\vec{B}'$ satisfies an identical equation. We see, therefore, that the electric and magnetic fields satisfy an equation for waves that propagate at the speed

$$c = \frac{1}{\sqrt{\mu_0 \epsilon_0}} \approx 2.99792 \times 10^8 \quad \text{metres/second} \,. \tag{1.7}$$

This is precisely the speed of light *in vacuo*, and these wave solutions describe the propagation of radio waves, light, etc.

With this preliminary, we are nearly ready to establish the relation between the SI units used in (1.1), and the Gaussian units used in (1.3). The procedure for doing this is to introduce constant factors α, β, γ and δ that relate the primed to the unprimed quantities,

$$\vec{E}' = \alpha \vec{E} \,, \quad \vec{B}' = \beta \vec{B} \,, \quad \rho' = \gamma \rho \,, \quad \vec{J}' = \delta \vec{J} \,, \tag{1.8}$$

and to fix the values of these constants by demanding that plugging (1.8) into (1.1) should give (1.3). It is essential, in doing so, that we have the relation (1.7) between c, μ_0 and ϵ_0. Elementary algebra then gives

$$\alpha = \frac{\gamma}{4\pi\epsilon_0} \,, \qquad \beta = \frac{\gamma}{4\pi} \sqrt{\frac{\mu_0}{\epsilon_0}} \,, \qquad \delta = \gamma \,. \tag{1.9}$$

Notice that the fact that the constants γ and δ, the rescalings of ρ and $\vec{J}$, turned out to be equal is a not a coincidence. These rescalings reflect the fact that the unit of *charge* is being rescaled in the transformation from SI to Gaussian units. Now the dimensions of charge density and current density are

$$[\rho] = (\text{Charge})\, L^{-3} \,, \qquad [\vec{J}] = (\text{Charge})\, L^{-2}\, T^{-1} \,, \tag{1.10}$$

so they must both scale the same way under a rescaling of the charge.

Observe that the value of the constant γ has not yet been determined. This means that we can choose any value for γ, and we may use this freedom in order to make some other equation as nice as possible by removing superfluous constant factors. Consider Coulomb's law, giving the force between two electric charges separated by a distance R. We again need to distinguish between the charges q_1' and q_2' expressed in SI units, and the charges q_1 and q_2 expressed in Gaussian units. Since we have the relation $\rho' = \gamma\rho$ between charge *densities*

in the two systems, and since the unit of *volume* is the same in the two systems, it follows that the charges will also be related by the same factor of γ:

$$q' = \gamma q\,. \tag{1.11}$$

Now, in the SI system the force between the two charges is given by

$$F = \frac{q'_1 q'_2}{4\pi\epsilon_0 R^2} = \frac{\gamma^2 q_1 q_2}{4\pi\epsilon_0 R^2}\,. \tag{1.12}$$

Clearly, since we are free to choose γ to be whatever we like, the most convenient choice is to take

$$\gamma = \sqrt{4\pi\epsilon_0}\,, \tag{1.13}$$

so that the force between charges q_1 and q_2 is simply

$$F = \frac{q_1 q_2}{R^2}\,. \tag{1.14}$$

This is precisely the choice made in the Gaussian system of units.

Going back to (1.9), and combining it with the additional relation (1.13), we see that the four constants α, β, γ and δ are now uniquely determined in terms of μ_0 and ϵ_0. Thus we arrive at the following "dictionary" for relating the SI (primed) quantities to the Gaussian (unprimed) quantities:

$$\begin{aligned} \vec{E}' &= \frac{1}{\sqrt{4\pi\epsilon_0}}\,\vec{E}\,, \qquad \vec{B}' = \sqrt{\frac{\mu_0}{4\pi}}\,\vec{B}\,, \\ \rho' &= \sqrt{4\pi\epsilon_0}\,\rho\,, \qquad \vec{J}' = \sqrt{4\pi\epsilon_0}\,\vec{J}\,, \qquad q' = \sqrt{4\pi\epsilon_0}\,q\,. \end{aligned} \tag{1.15}$$

With these relations established, we can happily proceed by using the more convenient Gaussian units in this course, and anyone who wishes to re-express things in SI units can do so using the *SI-Gauss Dictionary* (1.15).

1.3 Macroscopic media

In principle, every problem in classical electromagnetism can be viewed as a problem formulated in free space, together with a number of electric point charges carried by electrons, protons, etc. In practice, however, it is often the case that the number of individual point charges is so large that it would not be convenient to consider them all separately, and instead, it is preferable to make a "macroscopic approximation." One obvious example is the notion of a conductor: It would be very clumsy and unwieldy to treat every electrostatics problem involving a conducting surface as a problem involving 10^{23} or so positive

and negative point charges that are bound together in such a way as to make what we conventionally think of as a sheet of metal. Instead, we can typically just forget about the microscopic explanation of why the protons, neutrons and electrons have formed themselves into a metal, and instead simply abstract from this the macroscopic notion of a surface on which the electrostatic potential is constant.

Another example where a macroscopic viewpoint is very useful is when one considers materials (such as glass) that exhibit a dielectric permittivity, or else materials that exhibit a magnetic permeability. One certainly can give a microscopic understanding of why these materials behave as they do, but it is convenient not to have to delve into these details every time we want to work out the effect of a slab of glass in an electrostatics problem.

In order to give a macroscopic formulation of Maxwell's theory in the presence of media, we now interpret $\vec{E}$ and $\vec{B}$ as *averaged* values of the electric and magnetic fields, where the averaging is performed over the distance scale of order the interatomic spacing in the medium. The point here is that we don't want to get involved in looking at the (enormous) microscopic variations in the fields that occur on the atomic length scale as one moves around close to individual electrons and protons. Having performed this averaging, the meanings of $\vec{E}$ and $\vec{B}$ are the same as they are in free space. For example, $\vec{E}$ still measures the potential difference between neighbouring points divided by their spatial separation.

We must also introduce two new quantities, called $\vec{D}$ and $\vec{H}$, which are related to $\vec{E}$ and $\vec{B}$ respectively. The standard names for all four fields are:

$\vec{E}$:	Electric field
$\vec{D}$:	Electric displacement
$\vec{B}$:	Magnetic induction
$\vec{H}$:	Magnetic field

In free space, we have

$$\vec{D} = \vec{E}\,, \qquad \vec{H} = \vec{B}\,. \tag{1.16}$$

In a medium, on the other hand, $\vec{D}$ represents a "back-reacted" version of $\vec{E}$, which takes into account the fact that the positive and negative charges in the medium are displaced because of the presence of the externally-applied $\vec{E}$ field, and thus they feed back into the system. To leading order, the system of positive and negative charges in the medium (which is neutral on balance) distorts so that each atom or molecule acquires a small electric dipole moment, leading to a dipole moment density, or *polarisation*, $\vec{P}$, and

$$\vec{D} = \vec{E} + 4\pi\vec{P}\,. \tag{1.17}$$

In a similar way, if the medium has magnetic properties there will be a similar relation

$$\vec{H} = \vec{B} - 4\pi\vec{M}\,, \tag{1.18}$$

where $\vec{M}$ is a magnetic dipole density, or *magnetisation*, term.

The effect of all this is that the Maxwell equations are modified in the presence of the medium. Instead of the free-space equations (1.3) we shall now have

$$\begin{aligned} \vec{\nabla}\cdot\vec{D} &= 4\pi\rho\,, \qquad \vec{\nabla}\times\vec{H} - \frac{1}{c}\,\frac{\partial\vec{D}}{\partial t} = \frac{4\pi}{c}\,\vec{J}\,, \\ \vec{\nabla}\cdot\vec{B} &= 0\,, \qquad \vec{\nabla}\times\vec{E} + \frac{1}{c}\,\frac{\partial\vec{B}}{\partial t} = 0\,, \end{aligned} \tag{1.19}$$

Notice that it is the first two equations, the ones that have the ρ and $\vec{J}$ source-terms on the right-hand side, that are modified. The remaining two equations are completely unchanged from their free-space forms.

A common situation is when the medium is completely uniform and isotropic (meaning that it is the same everywhere, and the same in all directions), and for which $\vec{D}$ and $\vec{H}$ are simply constant multiples of $\vec{E}$ and $\vec{B}$ respectively:

$$\vec{D} = \epsilon\,\vec{E}\,, \qquad \vec{B} = \mu\,\vec{H}\,. \tag{1.20}$$

The constant ϵ is called the *relative permittivity* of the medium, and the constant μ is called the *relative permeability* of the medium. In free space, where (1.16) holds, we clearly have

$$\epsilon = 1\,, \qquad \mu = 1\,. \tag{1.21}$$

In this course, when we consider electromagnetism in a medium, we shall commonly assume the relations (1.20).[3] Note that the permittivity ϵ of the medium is also often called the *dielectric constant*.

1.4 Boundary conditions at media interfaces

A situation that one encounters frequently when studying physical problems in electromagnetism is where there is a boundary or interface between two different materials or media. The simplest such situation in electrostatics is the case where there is a conducting surface

[3]Note that the adjective "relative" preceding the permittivity or permeability is really superfluous when using the Gaussian system of units, since the permittivity and permeability of free space (i.e. vacuum) are both equal to unity. In the SI system the adjective "relative" is used in order to signify that it means the factor by which the permittivity or permeability is larger (or smaller) than that in the vacuum. We shall drop the prefix "relative" from now on.

in otherwise free space. Another example would be an interface between two materials with different dielectric constants. In fact the conductor in free space can just be viewed as a special case of the interface between two dielectric materials, with one of them (free space) having $\epsilon = 1$ and the other (the conductor) having $\epsilon = \infty$.

The boundary conditions on the electric and magnetic fields at an interface between two media can be determined by performing appropriate integrals of the Maxwell equations (1.19). Let us label the media by "1" and "2," and likewise place "1" and "2" subscripts on the various electric and magnetic fields on the two sides of the interface.

Beginning with $\vec{D}$, we can integrate $\vec{\nabla} \cdot \vec{D} = 4\pi\rho$ over a so-called "Gaussian pillbox" that straddles the interface. The pillbox is a like a very short length of circular cylinder, with the ends capped off so as to form a closed volume. One should imagine that the size of the whole pillbox is very small, and in fact eventually one takes the limit where the size tends to zero. At all stages in the limiting process, the height of the box (i.e. the length of the cylinder) is very small compared with its radius. The caps of the cylinder are taken to be parallel to the interface, with the interface slicing through the box; one cap on each side.

The *divergence theorem* states that for any vector field $\vec{v}$ we have

$$\int_V \vec{\nabla} \cdot \vec{v}\, dV = \int_S \vec{v} \cdot d\vec{S}\,, \tag{1.22}$$

where S is a closed surface enclosing a volume V. Integrating $\vec{\nabla} \cdot \vec{D} = 4\pi\rho$ over the pillbox and using (1.22), we therefore find

$$\int_V \vec{\nabla} \cdot \vec{D}\, dV = \int_S \vec{D} \cdot d\vec{S} = 4\pi \int_V \rho dV = 4\pi q\,, \tag{1.23}$$

where q is the charge inside the pillbox. Because the height of the pillbox is taken to be very small compared to its diameter, we can neglect the contributions to the $\vec{D}$ integral coming from the sides. Since the pillbox itself will eventually be taken to have infinitesimal size we can think of the interface where the pillbox is placed as being planar. Let $\vec{n}$ be the unit normal vector pointing from medium 1 into medium 2. If the cross-sectional area of the pillbox is ΔA, then (1.23) gives

$$\vec{n} \cdot (\vec{D}_2 - \vec{D}_1)\, \Delta A = 4\pi\sigma\, \Delta A\,, \tag{1.24}$$

where σ is the surface charge density[4] at the interface. Thus we have

$$\vec{n} \cdot (\vec{D}_2 - \vec{D}_1) = 4\pi\sigma \tag{1.25}$$

[4]The surface charge resides entirely in the surface of the interface; it is a charge per unit area. In the idealisation where the interface is a sharp boundary between the two media, it is an infinitesimally thin sheet of charge.

at the interface.

By the same same token, the integration of the Maxwell equation $\vec{\nabla} \cdot \vec{B} = 0$ over the same pillbox gives

$$\vec{n} \cdot (\vec{B}_2 - \vec{B}_1) = 0 \tag{1.26}$$

at the interface. The zero on the right-hand side reflects the fact that there are no magnetic charges.

Further boundary conditions follow by appropriately integrating the remaining two Maxwell equations across the interface. This time, we consider a rectangular loop formed by two parallel line elements, one on each side of the interface, joined into a loop by adding connecting lines at the two ends. In the discussion that follows we first take the lengths of the two end connecting lines to zero, meaning that the area of the loop will go to zero. Eventually, we also take the lengths of the long sides of the loop to zero also. We now make use of Stokes' theorem, which states that for any vector field $\vec{v}$ we have

$$\int_\Sigma (\vec{\nabla} \times \vec{v}\,) \cdot d\vec{S} = \oint_C \vec{v} \cdot d\vec{\ell}, \tag{1.27}$$

where Σ denotes an (open) surface whose boundary is the closed loop C.

Suppose again the unit normal from medium 1 to medium 2 is $\vec{n}$ at the chosen point on the interface that we are considering. Let $\vec{m}$ be a unit vector that is perpendicular to the recangular loop, and therefore it is tangent to the interface at the selected point. Integrating the Maxwell equation $\vec{\nabla} \times \vec{E} = -(1/c)\partial\vec{B}/\partial t$ over the area Σ of the loop and applying (1.27) gives

$$\int_\Sigma (\vec{\nabla} \times \vec{E}\,) \cdot d\vec{S} = \oint_C \vec{E} \cdot d\vec{\ell} = -\frac{1}{c}\frac{\partial}{\partial t} \int_\Sigma \vec{B} \cdot d\vec{S}\,. \tag{1.28}$$

Since $\vec{B}$ will be assumed to be finite, as also is $\partial\vec{B}/\partial t$, it follows that the right-hand side goes to zero as we send the lengths of the two end lines of the loop to zero, since the area of the rectangular loop will go to zero also. If the length of each of the two long sides of the loop is $\Delta\ell$, then it follows that

$$0 = \oint \vec{E} \cdot d\vec{\ell} = (\vec{m} \times \vec{n}) \cdot (\vec{E}_2 - \vec{E}_1)\Delta\ell\,, \tag{1.29}$$

(since the line integrations along the long sides of the rectangle are in the directions of the unit vectors $\pm\vec{m} \times \vec{n}$). Using the vector identity $\vec{a} \cdot (\vec{b} \times \vec{c}\,) = \vec{b} \cdot (\vec{c} \times \vec{a}\,)$ it follows that

$$\vec{m} \cdot [\vec{n} \times (\vec{E}_2 - \vec{E}_1)] = 0\,. \tag{1.30}$$

Now, we could have chosen a rectangular loop with any orientation for this discussion. That is to say, $\vec{m}$ could be a unit vector pointing in any direction within the plane of the interface. It must therefore be that[5]

$$\vec{n} \times (\vec{E}_2 - \vec{E}_1) = 0\,. \tag{1.31}$$

Finally, we perform an analogous integral over the last Maxwell equation, $\vec{\nabla} \times \vec{H} = 1/c\,\partial\vec{D}/\partial t + (4\pi/c)\vec{J}$. The finiteness of $\partial\vec{D}/\partial t$ means that its area integral over the loop goes to zero, but $\vec{J}$ can have a non-zero area integral in general, since there can be a surface current density[6] $\vec{K}$. Thus we find

$$\oint \vec{H} \cdot d\vec{\ell} = (\vec{m} \times \vec{n}) \cdot (\vec{H}_2 - \vec{H}_1)\Delta\ell = \frac{4\pi}{c}\,\vec{m} \cdot \vec{K}\Delta\ell\,, \tag{1.32}$$

and since the left-hand side can be rewritten as $\vec{m} \cdot [\vec{n} \times (\vec{H}_2 - \vec{H}_1)]$, and the equation must hold for all choices of direction for the tangent vector $\vec{m}$, we conclude that

$$\vec{n} \times (\vec{H}_2 - \vec{H}_1) = \frac{4\pi}{c}\,\vec{K}\,. \tag{1.33}$$

To sumarise, the boundary conditions we have derived above in (1.25), (1.26), (1.31) and (1.33) are

$$\begin{aligned} \vec{n} \cdot (\vec{D}_2 - \vec{D}_1) &= 4\pi\sigma\,, \qquad \vec{n} \cdot (\vec{B}_2 - \vec{B}_1) = 0\,, \\ \vec{n} \times (\vec{E}_2 - \vec{E}_1) &= 0\,, \qquad \vec{n} \times (\vec{H}_2 - \vec{H}_1) = \frac{4\pi}{c}\,\vec{K}\,. \end{aligned} \tag{1.34}$$

These give the junction conditions at the interface between medium 1 and medium 2, where $\vec{n}$ is the unit normal vector pointing from 1 to 2 at the interface, σ is the surface charge density and $\vec{K}$ is the surface current density. Note that the first line of (1.34) comprises conditions on the components of the fields *normal* to the interface, whilst the second line comprises conditions on the components *parallel* to the interface.

A special case of frequent interest arises for an electric field in free space, in the presence of a conducting surface. In the free-space region we have $\vec{D} = \vec{E}$, and the conductor can be viewed as the surface of a medium having infinite dielectric constant, which means that $\vec{E} = 0$ there. Thus the pillbox integration of $\vec{\nabla} \cdot \vec{D} = 4\pi\rho$ becomes just the integral of $\vec{\nabla} \cdot \vec{E} = 4\pi\rho$, with $\vec{E} = 0$ in "medium 1." Th upshot is that the first and third junction

[5]Note that although eqn (1.31) is written as a three-dimensional vector equation, the right-hand side in fact has only two non-vanishing components, since it is necessarily orthogonal to the unit vector $\vec{n}$. This is why we can deduce from eqn (1.30), which holds for all vectors $\vec{m}$ lying in the plane perpendicular to $\vec{n}$, that eqn (1.31) must hold.

[6]Like the surface charge density we discussed previously, here the surface current resides entirely within the infinitesimal thickness of the interface between the two media. It is a current per unit length.

conditions in (1.34) become (dropping the "2" subscript in the free-space region outside the conductor)

$$\vec{n} \cdot \vec{E} = 4\pi\sigma\,, \qquad \vec{n} \times \vec{E} = 0 \tag{1.35}$$

at the surface. The second equation says that there is no component of $\vec{E}$ *tangent* to the conducting surface, and the first equation says that the *normal* component of the electric field at the conductor is equal to $4\pi\sigma$.

It should be emphasised that the expression in eqn (1.35) for the surface charge density on the conductor assumes that there is an electric field only on one side of the conductor. If there were electric fields on both sides of a conductor in a vacuum, then the surface charge density would be given by $\sigma = (4\pi)^{-1}\,\vec{n} \cdot (\vec{E}_2 - \vec{E}_1)$, where the unit normal vector points into region 2.

1.5 Gauge potentials

When solving Maxwell's equations, it is often convenient to express the electric and magnetic fields in terms of potentials. This has the advantage that two of the four Maxwell equations are then explicitly solved from the outset, leaving just two more, which now become second order equations for the potentials.

Specifically, the two Maxwell equations that are solved by introducing potentials are the two that do not have ρ or $\vec{J}$ as sources, namely

$$\vec{\nabla} \cdot \vec{B} = 0\,, \qquad \vec{\nabla} \times \vec{E} + \frac{1}{c}\,\frac{\partial \vec{B}}{\partial t} = 0\,. \tag{1.36}$$

Notice that these two equations are exactly the same whether one is considering the free-space case (1.3) or the case where media are present (1.19).[7]

To introduce potentials we begin by considering $\vec{\nabla} \cdot \vec{B} = 0$. This can be solved by writing $\vec{B}$ as the curl of a vector:

$$\vec{B} = \vec{\nabla} \times \vec{A}\,, \tag{1.37}$$

since the divergence of the curl of any vector vanishes identically. Passing now to the second equation in (1.36), we plug in (1.37) and deduce (after using the fact that the partial time derivative commutes with $\vec{\nabla}$) that

$$\vec{\nabla} \times \Big(\vec{E} + \frac{1}{c}\,\frac{\partial \vec{A}}{\partial t}\Big) = 0\,. \tag{1.38}$$

[7] These two Maxwell equations are known as *Bianchi identities*. By contrast, the remaining two Maxwell equations are known as the *Maxwell field equations*.

If a vector has vanishing curl it can be written as the gradient of a function, and so we can write $\vec{E} + (1/c)\partial\vec{A}/\partial t = -\vec{\nabla}\phi$.

To summarise, we can write $\vec{E}$ and $\vec{B}$ in terms of a scalar potential ϕ and a vector potential $\vec{A}$ as

$$\boxed{\vec{E} = -\vec{\nabla}\phi - \frac{1}{c}\frac{\partial \vec{A}}{\partial t}\,, \qquad \vec{B} = \vec{\nabla}\times\vec{A}\,.} \tag{1.39}$$

The choice of the potentials ϕ and $\vec{A}$ that give rise to given $\vec{E}$ and $\vec{B}$ fields via (1.39) is not unique. Since the curl of a gradient vanishes identically, we get the same $\vec{B}$ if the gradient of an arbitrary function λ is added to $\vec{A}$. Thus if we define

$$\vec{A}' = \vec{A} + \vec{\nabla}\lambda\,, \tag{1.40}$$

then $\vec{A}'$ gives the same $\vec{B}$ as does $\vec{A}$:

$$\vec{B}' \equiv \vec{\nabla}\times\vec{A}' = \vec{\nabla}\times\vec{A} + \vec{\nabla}\times\vec{\nabla}\lambda = \vec{\nabla}\times\vec{A} = \vec{B}\,. \tag{1.41}$$

It is now evident that if, at the same time, we transform ϕ to

$$\phi' = \phi - \frac{1}{c}\frac{\partial\lambda}{\partial t}\,, \tag{1.42}$$

then we shall also find that ϕ' and $\vec{A}'$ give rise to the same $\vec{E}$, via (1.39), as do ϕ and $\vec{A}$.

To summarise, the expressions (1.39) for the electromagnetic fields give the *same* $\vec{E}$ and $\vec{B}$ if we transform the potentials ϕ and $\vec{A}$ according to

$$\boxed{\phi \longrightarrow \phi' = \phi - \frac{1}{c}\frac{\partial\lambda}{\partial t}\,, \qquad \vec{A} \longrightarrow \vec{A}' = \vec{A} + \vec{\nabla}\lambda\,,} \tag{1.43}$$

where λ is an *arbitrary* function of $\vec{r}$ and t. The transformations (1.43) are known as *gauge transformations*. The potentials ϕ and $\vec{A}$ are known as *gauge potentials*.

In this course, we shall mostly be concerned with the situation when the electric and magnetic fields are *static*, i.e. they are independent of time. It is evident from the Maxwell equations (1.3) that under these circumstances the electric and magnetic fields are totally decoupled from one another. Just because $\vec{E}$ and $\vec{B}$ are static, it does not necessarily mean that the the potentials ϕ and $\vec{A}$ have to be taken to be time independent.[8] However, one

[8]One way to see this is from the gauge transformations in eqns (1.43): we could start with time-independent gauge potentials $\phi(\vec{r})$ and $\vec{A}(\vec{r})$ describing the static fields $\vec{E}(\vec{r})$ and $\vec{B}(\vec{r})$, and then willfully transform to time-dependent gauge potentials $\phi'(\vec{r},t)$ and $\vec{A}'(\vec{r},t)$ by means of the gauge transformations in (1.43), by allowing the gauge parameter $\lambda(\vec{r},t)$ to have time dependence.

would have to be quite perverse to choose to complicate a time-independent problem by opting to describe it in terms of time-dependent potentials! Thus in practice, in the static case, we always choose to take ϕ and $\vec{A}$ to be time-independent, and so (1.39) becomes simply

$$\vec{E} = -\vec{\nabla}\phi\,, \qquad \vec{B} = \vec{\nabla}\times\vec{A}\,. \tag{1.44}$$

The residual part of the gauge transformations (1.43) that preserves the time-independence of the gauge potentials is given by taking the gauge parameter λ to be of the form

$$\lambda(\vec{r}, t) = -c\,k\,t + \lambda(\vec{r})\,, \tag{1.45}$$

where k is a constant and $\lambda(\vec{r})$ is an arbitrary function of position. (The inclusion of the factor of the speed of light c and the minus sign in the gauge transformation (1.45) is just for convenience, so that the final expression takes the simpled form (1.46) below.) Thus in the static case we have independent gauge transformations under which

$$\phi \longrightarrow \phi' = \phi + k\,, \qquad \vec{A} \longrightarrow \vec{A}' = \vec{A} + \vec{\nabla}\lambda(\vec{r})\,. \tag{1.46}$$

The gauge transformation for ϕ is just the familiar freedom to add an arbitrary constant to the electrostatic potential.

1.6 Electric field of a point charge; Coulomb's law

It was found experimentally long ago, by Cavendish, Coulomb and others, that the force between two charges q_1 and q_2 in free space was proportional to the product $q_1\,q_2$; was inversely proportional to the square of the distance bewteen them (let us assume point charges, for simplicity); and was directed along the line joining the two charges. Furthermore, the force is attractive if $q_1\,q_2$ is negative, and repulsive if $q_1\,q_2$ is positive. If we work in Gaussian units, then as discussed in section 1.2, the magnitude of the force is simply equal to $q_1\,q_2$ divided by the square of the separation. All this is summarised in the equation

$$\vec{F} = q_1\,q_2\,\frac{\vec{r}_1 - \vec{r}_2}{|\vec{r}_1 - \vec{r}_2|^3}\,, \tag{1.47}$$

which gives the force on q_1 due to q_2, where $\vec{r}_1$ and $\vec{r}_2$ are the position vectors of the two point charges q_1 and q_2.

Coulomb also found that the force on a charge q was given by

$$\vec{F} = q\,\vec{E}\,, \tag{1.48}$$

and so we can read off from (1.47) that the electric field at the point $\vec{r}$ due to a charge q_1 located at the point $\vec{r}_1$ is given by

$$\vec{E}(\vec{r}) = q_1 \, \frac{\vec{r} - \vec{r}_1}{|\vec{r} - \vec{r}_1|^3} \, . \tag{1.49}$$

A very important and fundamental feature of electromagnetism is that it is described by a system of *linear* equations (see (1.3)), and so it obeys the principal of superposition. In particular, this means that if there are N point charges q_a located at positions $\vec{r}_a$, then the total electric field at $\vec{r}$ is simply the sum of the individual contributions from each charge:

$$\vec{E}(\vec{r}) = \sum_{a=1}^{N} q_a \, \frac{\vec{r} - \vec{r}_a}{|\vec{r} - \vec{r}_a|^3} \, . \tag{1.50}$$

We can generalise this result to the case where there is a continuum of charge, with charge density ρ. In the infinitesimal volume $dxdydz$ in the neighbourhood of the point $\vec{r}$ there will be an infinitesimal charge $\rho(\vec{r})dxdydz$. For convenience, we may write the volume element $dxdydz$ as $dxdydz \equiv d^3\vec{r}$. Then, we simply generalise the discrete sum (1.50) to an integral, and obtain

$$\vec{E}(\vec{r}) = \int \rho(\vec{r}\,') \frac{\vec{r} - \vec{r}\,'}{|\vec{r} - \vec{r}\,'|^3} \, d^3\vec{r}\,' \, . \tag{1.51}$$

The inverse process, of passing from the continuum result (1.51) to the discrete sum (1.50), can be achieved by means of the Dirac delta function. In one dimension the Dirac delta function $\delta(x-a)$ is a "spike" of zero width, infinite height, and total area=1 that has the property

$$f(a) = \int_{x_1}^{x_2} f(x)\delta(x-a)dx \tag{1.52}$$

for any function $f(x)$, provided that the integration interval $[x_1, x_2]$ encompasses the point $x = a$. (The integral would give zero if a did not lie inside the integration interval.) We then define the three-dimensional delta function $\delta^3(\vec{r} - \vec{a}\,)$ as

$$\delta^3(\vec{r} - \vec{a}\,) \equiv \delta(x - a_1)\delta(y - a_2)\delta(z - a_3) \, , \tag{1.53}$$

where a_1, a_2 and a_3 are the x, y and z components of the vector $\vec{a}$: i.e. $\vec{a} = (a_1, a_2, a_3)$. (In a commonly-used notation a vector $\vec{a}$ would be written in terms of its Cartesian components a_1, a_2 and a_3 as $\vec{a} = a_1\,\vec{i} + a_2\,\vec{j} + a_3\,\vec{k}$, where $\vec{i}$, $\vec{j}$ and $\vec{k}$ are unit vectors along the x, y and z directions. The notation $\vec{a} = (a_1, a_2, a_3)$ is equivalent, but more economical.) Clearly, the three-dimensional delta function has the property that

$$f(\vec{a}\,) = \int_V f(\vec{r})\delta^3(\vec{r} - \vec{a}\,)d^3\vec{r}, \tag{1.54}$$

provided that the integration volume V encompasses the point $\vec{r} = \vec{a}$ where the delta function has its "spike."

It should be noted that the Dirac delta function, and its three-dimensional generalisation that we discussed above, are both *symmetric* functions, in the sense that

$$\delta(-x) = \delta(x)\,, \qquad \delta^3(-\vec{r}) = \delta^3(\vec{r})\,. \tag{1.55}$$

Using the delta function, we can then write the charge density ρ for the set of charges in (1.50) as

$$\rho(\vec{r}) = \sum_{a=1}^{N} q_a\, \delta^3(\vec{r} - \vec{r}_a)\,. \tag{1.56}$$

Substituting this into (1.51), and using (1.54), we indeed recover (1.50).

1.7 Gauss's law

If we are considering electrostatics, i.e. the situation where there is a time-independent electric field and no magnetic field, the Maxwell equations (1.3) in free space reduce to

$$\vec{\nabla}\cdot\vec{E} = 4\pi\rho\,, \qquad \vec{\nabla}\times\vec{E} = 0\,. \tag{1.57}$$

By integrating the first equation over a volume V, and using the divergence theorem (1.22), we obtain Gauss's law

$$\int_S \vec{E}\cdot d\vec{S} = 4\pi Q\,, \tag{1.58}$$

where S is the closed surface surrounding the volume V, and Q is the total charge contained within the volume V:

$$Q = \int_V \rho dV\,. \tag{1.59}$$

(Actually, historically, the Maxwell equation $\vec{\nabla}\cdot\vec{E} = 4\pi\rho$ was discovered experimentally in its equivalent integrated form (1.58).)

It is instructive to examine Gauss's law in the special case of a single point charge q. Since, when we set up a Cartesian coordinate system we can choose the origin to be at any arbitrary point, it is convenient to choose it so that the charge sits at the origin. Using (1.49) we see that the electric field of the charge will be given by

$$\vec{E} = \frac{q\,\vec{r}}{r^3}\,. \tag{1.60}$$

Let us check that this is consistent with the Maxwell equation $\vec{\nabla}\cdot\vec{E} = 4\pi\rho$, and its integrated form (1.58).

First, we calculate the divergence of $\frac{\vec{r}}{r^3}$. Clearly we have $\vec{\nabla}\cdot\vec{r} = \frac{\partial x}{\partial x} + \frac{\partial y}{\partial y} + \frac{\partial z}{\partial z} = 3$, and since $r^2 = x^2 + y^2 + z^2$ we have, differentiating, $2r\,\frac{\partial r}{\partial x} = 2x$, etc., and hence $\frac{\partial r}{\partial x} = \frac{x}{r}$, etc. Thus we have

$$\vec{\nabla}\cdot\vec{r} = 3\,, \qquad \vec{\nabla} r = \frac{\vec{r}}{r}\,, \tag{1.61}$$

and so[9]

$$\vec{\nabla}\cdot\Big(\frac{\vec{r}}{r^3}\Big) = \frac{\vec{\nabla}\cdot\vec{r}}{r^3} - \frac{3\vec{r}}{r^4}\cdot\frac{\vec{r}}{r} = \frac{3}{r^3} - \frac{3}{r^3} = 0\,. \tag{1.62}$$

This shows that for (1.60) we have in general $\vec{\nabla}\cdot\vec{E} = 0$. This is perfectly correct away from the origin, since we hope to find $\vec{\nabla}\cdot\vec{E} = 4\pi\rho$ and indeed $\rho = 0$ away from the origin. However, at the origin the calculation (1.62) is not trustworthy, because there are denominators that go to zero at $r = 0$. The safe way to handle this is to consider the integrated form of the equation, in (1.58).

Let us take the volume V in (1.58) to be a sphere of radius R centred on the origin. Plugging in (1.60), we shall then have on the left-hand side

$$q\int_S \frac{\vec{r}}{r^3}\cdot d\vec{S} = q\int_S \frac{\vec{n}\cdot d\vec{S}}{R^2} = q\int d\Omega = 4\pi q\,, \tag{1.63}$$

where we have defined the unit vector $\vec{n} \equiv \vec{r}/r$ which is the unit outward normal on the sphere of radius r, and where $d\Omega$ is the area element on the unit sphere (i.e. the solid angle element).

We obtained the result (1.63) by choosing to integrate over a sphere of radius R centred on the origin, but it is obvious that the result would be the same for *any* closed surface that surrounded the origin, and that instead we would get zero if the integration surface did not surround the origin. This can be seen as follows. Let the surface of the sphere of radius R be denoted by S, and now consider also some other closed surface S', of arbitrary shape. For the sake of definiteness let's suppose that S' lies entirely outside S. We shall use $\widetilde{V}$ to denote the 3-volume between the surfaces S and S'. Now integrate the divergence of the electric field $\vec{E}$ over the volume $\widetilde{V}$, and use the divergence theorem:

$$\int_{\widetilde{V}} \vec{\nabla}\cdot\vec{E}\, d^3\vec{r} = \int_{S'} \vec{E}\cdot d\vec{S} - \int_S \vec{E}\cdot d\vec{S}\,. \tag{1.64}$$

Note that the total surface bounding the volume $\widetilde{V}$ is the sum of the outer surface S' and the inner surface S. When applying the divergence theorem the area element should always be directed *outwards* from the volume, which is why the second integral on the right-hand side of (1.64) has a minus sign. Now in the volume integral on the left-hand side of (1.64)

[9] We are using the fact that for any scalar field f and vector field $\vec{v}$, we have $\vec{\nabla}\cdot(f\vec{v}) = (\vec{\nabla} f)\cdot\vec{v} + f\vec{\nabla}\cdot\vec{v}$.

we can use the result (1.62) that $\vec{\nabla}\cdot\vec{E}=0$, since the volume $\widetilde{V}$ in (1.64) does not include the origin. Hence we have

$$\int_{S'}\vec{E}\cdot d\vec{S}=\int_{S}\vec{E}\cdot d\vec{S}\,, \tag{1.65}$$

thus proving that the result in (1.63) is independent of how we choose the surface of integration, as long as it encloses the origin.

We have seen that the volume integral of $\vec{\nabla}\cdot\vec{E}$ for the point charge at the origin will give $4\pi q$ if the volume of integration includes the origin, whilst it will instead give zero if the volume does not include the origin. The conclusion from this is that the function ρ on the right-hand side of the Maxwell equation $\vec{\nabla}\cdot\vec{E}=4\pi\rho$ must be a three-dimensional delta function centred on the origin, and that therefore the precise calculation of $\vec{\nabla}\cdot\vec{E}$ gives

$$\vec{\nabla}\cdot\vec{E}=4\pi q\,\delta^3(\vec{r})\,,\qquad \text{for}\quad \vec{E}=\frac{q\,\vec{r}}{r^3}\,. \tag{1.66}$$

In other words, a point charge q located at the origin is described by the charge density

$$\rho(\vec{r})=q\,\delta^3(\vec{r})\,. \tag{1.67}$$

1.8 Electrostatic potential

In section 1.5 we introduced gauge potentials in terms of which the electric and magnetic fields could be expressed. For the case of electrostatics, we have the particularly simple situation that the electric field is written purely in terms of a scalar potential ϕ, with

$$\vec{E}=-\vec{\nabla}\phi\,, \tag{1.68}$$

where an arbitrary constant k can clearly be added to ϕ without altering $\vec{E}$. We can see by inspection that the second Maxwell equation in (1.57) is identically satisfied when we write $\vec{E}=-\vec{\nabla}\phi$, and also that the first Maxwell equation in (1.57) becomes the Poisson equation

$$\nabla^2\phi=-4\pi\rho\,. \tag{1.69}$$

Solving any problem in electrostatics is therefore reduced to solving the Poisson equation (1.69) for a given charge density ρ, subject to given boundary conditions. This is called a *Boundary Value Problem*.

First, let us consider the potential due to a point charge q in an infinite free space. Taking the charge to be located at the origin for convenience, we therefore need to find a function ϕ such that $-\vec{\nabla}\phi$ is equal to $\vec{E}=q\,\frac{\vec{r}}{r^3}$ as in (1.60). From the second equation in (1.61) we see that $\vec{\nabla}(\frac{1}{r})=-\frac{\vec{r}}{r^3}$, and so we may take the electrostatic potential to be

$$\phi=\frac{q}{r}\,. \tag{1.70}$$

Of course having the charge located at the origin was an inessential convenience, and we can readily write down the answer in the case where the charge is instead located at the point $\vec{r}_1$ just by shifting the origin of the Cartesian axes, giving

$$\phi(\vec{r}) = \frac{q}{|\vec{r} - \vec{r}_1|}\,. \tag{1.71}$$

Note that because of the residual gauge symmetry $\phi \longrightarrow \phi + \text{constant}$ we can add an arbitrary constant to ϕ if we wish, but in fact the choice of this "integration constant" that has been made in writing (1.71) is already rather natural, since it means that $\phi(\vec{r})$ goes to zero as $\vec{r}$ goes to infinity.

Because of the linearity of the equations the generalisation to a system of N point charges q_a located at positions $\vec{r}_a$ is immediate:

$$\phi(\vec{r}) = \sum_{a=1}^{N} \frac{q_a}{|\vec{r} - \vec{r}_a|}\,. \tag{1.72}$$

For the case of a continuous charge distribution, we similarly have

$$\boxed{\phi(\vec{r}) = \int \frac{\rho(\vec{r}\,')}{|\vec{r} - \vec{r}\,'|}\, d^3\vec{r}\,'\,.} \tag{1.73}$$

These expressions are analogous to equations (1.50) and (1.51) for the electric field. Indeed, one can easily explicitly see that calculating $-\vec{\nabla}\phi$ for (1.72) or (1.73) gives (1.50) or (1.51) respectively. To do these calculations, one must be careful to note that $\vec{\nabla}$ means the gradient with respect to the coordinates in $\vec{r}$, and that $\vec{r}_a$ in (1.72), or $\vec{r}\,'$ in (1.73), are treated as constants in the differentiation. Thus one uses the result that

$$\vec{\nabla}\frac{1}{|\vec{r} - \vec{r}\,'|} = -\frac{\vec{r} - \vec{r}\,'}{|\vec{r} - \vec{r}\,'|^3}\,. \tag{1.74}$$

We had previously obtained the result (1.66) that the electric field for a point charge q at the origin, namely $\vec{E} = q\,\vec{r}/r^3$, satisfies $\vec{\nabla}\cdot\vec{E} = 4\pi q\,\delta^3(\vec{r})$. Since $\vec{E} = -\vec{\nabla}\phi$ with $\phi = q/r$, we therefore have

$$\nabla^2\phi = -4\pi q\,\delta^3(\vec{r})\,. \tag{1.75}$$

More generally, if the charge q is located at $\vec{r} = \vec{r}\,'$ so that the potential is given by $\phi(\vec{r}) = q\,|\vec{r} - \vec{r}\,'|^{-1}$, we shall therefore have

$$\nabla^2\phi = -4\pi q\,\delta^3(\vec{r} - \vec{r}\,')\,. \tag{1.76}$$

From this we can read off a result that will be useful frequently in the future, namely that

$$\boxed{\nabla^2\,\frac{1}{|\vec{r} - \vec{r}\,'|} = -4\pi\,\delta^3(\vec{r} - \vec{r}\,')\,.} \tag{1.77}$$

1.9 Electrostatic energy

The force on a charge q in an electric field $\vec{E}$ is given by $\vec{F} = q\vec{E}$. It follows that the work done in moving it from a point P_1 to a point P_2 is given by

$$\Delta W = -\int_{P_1}^{P_2} \vec{F} \cdot d\vec{\ell} = -q \int_{P_1}^{P_2} \vec{E} \cdot d\vec{\ell}. \tag{1.78}$$

Using $\vec{E} = -\vec{\nabla}\phi$ therefore gives

$$\Delta W = q \int_{P_1}^{P_2} \vec{\nabla}\phi \cdot d\vec{\ell} = q\phi(P_2) - q\phi(P_1)\,. \tag{1.79}$$

From this, we can read off the potential energy U of the charge q in the electrostatic field as being

$$\boxed{U = q\,\phi\,.} \tag{1.80}$$

If we consider a system of N charges q_a at point $\vec{r}_a$ in free space, then the the electrostatic energy of a particular charge q_a in to the potential due to the other $N-1$ charges will therefore be

$$U_a = q_a \sum_{b\neq a} \frac{q_b}{|\vec{r}_a - \vec{r}_b|}\,, \tag{1.81}$$

where the sum is taken over the remaining $N-1$ charges. The total potential energy will then be given by

$$U = \sum_{a=1}^{N} \sum_{b<a} \frac{q_a q_b}{|\vec{r}_a - \vec{r}_b|}\,. \tag{1.82}$$

(The second summation is over $b < a$ rather than over all $b \neq a$ to avoid a double counting: the energy of charge 1 in the field of charge 2 is the same as the energy of charge 2 in the field of charge 1, but this energy counts *only once* to the total energy, not twice.) One way to view this is that we assemble the charges one at a time. Bringing up the first charge requires no energy, since there is no electrostatic potential (apart from its own) at that step. Bringing up the second charge requires doing work against the potential of the first charge that is already there. Bringing up the third charge requires doing work against the potentials of the first and second charges. And so on.

It is easy to see that another way of writing (1.82) is

$$U = \tfrac{1}{2} \sum_{a} \sum_{b\neq a} \frac{q_a q_b}{|\vec{r}_a - \vec{r}_b|}\,. \tag{1.83}$$

where it is understood that, aside from the exclusion $b \neq a$, the indices a and b range over 1 to N. We can rewrite the summations like this because the quantity $\frac{q_a\,q_b}{|\vec{r}_a - \vec{r}_b|}$ that is being

summed over is symmetric under exchanging the labels a and b. Note that the excluded terms where $b = a$ would be infinite, since the denominator $|\vec{r}_a - \vec{r}_b|$ would be zero. In some sense one could think of the excluded terms as representing the "self-energies" of the individual charges in their own electrostatic potentials. Note that although these terms would be infinite, they would in a sense be harmless even if they were included in the double summation, since they would be independent of the locations of the charges. What one really cares about when considering a potential energy is how it *changes* when the charges are moved around, since the utility of the concept of potential energy is to calculate forces by seeing how it changes when the charges are displaced. Thus an additive constant term (albeit infinite) in the expression for the potential energy, that is unaltered as the charges are moved, could just be neglected, or subtracted out, without really affecting anything physical.[10]

Generalising (1.83) to the case of a continuous charge distribution clearly gives[11]

$$U = \tfrac{1}{2} \int \int \frac{\rho(\vec{r})\rho(\vec{r}\,')}{|\vec{r} - \vec{r}\,'|} \, d^3\vec{r} \, d^3\vec{r}\,' \, . \tag{1.84}$$

In view of the fact that the potential $\phi(\vec{r})$ is given by (1.73), we can rewrite (1.84) as

$$U = \tfrac{1}{2} \int \rho(\vec{r})\phi(\vec{r}) d^3\vec{r}. \tag{1.85}$$

We can also rewrite the energy purely in terms of the electric field. To do this, we use the Poisson equation (1.69) to write ρ in (1.85) as $\rho = -\frac{1}{4\pi} \nabla^2\phi$, giving

$$\begin{aligned} U &= -\frac{1}{8\pi} \int \phi \nabla^2 \phi d^3\vec{r}, \\ &= -\frac{1}{8\pi} \int \vec{\nabla} \cdot (\phi \vec{\nabla}\phi) d^3\vec{r} + \frac{1}{8\pi} \int |\vec{\nabla}\phi|^2 \, d^3\vec{r}, \\ &= -\frac{1}{8\pi} \int_S \phi \vec{\nabla}\phi \cdot d\vec{S} + \frac{1}{8\pi} \int |\vec{\nabla}\phi|^2 \, d^3\vec{r}, \\ &= \frac{1}{8\pi} \int |\vec{\nabla}\phi|^2 \, d^3\vec{r}. \end{aligned} \tag{1.86}$$

[10] A very similar issue of infinities arises in the quantum theory of the electromagnetic field. This also is resolved, in a procedure known as renormalisation, by being careful to distinguish between quantities that are truly physical and those that are not.

[11] Note that here the same issue of infinities from "self-energies"arises again, and here one cannot simply avoid the problem by saying $\vec{r} \neq \vec{r}\,'$ (the analogue of the exclusion $b \neq a$ in (1.83)), since $\vec{r}$ can be very close to $\vec{r}\,'$, without being equal, and still the integrand would be very large and liable to lead to a divergent expression for the energy. The important point, again, is that the divergent term in the expression for the energy would be a pure constant that could be subtracted out without upsetting any physically-meaningful calculation.

Note that after performing the integration by parts here, we have dropped the surface term coming from the "sphere at infinity" since the fields are assumed to go to zero there; see the discussion below.

Let us pause here to comment on two of the concepts employed in the previous paragraph, namely integration by parts, and also the "sphere at infinity." In the familiar case of a one-dimensional integral, integration by parts amounts to saying that in the integral

$$\int_a^b u(x)\, v'(x)\, dx\,, \tag{1.87}$$

where $v'(x)$ means $dv(x)/dx$, we can write $u\, v'$ as[12]

$$u\, v' = (u\, v)' - u'\, v\,, \tag{1.88}$$

and so, plugging into the integral (1.87), we get

$$\begin{aligned} \int_a^b u\, v'\, dx &= \int_a^b (u\, v)'\, dx - \int_a^b u'\, v\, dx\,, \\ &= \Big[u(x)\, v(x)\Big]_a^b - \int_a^b u'\, v\, dx\,. \end{aligned} \tag{1.89}$$

In the calculation in eqn (1.86) above, we did something analogous, but in three dimensions. Stated generally, for an arbitrary scalar field ψ and vector field $\vec{v}$, we write $\psi\, \vec{\nabla}\cdot\vec{v}$ as the total derivative $\vec{\nabla}\cdot(\psi\,\vec{v})$ minus the correction term $\vec{\nabla}\psi\cdot\vec{v}$. Integrating over a volume V bounded by the surface S, and using the divergence theorem, we have

$$\begin{aligned} \int_V \psi\, \vec{\nabla}\cdot\vec{v}\, dV &= \int_V \vec{\nabla}\cdot(\psi\,\vec{v})\, dV - \int_V \vec{\nabla}\psi\cdot\vec{v}\, dV\,, \\ &= \int_S \psi\,\vec{v}\cdot d\vec{S} - \int_V \vec{\nabla}\psi\cdot\vec{v}\, dV\,. \end{aligned} \tag{1.90}$$

(In the case of (1.86) above, ψ is equal to ϕ and $\vec{v}$ is equal to $\vec{\nabla}\phi$.)

Note that we shall frequently be making use of this kind of three-dimensional integration by parts. It is helpful to keep in mind, therefore, that the notion of "integration by parts" is much broader than the special case in (1.89) of integration by parts in one dimension.

The other thing we did in the derivation in (1.86) was to drop the boundary term after integrating by parts. The point here is that in our volume integrations in (1.86), we are integrating over all space. One way to make precise what is going on is to think of first integrating just over the volume V interior to a sphere of radius R. So the boundary surface S is then the surface of the sphere of radius R. At the end of the day, we shall send R

[12] That is, we write $u\, v'$ as the total derivative $(u\, v)'$ minus the correction term $u'\, v$.

to infinity, meaning that V will then be the whole of space, and S will be the "sphere at infinity." We can see that in the calculation above it is legitimate to drop that surface integral over the sphere at infinity because the integrand, i.e. $\phi\,\vec{\nabla}\phi$, falls off sufficiently fast as R is sent to infinity. Specifically, our localised charge distribution will have a scalar potential that falls of like $\frac{1}{r}$ at large distance (just like the potential of a single point charge). This means that $\vec{\nabla}\phi$ will fall off like $\frac{1}{r^2}$, and thus $\phi\,\vec{\nabla}\phi$ will fall off like $\frac{1}{r^3}$. On the other hand, the area element $d\vec{S}$ of a sphere of radius r will grow like r^2. Thus on the spherical surface S at radius R, the integral $\int_S \phi\,\vec{\nabla}\phi\cdot d\vec{S}$ will be falling off like $\frac{1}{R}$, and this means it will go to zero as R goes to infinity. This is why we can drop the surface term in the penultimate line of (1.86) coming from the "sphere at infinity."

Since $\vec{E}=-\vec{\nabla}\phi$, the energy U in (1.86) can be written as

$$U=\frac{1}{8\pi}\int |\vec{E}|^2 d^3\vec{r}, \tag{1.91}$$

integrated over all space. This leads naturally to the definition of *energy density* in the electric field as

$$w=\frac{1}{8\pi}\,|\vec{E}|^2\,. \tag{1.92}$$

It is of interest to apply this result to the electric field at the surface of a conductor. Recall from section 1.4 that we found that the electric field is always normal to the surface of a conductor, and that there is a surface-charge density σ given by $\vec{n}\cdot\vec{E}=4\pi\sigma$ (see(1.35)). Furthermore, at the conductor $\vec{n}\times\vec{E}=0$, so there is no tangential component of $\vec{E}$. This means that very close to the conductor $\vec{E}$ must be very nearly parallel to the normal vector $\vec{n}$, so $\vec{E}\approx(\vec{n}\cdot\vec{E})\,\vec{n}=4\pi\sigma\,\vec{n}$. Thus in the neighbourhood of the conductor there is an energy density

$$w=\frac{1}{8\pi}\,|\vec{E}|^2=2\pi\sigma^2\,. \tag{1.93}$$

If an element of the surface of area δA is displaced outwards by a distance δx, the electrostatic energy will then change by an amount

$$\delta U=-w\delta A\,\delta x=-2\pi\sigma^2\delta A\,\delta x\,, \tag{1.94}$$

and so there is an outward force δF given by $\delta U=-\delta F\,\delta x$. This implies an outward force per unit area (i.e. pressure) given by $p=\delta F/\delta A$, with

$$p=2\pi\sigma^2=\frac{1}{8\pi}\,|\vec{E}|^2\,. \tag{1.95}$$

Note that this formula for the pressure assumes that there is an electric field on only one side of the conducting surface. If there were electric fields $\vec{E}_1$ on one side and $\vec{E}_2$ on the

other side, then there would be a pressure $p_1 = \frac{1}{8\pi}\,|\vec{E}_1|^2$ exerted by the electric field $\vec{E}_1$ and a pressure $p_2 = \frac{1}{8\pi}\,|\vec{E}_2|^2$ exerted by the electric field $\vec{E}_2$. The pressure p_1 would be directed outwards from the region 1 (i.e. towards region 2), and the pressure p_2 would be directed outwards from region 2 (i.e. towards region 1). Thus the total pressure $p = p_1 - p_2$ on the conductor (in the direction from region 1 to region 2) cannot be written in terms of total surface charge density $\sigma = \frac{1}{4\pi}\,\vec{n}\cdot(\vec{E}_2 - \vec{E}_1)$ in the case when there are electric fields on both sides of the conductor. It is only in the case that there is an electric field on just one side that the formula (1.95) can be applied.

1.10 Capacitance

For a system of N conductors X_a, held at potentials V_a (for $1 \leq a \leq N$), and carrying charges Q_a, the total electrostatic potential energy, following from (1.85), is given by

$$U = \tfrac{1}{2}\sum_{a=1}^{N} Q_a\, V_a\,. \tag{1.96}$$

Since the electrostatic potential is always a linear function of the charge density, it follows that for this configuration of charged conductors there must exist a linear relation of the form

$$V_a = \sum_{b=1}^{N} \beta_{ab}\, Q_b\,, \tag{1.97}$$

for a certain set of constants β_{ab}. These constants will be characteristic of the geometry of the set of conductors, independent of the charges and the potentials. The set of N equations (1.97) can be inverted, to express the charges as functions of the potentials:

$$Q_a = \sum_{b=1}^{N} C_{ab}\, V_b\,. \tag{1.98}$$

Each diagonal element C_{aa} is called the *capacitance* of the corresponding conductor X_a, and the off-diagonal elements C_{ab} with $a \neq b$ are called the *coefficients of induction*.

In particular, the capacitance of a given conductor is equal to the charge on that conductor when it is held at unit potential, with all the other conductors held at zero potential.

Going back to the expression (1.96) for the energy U of the system of conductors, it follows from (1.98) that we may express it purely in terms of C_{ab} and the potentials V_a, as the quadratic form

$$U = \tfrac{1}{2}\sum_{a=1}^{N}\sum_{b=1}^{N} C_{ab}\, V_a\, V_b\,. \tag{1.99}$$

2 Uniqueness Theorem, Green Functions and Method of Images

2.1 Uniqueness theorem

Whenever one is solving a differential equation, such as the Poisson equation $\nabla^2\phi = -4\pi\rho$ that we encounter in electrostatics, the question arises as to what boundary conditions one must impose in order to obtain a unique solution. Expressed more physically, one may ask how much boundary information must be specified in order to pin down the physics of the problem completely.

One answer for Poisson's equation is that the solution for the potential ϕ inside a volume V will be uniquely determined if its value at all points on the (closed) surface S that bounds V is specified. For example, if we are solving $\nabla^2\phi = -4\pi\rho$ inside a sphere, then the solution will be uniquely determined if the value of ϕ at every point on the surface of the sphere is specified. This type of boundary condition, in which ϕ is specified on S, is known as a *Dirichlet boundary condition*.

An alternative possibility is to specify not ϕ itself but its *normal derivative* $\partial\phi/\partial n$ on the boundary, where

$$\frac{\partial\phi}{\partial n} \equiv \vec{n}\cdot\vec{\nabla}\,\phi\,. \tag{2.1}$$

That is, the derivative of ϕ along the direction $\vec{n}$ of the normal to the surface S. (Recall $\vec{n}$ is the unit-length normal vector.) This boundary condition, where $\partial\phi/\partial n$ is specified on S, is known as a *Neumann boundary condition*. Since the electric field is given in terms of the potential by $\vec{E} = -\vec{\nabla}\phi$, this means that for the Neumann boundary condition, the normal component of the electric field is specified on the boundary S. In this case the solution for ϕ is again unique, except for the (trivial) point that an arbitrary additive constant is undetermined. Thus, for the two cases the boundary data will be of the form

$$\text{Dirichlet:} \qquad \phi(\vec{r})\Big|_S = f_D\,, \tag{2.2}$$

$$\text{Neumann:} \qquad \frac{\partial\phi(\vec{r})}{\partial n}\Big|_S = f_N\,, \tag{2.3}$$

where f_D or f_N is a function, defined for all points *on the boundary* S, which characterises the boundary data one wishes to impose.

One can also consider mixed boundary conditions, which are Dirichlet on parts of the boundary and Neumann on the rest. We could also allow a more general possibility known

as *Robin boundary conditions*,[13] where the potential ϕ on the boundary obeys

$$\text{Robin:} \qquad \Big(\frac{\partial\phi}{\partial n} + a\,\phi\Big)\Big|_S = f_R\,, \tag{2.4}$$

where f_R is a freely-specifiable function on the boundary, and where a is a strictly non-negative, but otherwise freely specifiable, function on the boundary. So in this case, the functions f_R and a comprise the boundary data for the Robin boundary-value problem. The reason for the restriction that $a \geq 0$ at all points on S will become evident shortly.

The uniqueness statements can be proven by supposing the contrary; that is, we suppose that for the given boundary condition on S and a given charge density ρ, there exist two *different* solutions to $\nabla^2\phi = -4\pi\rho$. Let these solutions be ϕ_1 and ϕ_2. The idea will be to try to prove that actually $\phi_1 = \phi_2$, and so the solution is unique. With

$$\nabla^2\phi_1 = -4\pi\rho \qquad \text{and} \qquad \nabla^2\phi_2 = -4\pi\rho\,, \tag{2.5}$$

it follows by subtraction that the function ψ defined by

$$\psi \equiv \phi_1 - \phi_2 \tag{2.6}$$

will satisfy Laplace's equation

$$\nabla^2\psi = 0 \tag{2.7}$$

in the volume V.

Let us consider the cases of pure Dirichlet or pure Neumann boundary conditions first. Since ϕ_1 and ϕ_2 by definition satisfy identical boundary conditions on S, either (2.2) in the Dirichlet case or (2.3) in the Neumann case, it follows that ψ will satisfy either $\psi = 0$ (Dirichlet) or $\partial\psi/\partial n = 0$ (Neumann) on S.

We now multiply (2.7) by ψ, integrate over V, and then perform an integration by parts:

$$\begin{aligned} 0 &= \int_V \psi\nabla^2\psi\,dV\,, \\ &= \int_V \Big[\vec\nabla\cdot(\psi\vec\nabla\psi) - \vec\nabla\psi\cdot\vec\nabla\psi\Big]dV\,, \\ &= \int_S \psi\vec\nabla\psi\cdot d\vec S - \int_V |\vec\nabla\psi|^2\,dV\,, \\ &= \int_S \psi\,\frac{\partial\psi}{\partial n}\,dS - \int_V |\vec\nabla\psi|^2\,dV\,. \end{aligned} \tag{2.8}$$

(The first term on the penultimate line comes by using the divergence theorem (1.22).) Now the area element $d\vec S$ in the surface integral is in the direction of the unit normal $\vec n$

[13] Pronounced roughly like "Roe-ban;" Victor Gustave Robin was a 19th century French mathematician.

(i.e. $d\vec{S} = \vec{n}\, dS$), and so we see that since either ψ or its normal derivative $\partial\psi/\partial n \equiv \vec{n}\cdot\vec{\nabla}\psi$ vanishes at all points on S, we are left with

$$\int_V |\vec{\nabla}\psi|^2\, dV = 0\,. \tag{2.9}$$

The integrand is everywhere non-negative, and so the integral can only be zero if the integrand vanishes everywhere in V. But if $|\vec{\nabla}\psi|^2$=0 it follows that

$$\vec{\nabla}\psi = 0 \tag{2.10}$$

everywhere in V, and so we conclude that $\psi =$constant everywhere in V. In other words, we have proved that $\psi = k$ and hence

$$\phi_1 = \phi_2 + k\,, \tag{2.11}$$

where k is a constant.

In the case of Dirichlet boundary conditions we know that $\phi_1 = \phi_2$ on S, and so the constant k must be zero. This proves that $\phi_1 = \phi_2$ everywhere in V, thus establishing that the solution is unique.

In the case of Neumann boundary conditions, where only the normal derivative is specified on S, it is clear that the constant k can never be determined. This is of no consequence, since ϕ and $\phi + k$ give rise to the same physical $\vec{E}$ field in any case. So in the Neumann case too, the solution for the electric field $\vec{E}$ is unique.

Note that the results above can apply not only to the problem of solving for ϕ inside a finite volume V with finite-sized closed boundary S, but also in the case where the volume V is infinite. A typical example would be when there is a finite-sized surface S_1 (for example a spherical conductor) and the volume V is taken to be the entire infinite space outside it. In this case there is no actual boundary at infinity, but we can treat the problem by imagining that we introduce a spherical boundary surface S_2 at some very large radius R, and eventually we send R to infinity. When R is large but finite, we have a finite volume V bounded by the disconnected sum of the two surfaces S_1 (in the middle) and S_2 (at large distance).

The uniqueness arguments discussed above can then be applied to this situation, with the surface integral in (2.8) becoming the sum of two integrals, one over the component S_1 of the total boundary and the other over the component S_2. Dirichlet or Neumann boundary conditions are specified on S_1, and so that contribution to the surface integral will vanish. The surface integral over S_2 will become zero in the limit that the radius R is

sent to infinity, provided that ϕ goes to zero sufficiently fast at infinity. Thus in practice we think of S_2 as "the sphere at infinity," and we impose the boundary condition that ϕ goes to zero sufficiently rapidly at infinity, thereby ensuring that the S_2 component of the surface integral in (2.8) will vanish too. This ensures that again we are left with just the volume integral (2.9), and so the uniqueness proof goes through as before.

Note also that we can allow multiple disconnected surfaces at finite distance, provided that Dirichlet or Neumann boundary conditions are imposed on all of them.

In summary, therefore, we have a uniqueness proof too in the case where the volume V is infinite, provided that we not only impose Dirichlet or Neumann boundary conditions on any boundary surfaces at finite distance, but we also impose a fall-off condition on the potential at infinity.

Note that we have established uniqueness of the solution subject to the imposition of *either* Dirichlet *or* Neumann boundary conditions at each point on the boundary. It could be Dirichlet for some points, and Neumann for others, but at any given point one can specify only one of Dirichlet or Neumann.

In the case of Robin boundary conditions, the difference $\psi = \phi_1 - \phi_2$ between two solutions of $\nabla^2\phi = -4\pi\rho$ obeying the *same* Robin boundary condition (2.4) (and so the same functions f_R and a for each of ϕ_1 and ϕ_2) will therefore satisfy $\partial\psi/\partial n + a\psi = 0$ on the boundary surface S; in other words $\partial\psi/\partial n = -a\psi$ on S. The vanishing of the last line in the uniqueness proof following from integrating $\psi\nabla^2\psi = 0$ as in (2.8) will give (after multiplying by -1 for convenience)

$$\int_S a\,\psi^2\,dS + \int_V |\vec{\nabla}\psi|^2\,dV = 0\,. \tag{2.12}$$

In this case, then, the surface integral is not immediately zero like it was in the Dirichlet or Neumann cases, but instead we have that the integrand in the surface integral is non-negative everywhere on S. (Recall that the function a in (2.4) was required to be everywhere non-negative on the boundary S.) Thus, by an extension of the argument we used in the Dirichlet and Neumann cases, we now have two integrals, each of which has a pointwise non-negative integrand, summing up to zero. This can only be true if each integrand vanishes at each point within its integration domain, so the volume integral tells us that $\vec{\nabla}\psi = 0$ in V and hence $\psi = k$ where k is a constant in V, while the surface integral tells us that $\psi = 0$ on S. Thus we see that k must equal zero and hence we conclude that $\psi = 0$ everywhere, implying $\phi_1 = \phi_2$ and hence uniqueness of the solution.

With any of the Dirichlet, Neumann, mixed or Robin boundary conditions specified, the problem is said to be *well posed*. This means that these boundary conditions are neither

too weak, leaving the problem underdetermined and not fully pinned down, nor are they too strong, leaving the problem overdetermined and therefore admitting no solution.

An example of an overdetermined problem would be if one tried to impose *both* Dirichlet *and* Neumann boundary conditions at each point on S. In other words, if one tried to specify both the potential *and* its normal derivative at each point on S. Specifying both ϕ and $\partial\phi/\partial n$ on S is known as specifying *Cauchy* boundary conditions. That this would be an overdetermination is obvious from the fact that Dirichlet conditions alone are sufficient to give a unique solution. And, on the other hand, Neumann conditions alone are sufficient to give another unique solution. Except in the unlikely event that one picked precisely the "matching" set of Neumann conditions that would reproduce the solution with the Dirichlet conditions, there will be a conflict between the two, implying that no solution would exist.

2.2 Green's theorem

In section 1.8 we gave the expression (1.73) for the electrostatic potential due to a distribution of charge with charge density ρ:

$$\phi(\vec{r}) = \int \frac{\rho(\vec{r}\,')}{|\vec{r} - \vec{r}\,'|}\, d^3\vec{r}\,' \tag{2.13}$$

This result assumes that the charge distribution exists in otherwise free space, with no conductors or other boundaries present. In practice, as we have already remarked, a typical realistic situation is one where there are other conductors, etc., on which boundary conditions are specified.

To handle the case where there are boundaries, the following procedure can be useful. We first derive a simple result known as *Green's theorem*, and then apply it to the case of interest.

Let $\phi(\vec{r})$ and $\psi(\vec{r})$ be two scalar functions. We can then consider

$$\begin{aligned} \vec{\nabla}\cdot(\phi\vec{\nabla}\psi - \psi\vec{\nabla}\phi) &= \vec{\nabla}\phi\cdot\vec{\nabla}\psi + \phi\nabla^2\psi - \vec{\nabla}\psi\cdot\vec{\nabla}\phi - \psi\nabla^2\phi\,, \\ &= \phi\nabla^2\psi - \psi\nabla^2\phi\,. \end{aligned} \tag{2.14}$$

Integrating this over a volume V bounded by surface S, and using the divergence theorem (1.22), we therefore find

$$\boxed{\int_V \Big(\phi(\vec{r})\nabla^2\psi(\vec{r}) - \psi(\vec{r})\nabla^2\phi(\vec{r})\Big)d^3\vec{r} = \int_S \Big(\phi(\vec{r})\vec{\nabla}\psi(\vec{r}) - \psi(\vec{r})\vec{\nabla}\phi(\vec{r})\Big)\cdot d\vec{S}\,.} \tag{2.15}$$

This is Green's theorem. Actually, we shall apply it to functions of a primed position vector $\vec{r}\,'$ rather than $\vec{r}$, so we shall use

$$\int_V \Big[\phi(\vec{r}\,')\nabla'^2\psi(\vec{r}\,')-\psi(\vec{r}\,')\nabla'^2\phi(\vec{r}\,')\Big]d^3\vec{r}\,' = \int_S \Big[\phi(\vec{r}\,')\vec{\nabla}\,'\psi(\vec{r}\,')-\psi(\vec{r}\,')\vec{\nabla}\,'\phi(\vec{r}\,')\Big]\cdot d\vec{S}\,', \quad (2.16)$$

where $\vec{\nabla}\,'$ denotes the gradient with respect to the primed position vector coordinates, and $d\vec{S}\,'$ denotes the area element using the primed variables also.

We now apply (2.16) to our electrostatics problem by taking ϕ to be the electrostatic potential satisfying Poisson's equation (1.69), but written using the primed coordinates (i.e. $\nabla'^2\,\phi(\vec{r}\,') = -4\pi\,\rho(\vec{r}\,')$), and taking

$$\psi(\vec{r}\,') = \frac{1}{|\vec{r}-\vec{r}\,'|}\,. \quad (2.17)$$

Note that the unprimed position vector $\vec{r}$ just behaves as a constant in this discussion. Thus (2.16) becomes

$$\begin{aligned}&\int_V \Big[\phi(\vec{r}\,')\,\nabla'^2\frac{1}{|\vec{r}-\vec{r}\,'|} - \frac{1}{|\vec{r}-\vec{r}\,'|}\,\nabla'^2\,\phi(\vec{r}\,')\Big]\,d^3\vec{r}\,' \\ &= \int_S \Big[\phi(\vec{r}\,')\vec{\nabla}'\,\frac{1}{|\vec{r}-\vec{r}\,'|} - \frac{1}{|\vec{r}-\vec{r}\,'|}\,\vec{\nabla}'\,\phi(\vec{r}\,')\Big]\cdot d\vec{S}'\,. \end{aligned} \quad (2.18)$$

Note from (1.77) that we shall have

$$\nabla'^2\,\frac{1}{|\vec{r}-\vec{r}\,'|} = -4\pi\delta^3(\vec{r}\,'-\vec{r}\,)\,. \quad (2.19)$$

(Both the function $|\vec{r}-\vec{r}\,'|^{-1}$ and the delta function $\delta^3(\vec{r}-\vec{r}\,')$ are *symmetric* under the exchange of $\vec{r}$ and $\vec{r}\,'$.) This can be used on the left-hand side of (2.18), and we also use $\nabla'^2\phi(\vec{r}\,') = -4\pi\rho(\vec{r}\,')$. Thus we obtain

$$\boxed{\phi(\vec{r}\,) = \int_V \frac{\rho(\vec{r}\,')}{|\vec{r}-\vec{r}\,'|}\,d^3\vec{r}\,' + \frac{1}{4\pi}\int_S \Big[\frac{1}{|\vec{r}-\vec{r}\,'|}\,\vec{\nabla}'\phi(\vec{r}\,') - \phi(\vec{r}\,')\,\vec{\nabla}'\Big(\frac{1}{|\vec{r}-\vec{r}\,'|}\Big)\Big]\cdot d\vec{S}'\,, \quad (2.20)}$$

where the term with the delta function that arose upon using (2.19) has been used in order to perform the volume integration on that term, thus giving the $\phi(\vec{r}\,)$ after having divided everything by (-4π). The first term on the right-hand side of (2.20) is of the same form as the expression (2.13) that held in free space. The surface integrals in (2.20) represent the contribution from charge distributions on the boundary S that we are now including.

Rewriting eqn (2.20) in terms of the normal derivative introduced in eqn (2.1) (now placing a prime on the unit normal vector because the surface integral is over the primed

area element $d\vec{S}\,'$), we have

$$\phi(\vec{r}) = \int_V \frac{\rho(\vec{r}\,')}{|\vec{r}-\vec{r}\,'|}\, d^3\vec{r}\,' + \frac{1}{4\pi}\int_S \Big[\frac{1}{|\vec{r}-\vec{r}\,'|}\,\frac{\partial\phi(\vec{r}\,')}{\partial n'} - \phi(\vec{r}\,')\,\frac{\partial}{\partial n'}\Big(\frac{1}{|\vec{r}-\vec{r}\,'|}\Big)\Big]\, dS'\,, \quad (2.21)$$

We can interpret (2.20) or (2.21) as giving the expression for the potential everywhere in V in terms of the charge density ρ in V and the potential (and its normal derivative) on S. However, we cannot view (2.20) in itself as providing the answer we are seeking for how to solve for the potential in a general electrostatics problem. It can be seen from (2.21) that we would need to feed in the information about ϕ on the boundary *and also* about $\partial\phi/\partial n$ on the boundary in order to obtain the expression for ϕ in V. But we saw in the discussion of the uniqueness theorem that we are not allowed to specify independently the values of ϕ and also of its normal derivative on the boundary; that would give an overdetermined problem that admitted no solution.

Thus we can only regard (2.21) as an integral equation which will tell us what ϕ is everywhere, once we know what it and its normal derivative are on the boundary. To solve the general boundary-value problem we will need to introduce another tool, which is called the *Green function*.

2.3 Green functions and the boundary-value problem

Although (2.20) solves the desired Poisson equation (1.69) with the desired boundary conditions, it is unsatisfactory as a solution of the boundary-value problem because one would have to know both ϕ and its normal derivative on the boundary, whereas in fact these are not independent pieces of information and so they cannot be independently specified. The difficulty would be overcome if we could somehow arrange that only one of the two surface-integral terms in (2.20) were present. This can be achieved by changing the choice for the function ψ in (2.17) that we inserted into Green's theorem (2.15) in order to obtain (2.20). Instead of taking ψ to be simply given by (2.17), we need to be a little more ingenious. The function we require is known as a *Green function*.[14]

The key point about the function $\psi = |\vec{r}-\vec{r}\,'|^{-1}$ that we needed in deriving the result (2.20) was that it satisfied (2.19). In fact there is a much broader class of functions that

[14] One sometimes sees this referred to as "a Green's function." This is a grammatical abomination, which is as wrong as calling an apple pie "an apple's pie," or speaking of "a Bessel's function" or "a Laplace's transform." There are numerous offenders in the physics community. J.D. Jackson used to be one of them, but although he referred to "a Green's function" in the first edition of *Classical Electrodynamics*, he had reformed by the time the second edition appeared.

satisfy equation (2.19). This is an inhomogeneous equation with the delta function providing a source on the right-hand side, and so we can add to the solution (2.17) an arbitrary solution of the homogeneous equation. Thus we may take ψ in (2.15) to be any function of the form

$$G(\vec{r},\vec{r}\,') = \frac{1}{|\vec{r}-\vec{r}\,'|} + F(\vec{r},\vec{r}\,')\,, \tag{2.22}$$

where $F(\vec{r},\vec{r}\,')$ is any solution of the homogeneous equation

$$\nabla'^2 F(\vec{r},\vec{r}\,') = 0\,. \tag{2.23}$$

(i.e. $F(\vec{r},\vec{r}\,')$ is an *harmonic* function.) Thus we have

$$\nabla'^2 G(\vec{r},\vec{r}\,') = -4\pi\delta^3(\vec{r}-\vec{r}\,')\,. \tag{2.24}$$

The idea now is that we will *choose* $F(\vec{r},\vec{r}\,')$ so that $G(\vec{r},\vec{r}\,')$, which is called a *Green function*, satisfies appropriate boundary conditions.

To see how this works, we first note that there is an analogous result to (2.20) where we take $\psi(\vec{r}\,')$ in (2.16) to be $G(\vec{r},\vec{r}\,')$ rather than $|\vec{r}-\vec{r}\,'|^{-1}$, namely

$$\begin{aligned}&\int_V \Big[\phi(\vec{r}\,')\,\nabla'^2\,G(\vec{r}\,\vec{r}\,') - G(\vec{r},\vec{r}\,')\,\nabla'^2\,\phi(\vec{r}\,')\Big]\,d^3\vec{r}\,'\\ &= \int_S \Big[\phi(\vec{r}\,')\vec{\nabla}'\,G(\vec{r},\vec{r}\,') - G(\vec{r},\vec{r}\,')\,\vec{\nabla}'\,\phi(\vec{r}\,')\Big]\cdot d\vec{S}'\,.\end{aligned} \tag{2.25}$$

Using (2.24) together with $\nabla'^2\phi(\vec{r}\,') = -4\pi\rho(\vec{r}\,')$, we therefore find

$$\phi(\vec{r}) = \int_V \rho(\vec{r}\,')\,G(\vec{r},\vec{r}\,')d^3\vec{r}\,' + \frac{1}{4\pi}\int_S \Big[G(\vec{r},\vec{r}\,')\,\vec{\nabla}'\phi(\vec{r}\,') - \phi(\vec{r}\,')\,\vec{\nabla}'G(\vec{r},\vec{r}\,')\Big]\cdot d\vec{S}'\,. \tag{2.26}$$

In other words, we have

$$\phi(\vec{r}) = \int_V \rho(\vec{r}\,')\,G(\vec{r},\vec{r}\,')d^3\vec{r}\,' + \frac{1}{4\pi}\int_S \left[G(\vec{r},\vec{r}\,')\,\frac{\partial\phi(\vec{r}\,')}{\partial n'} - \phi(\vec{r}\,')\,\frac{\partial G(\vec{r},\vec{r}\,')}{\partial n'}\right] dS'\,. \tag{2.27}$$

Consider first the case where we wish to specify Dirichlet boundary conditions for the potential on the surface S. We achieve this by choosing the harmonic function $F(\vec{r},\vec{r}\,')$ in (2.22) so that $G(\vec{r},\vec{r}\,')$ vanishes when $\vec{r}\,'$ lies in the surface S. Thus, denoting this *Dirichlet Green function* by $G_D(\vec{r},\vec{r}\,')$, we have

$$G_D(\vec{r},\vec{r}\,') = 0 \qquad \text{when} \quad \vec{r}\,' \in S\,. \tag{2.28}$$

Using $G_D(\vec{r},\vec{r}\,')$ in (2.27) we therefore obtain

$$\boxed{\phi(\vec{r}) = \int_V \rho(\vec{r}\,')\,G_D(\vec{r},\vec{r}\,')d^3\vec{r}\,' - \frac{1}{4\pi}\int_S \phi(\vec{r}\,')\,\frac{\partial G_D(\vec{r},\vec{r}\,')}{\partial n'}\,dS'\,.} \tag{2.29}$$

This has achieved the goal of giving an expression for $\phi(\vec{r})$ everywhere in the volume V, expressed in terms of the given charge density $\rho(\vec{r})$ and the values of $\phi(\vec{r})$ on the boundary surface S. Thus, we may say that the Dirichlet boundary-value problem is solved, albeit somewhat formally.

One might worry that (2.29) has done little more than replace one difficult problem (solving $\nabla^2\phi = 4\pi\rho$ for ϕ) by another equivalently difficult problem (solving for the harmonic function $F(\vec{r}, \vec{r}\,')$ that is needed in order to ensure the Green function satisfies the all-important boundary condition (2.28)). However, this is not quite true, and moreover, a very important advantage of adopting this Green-function approach is that solving just once for the Green function for the given geometry then allows us to solve many different boundary-value problems with different choices for the charge density $\rho(\vec{r})$ and different choices for the boundary value of the potential $\phi(\vec{r})$.

The point is the following. Having once solved for $G_D(\vec{r}, \vec{r}\,')$ in the given geometry (i.e. for the specified boundary surface S), one can now construct the solution for $\phi(\vec{r})$ for *any* choice of charge density $\rho(\vec{r})$ and for *any* choice of the boundary-value potential on the surface S. Thus finding the Dirichlet Green function just once for the chosen geometry allows us to solve *any* Dirichlet boundary-value problem for that geometry.

The solution to the Neumann problem goes rather similarly, although with a minor subtlety. One might think that now one should choose $F(\vec{r}, \vec{r}\,')$ in (2.22) so that the normal derivative of the Green function vanished on S,

$$\frac{\partial G_N(\vec{r}, \vec{r}\,')}{\partial n'} = 0 \qquad \text{when} \quad \vec{r}\,' \in S\,. \tag{2.30}$$

But this would in general lead to a contradiction, since integrating (2.24) over V and using the divergence theorem (1.22) gives

$$\int_S \vec{\nabla}' G_N(\vec{r}, \vec{r}\,') \cdot d\vec{S}' = -4\pi\,, \qquad \text{i.e.} \qquad \int_S \frac{\partial G_N(\vec{r}, \vec{r}\,')}{\partial n'}\, dS' = -4\pi\,, \tag{2.31}$$

and so we cannot impose the boundary condition (2.30) on S. The simplest way around this is to impose instead

$$\frac{\partial G_N(\vec{r}, \vec{r}\,')}{\partial n'} = -\frac{4\pi}{A} \qquad \text{when} \quad \vec{r}\,' \in S\,, \tag{2.32}$$

where $A = \int_S dS$ is the area of the boundary S.[15]

[15]Note that any other choice for $\frac{\partial G_N(\vec{r}, \vec{r}\,')}{\partial n'}$ when $\vec{r}\,'$ lies in S would be equally acceptable, provided that it gives the same result -4π when integrated over the whole of S.

Substituting (2.32) into (2.27), we therefore find

$$\boxed{\phi(\vec{r}\,) = \langle\phi\rangle_S + \int_V \rho(\vec{r}\,')\, G_N(\vec{r},\vec{r}\,')d^3\vec{r}\,' + \frac{1}{4\pi}\int_S G_N(\vec{r},\vec{r}\,')\,\frac{\partial\phi(\vec{r}\,')}{\partial n'}\,dS'\,,} \tag{2.33}$$

where $\langle\phi\rangle_S$ denotes the average value of ϕ over the surface S,

$$\langle\phi\rangle_S = \frac{1}{A}\int_S \phi\, dS\,. \tag{2.34}$$

Note that $\langle\phi\rangle$ is just a pure constant, so it is not in any case important as far as the calculation of physical quantities such as the electric field $\vec{E} = -\vec{\nabla}\phi$ is concerned. It is just a manifestation of the usual minor inconvenience that the additive constant part of the potential is not pinned down when Neumann boundary conditions are imposed. If another choice, different from (2.32), had been made for the values of $\frac{\partial G_N(\vec{r},\vec{r}\,')}{\partial n'}$ when $\vec{r}\,'$ lies in S (still obeying the requirement that (2.31) hold), the final result would be that $\langle\phi\rangle_S$ in (2.33) would be replaced by a different constant, corresponding to some differently-weighted averaging of ϕ over S. This would again be of no importance when calculating physical quantities like the electric field.

The solutions (2.29) or (2.33) for ϕ in terms of the Dirichlet or Neumann Green function provide at least a formal solution to the boundary-value problem. How useful they are in practice depends upon the details of the geometry of the problem. It all comes down to the question of whether one can solve explicitly for the Green function $G_D(\vec{r},\vec{r}\,')$ or $G_N(\vec{r},\vec{r}\,')$. For a boundary S of some generic type it will certainly be impossible in closed form. In certain special cases one can obtain closed-form expressions. We shall meet examples shortly where this can be done, in the case of an infinite planar boundary S, and in the case of a spherical boundary.

It is worth making a few closing remarks about the physical interpretation of the Green function. The simplest example is when

$$G(\vec{r},\vec{r}\,') = \frac{1}{|\vec{r}-\vec{r}\,'|}\,, \tag{2.35}$$

which is, as we have seen, the Green function for the Dirichlet problem where the only boundary is the "sphere at infinity." We can recognise (2.35) as being the electric potential at the point $\vec{r}\,'$ due to a unit-strength point charge at the point $\vec{r}$. In fact (2.35) is symmetrical under the exchange of $\vec{r}$ and $\vec{r}\,'$, and so it can be equivalently viewed as the potential at $\vec{r}$ due to a unit charge at $\vec{r}\,'$.

In the more general case, the Green function is of the form (2.22), where $F(\vec{r},\vec{r}\,')$ is an harmonic function. This means that $G(\vec{r},\vec{r}\,')$ again has the interpretation of being the

potential at $\vec{r}\,'$ due to a unit charge at $\vec{r}$, but now in the more complicated situation where $G(\vec{r},\vec{r}\,')$ vanishes on S (in the Dirichlet case), or its normal derivative obeys (2.31) (in the Neumann case).

One can show quite generally that in the case of Dirichlet boundary conditions, the Green function $G_D(\vec{r},\vec{r}\,')$ is necessarily symmetrical under the exchange of $\vec{r}$ and $\vec{r}\,'$:

$$G_D(\vec{r},\vec{r}\,') = G_D(\vec{r}\,',\vec{r})\,. \tag{2.36}$$

(This can be seen by using Green's theorem (2.15); see homework 2.) In the case of Neumann boundary conditions, symmetry under the exchange of $\vec{r}$ and $\vec{r}\,'$ is not automatic, but it can always be imposed.

2.4 Dirichlet and Neumann Green functions for infinite planar conductor

In some cases, if the boundary surface in a boundary value problem has particularly simple and symmetrical geometry, it is possible to solve for the Green function in closed form, by using the *Method of Images*. (Recall from the previous discussion that solving for the Green function amounts to solving for the potential due to a unit point charge in the volume V, subject to the appropriate boundary condition on the boundary surface S.)

Suppose, for example, the potential is specified to be zero on a surface S in otherwise free space, and that one wishes to calculate the potential everywhere outside (in the volume V) due to a point charge located outside the surface. If S is suitably symmetrical, it may be possible to "mock up" the same zero-potential surface by considering a totally free space, with no actual boundary surfaces anywhere, but with one or more additional *image charges* judiciously introduced in the region of the total space that lies *outside* the volume V. (i.e. the additional image charges are on the "other side" of the surface S in the actual physical problem, disconnected from the volume V.) Suppose that by introducing image charges in this way, one can arrange that the total potential due to the original charge plus the image charges is zero on the "virtual" surface S in the mocked-up situation. It is then clear, by invoking the uniqueness theorem (see section 2.1), that the potential at all points in V must be the same in the image-charge "mock-up" and in the original problem with the actual physical conducting surface.

The challenge is to figure out how to achieve the "virtual" zero-potential surface by means of image charges. In practice, there are very few cases where it can be done. We shall discuss two of them now.

The simplest example where the method of images can be employed is in the case of an

infinite planar conductor. Let us suppose, for convenience, that Cartesian axes are chosen so that the conductor lies in the plane $z = 0$. We shall take the volume V that is of physical interest to be the whole of the half-space $z > 0$. As is required for the Dirichlet Green function, the conductor is fixed at zero potential. That is, we wish to find the potential at a point $\vec{r}$ in the upper half-space $z > 0$ due to a unit-strength point charge located at a point $\vec{r}\,'$ in the upper half-space, in the case that the whole plane at $z = 0$ is a an infinite grounded planar conductor.

We can solve this problem by turning to the "mock-up" problem where the infinite grounded conductor is removed, and we instead place an image charge at the point in the lower half-plane that is located at the mirror reflection point of the actual charge in the upper half-plane. To be specific, if a point charge q is located at some point in V, then it is obvious that if an image charge $-q$ is placed "behind" the conductor at precisely the mirror-image location, then by symmetry it must be the case that the total potential of original plus image charge, now taken to be in a completely free space with no conductor at all, will vanish on the plane $z = 0$. Therefore, the potential at any point in V in the original problem with conductor will be given by the total potential in the image-charge "mock up."

To be more precise, let us suppose that the original charge q is located at

$$\vec{r}_1 = (x_1, y_1, z_1)\,, \qquad z > 0\,. \tag{2.37}$$

The image charge $-q$ will then be located at

$$\vec{r}_2 = (x_2, y_2, z_2) = (x_1, y_1, -z_1)\,. \tag{2.38}$$

Therefore, the total potential is given by

$$\begin{aligned} \phi(\vec{r}) &= \frac{q}{|\vec{r} - \vec{r}_1|} - \frac{q}{|\vec{r} - \vec{r}_2|}\,, \\ &= \frac{q}{\sqrt{(x-x_1)^2 + (y-y_1)^2 + (z-z_1)^2}} - \frac{q}{\sqrt{(x-x_1)^2 + (y-y_1)^2 + (z+z_1)^2}}\,. \end{aligned} \tag{2.39}$$

Clearly this potential indeed vanishes on the surface $z = 0$, and so therefore by the uniqueness theorem $\phi(\vec{r})$ describes the potential, at all points with $z > 0$, of the single charge q in the presence of the infinite conductor at $z = 0$.

2.4.1 Dirichlet Green function for infinite planar boundary

We are now in a position to construct the Dirichlet and Neumann Green functions for this case. Recall that the Dirichlet Green function $G_D(\vec{r}, \vec{r}\,')$ is defined to be the potential at $\vec{r}$

due to a unit strength charge at $\vec{r}\,'$, subject to the condition that $G_D(\vec{r},\vec{r}\,')$ should vanish on the boundary S. Thus we can read off from (2.39) that the Dirichlet Green function in the case of the infinite planar boundary at $z=0$ is given by

$$G_D(x,y,z;x',y',z') \;=\; \frac{1}{\sqrt{(x-x')^2+(y-y')^2+(z-z')^2}} - \frac{1}{\sqrt{(x-x')^2+(y-y')^2+(z+z')^2}}\,,$$

This is indeed of the form

$$G_D(x,y,z;x',y',z') = \frac{1}{|\vec{r}-\vec{r}\,'|} + F(\vec{r},\vec{r}\,')\,, \tag{2.40}$$

with

$$F(\vec{r},\vec{r}\,') = -\frac{1}{\sqrt{(x-x')^2+(y-y')^2+(z+z')^2}}\,. \tag{2.41}$$

Note that, as asserted in section 2.3, G_D is indeed symmetric under the exchange of $\vec{r}$ and $\vec{r}\,'$. The first term in (2.40) is a solution of the inhomogeneous equation $\nabla'^2|\vec{r}-\vec{r}\,'|^{-1} = -4\pi\delta^3(\vec{r}-\vec{r}\,')$ (and, since it is symmetric, also $\nabla^2|\vec{r}-\vec{r}\,'|^{-1} = -4\pi\delta^3(\vec{r}-\vec{r}\,')$), while the second term is a solution of the homogeneous equation $\nabla'^2 F(\vec{r},\vec{r}\,') = 0$ (and also therefore $\nabla^2 F(\vec{r},\vec{r}\,') = 0$) in the region of interest, i.e. when z and z' lie in the upper half-space.[16] As discussed in chapter 2.3, the solution $F(\vec{r},\vec{r}\,')$ is added in order to ensure that the total expression $G_D(\vec{r},\vec{r}\,') = |\vec{r}-\vec{r}\,'|^{-1} + F(\vec{r},\vec{r}\,')$ obeys the required Dirichlet boundary condition on the plane $z=0$.

To use (2.40) in order to solve Dirichlet boundary-value problems in this geometry, we just plug it into the general expression (2.29). This requires that we evaluate the normal derivative of G_D on the boundary, which in the present case means that we need

$$-\frac{\partial G_D(x,y,z;x',y',z')}{\partial z'}\Big|_{z'=0} = -\frac{2z}{[(x-x')^2+(y-y')^2+z^2]^{3/2}}\,. \tag{2.42}$$

(The normal derivative in (2.29) is directed *outwards* from the volume V, which means in this case in the *negative* direction along the z' axis.) Suppose, for simplicity, we consider the situation where there is no charge distribution ρ in the problem, and so the entire contribution to the potential $\phi(\vec{r})$ comes from the boundary contribution in (2.29). Then we find

$$\phi(x,y,z) = \frac{z}{2\pi}\int_{-\infty}^{\infty} dx' \int_{-\infty}^{\infty} dy' \, \frac{\phi(x',y',0)}{[(x-x')^2+(y-y')^2+z^2]^{3/2}}\,. \tag{2.43}$$

Thus we have constructed the solution of the source-free equation $\nabla^2\phi = -4\pi\rho = 0$, in which the boundary value of ϕ on the infinite planar surface $z'=0$ is specified.

[16] $F(\vec{r},\vec{r}\,')$ actually satisfies $\nabla^2 F(\vec{r},\vec{r}\,') = \nabla'^2 F(\vec{r}-\vec{r}\,') = 4\pi\delta(x-x')\,\delta(y-y')\,\delta(z+z')$ in general, since it is the potential due to a negative unit charge at $(x',y',-z')$ in free space. In the region of space we are interested in, namely $z \geq 0$, it is indeed therefore harmonic, obeying $\nabla'^2 F(\vec{r},\vec{r}\,') = \nabla^2 F(\vec{r}-\vec{r}\,') = 0$.

It should be stressed here that the potential ϕ appearing in (2.43) is the potential that we are wishing to solve for in the general Dirichlet boundary-value problem in the upper half-space. It has nothing whatever to do with the potential ϕ that we found in eqn (2.39), which was the potential at $\vec{r}$ due to a point charge q above the grounded plane. When one is solving potential theory problems in electrostatics one almost always uses the symbol ϕ for the potential. The potential ϕ in (2.39) was an intermediate step in arriving at the Dirichlet Green function written in (2.40). Having got that result, we have now wiped the slate clean and are free to use the symbol ϕ for a different purpose, which is what we are doing in (2.43).

2.4.2 Neumann Green function for infinite planar boundary

We can also easily construct the Neumann Green function $G_N(\vec{r}, \vec{r}\,')$ for this geometry. In this case, it is defined to be the potential at $\vec{r}$ due to a unit strength charge at $\vec{r}\,'$, subject to the condition that the normal derivative of G_N should vanish on the plane $z' = 0$.[17] This time, we can suspect that a small modification of the image-charge trick should give us the required result. Indeed this works, and all we need to do is to replace the minus sign in front of the second term in (2.40) by a plus sign, to give

$$G_N(x, y, z; x', y', z') = \frac{1}{\sqrt{(x-x')^2 + (y-y')^2 + (z-z')^2}} + \frac{1}{\sqrt{(x-x')^2 + (y-y')^2 + (z+z')^2}}\,.$$

It is easily seen that this satisfies the required condition that

$$\frac{\partial G_N(x, y, z; x', y', z')}{\partial z'}\Big|_{z'=0} = 0\,. \tag{2.44}$$

We now plug the Neumann Green function into (2.33), in order to solve the general class of boundary-value problems in which the normal derivative of ϕ is specified on the infinite planar surface $z = 0$. Suppose again, for simplicity, we consider the case where $\rho = 0$. Plugging (2.44) into (2.33) then gives

$$\phi(x, y, z) = -\frac{1}{2\pi} \int_{-\infty}^{\infty} dx' \int_{-\infty}^{\infty} dy' \, \frac{1}{\sqrt{(x-x')^2 + (y-y')^2 + z^2}} \left(\frac{\partial \phi(x', y', z')}{\partial z'}\Big|_{z'=0} \right). \tag{2.45}$$

[17] In our general discussion for the Green function for Neumann boundary conditions, we had the requirement (2.32) that the normal derivative should equal $-4\pi/A$, where A was the area of the boundary. In the present case this area is infinite, and so we can simply require that the normal derivative of G_N should vanish.

(The minus sign arises because the outward normal derivative at $z' = 0$ is $-\partial/\partial z'$.) Of course, since $\vec{E} = -\vec{\nabla}\phi$, we may write (2.46) as

$$\phi(x,y,z) = \frac{1}{2\pi}\int_{-\infty}^{\infty} dx' \int_{-\infty}^{\infty} dy' \, \frac{E_z(x',y',0)}{\sqrt{(x-x')^2+(y-y')^2+z^2}}\,. \tag{2.46}$$

2.5 Dirichlet Green function for spherical conductor

2.5.1 Method of images for spherical conductor

A slightly more subtle example where the method of images can be employed is in the case of a spherical conductor. Suppose a conducting sphere of radius a is held at zero potential, and that a charge q is placed outside the sphere, at a distance b from its origin. It turns out that this situation can be "mocked up" by considering instead entirely free space containing the original charge and also a certain charge q' placed at a certain distance c from the origin of the sphere, on the line joining the charge q and the origin.

A quick way to derive this result is as follows. Imagine that the sphere is centred on the origin of Cartesian coordinates, and that the charge q is placed at distance b along the z axis, i.e. at $(x,y,z) = (0,0,b)$. The claim is that the image charge q' should also lie on the z axis, at some point $(x,y,z) = (0,0,c)$. If this does indeed give rise to a spherical surface of radius a that has zero potential, then in particular it must be that the potential is zero at the two points $(0,0,a)$ and $(0,0,-a)$ on the sphere (the north pole and the south pole). Since these two points are aligned on the same axis as the charges, it is particularly easy to write down the conditions that the potential should be zero at each pole:

$$\textbf{S. pole}: \quad \frac{q}{b+a} + \frac{q'}{a+c} = 0\,, \qquad \textbf{N. pole}: \quad \frac{q}{b-a} + \frac{q'}{a-c} = 0\,. \tag{2.47}$$

These two conditions determine q' and c, giving

$$q' = -\frac{aq}{b}\,, \qquad c = \frac{a^2}{b}\,. \tag{2.48}$$

Observe that since $b > a$, we have $c = a(a/b) < a$. Thus, as one would expect, the image charge is *inside* the spherical surface. It remains to verify that the potential then vanishes for an arbitrary point on the sphere. The problem has rotational symmetry around the z axis, so it suffices to consider a point P at angle θ from the z-axis. If the distance from q to P is ℓ_q, and the distance from q' to P is $\ell_{q'}$, then the cosine rule gives

$$\ell_q^2 = a^2 + b^2 - 2ab\cos\theta\,, \qquad \ell_{q'}^2 = a^2 + c^2 - 2ac\cos\theta\,. \tag{2.49}$$

After using (2.48) we see that $\ell_{q'} = (a/b)\,\ell_q$ and that indeed we have

$$\frac{q}{\ell_q} + \frac{q'}{\ell_{q'}} = 0 \tag{2.50}$$

for all θ, and so the potential vanishes everywhere on the sphere $x^2 + y^2 + z^2 = a^2$.

It is useful also to give the result in a more general fashion, in which the original charge q is placed at an arbitrary point $\vec{r}_1$ located outside the sphere, rather than lying specifically on the z axis.

Clearly, if the charge q lies at $\vec{r}_1$ then the charge q' must lie at a point $\vec{r}_2$ along the same direction, and since the second relation in (2.48) can be written as $c = (a^2/b^2)\, b$, we must have

$$\vec{r}_2 = \frac{a^2}{{r_1}^2}\, \vec{r}_1\,, \qquad q' = -\frac{aq}{r_1}\,. \tag{2.51}$$

Thus, the potential at $\vec{r}$ outside the zero-potential sphere at $r = a$ due to a charge q located at $\vec{r}_1$ outside the sphere is given by

$$\phi(\vec{r}) = \frac{q}{|\vec{r} - \vec{r}_1|} - \frac{qa/r_1}{|\vec{r} - (a^2/r_1^2)\, \vec{r}_1|}\,. \tag{2.52}$$

If we define γ to be the angle between $\vec{r}$ and $\vec{r}_1$, so that $\vec{r} \cdot \vec{r}_1 = r r_1 \cos\gamma$, then (2.52) can be expressed as

$$\phi(\vec{r}) = \frac{q}{(r^2 + r_1^2 - 2 r r_1 \cos\gamma)^{1/2}} - \frac{qa}{(r^2 r_1^2 + a^4 - 2a^2\, r r_1 \cos\gamma)^{1/2}}\,. \tag{2.53}$$

Using the expression for the surface charge density on a conductor, $\sigma = \vec{n} \cdot \vec{E}/(4\pi)$ (see (1.35)), we have

$$\sigma = -\frac{1}{4\pi}\, \frac{\partial\phi}{\partial r}\Big|_{r=a} = -\frac{q\,(r_1^2 - a^2)}{4\pi a (a^2 + r_1^2 - 2a\, r_1 \cos\gamma)^{3/2}}\,. \tag{2.54}$$

If (2.54) is integrated over the area of the sphere, it gives a total charge $q' = -qa/r_1$. In other words, the total induced charge on the surface of the sphere is equal to the *image* charge. This is in accordance with Gauss's law.

Some simple generalisations:

Because of the linearity of the Maxwell equations, it is straightforward to generalise the above result in a variety of ways. For example, instead of taking the conducting sphere to be at zero potential, we could consider a situation where it is held at a non-zero potential V (relative to zero at infinity, still). All that need be done is to add another term to the potential (2.52), corresponding to the introduction of a point charge at the origin. Thus if we now take

$$\phi(\vec{r}) = \frac{q}{|\vec{r} - \vec{r}_1|} - \frac{qa/r_1}{|\vec{r} - (a^2/r_1^2)\, \vec{r}_1|} + \frac{Q}{r}\,, \tag{2.55}$$

then the potential on the surface of the sphere becomes $\phi = Q/a$. Choosing $Q = Va$ therefore gives the required result.

As another generalisation, we can calculate the solution for a grounded sphere placed in a previously-uniform electric field.[18] Without loss of generality, let us take the electric field to be directed along the z axis. The uniform field can be achieved via a limiting process in which two point charges $\pm Q$ are placed at $z = \mp b$ respectively. Close to the origin, there will therefore be an approximately uniform electric field $E_0 \approx 2Q/b^2$ directed along z. Eventually, we take b to infinity, whilst holding $E_0 = 2Q/b^2$ fixed, and the approximation becomes exact.

In the presence of the grounded sphere, each of the above charges will have its image charge, with $+Q$ at $-b$ having an image charge $-Qa/b$ at $z = a^2/b$, and $-Q$ at $+b$ having an image charge $+Qa/b$ at $z = -a^2/b$. If we use spherical polar coordinates to write

$$\vec{r} = (x, y, z) = (r\sin\theta\cos\varphi, r\sin\theta\sin\varphi, r\cos\theta)\,, \tag{2.56}$$

then from (2.53) we deduce that the total potential for the system we are considering will be

$$\begin{aligned}\phi(r,\theta,\varphi) \;\; = \;\; & \frac{Q}{(r^2+b^2+2rb\cos\theta)^{1/2}} - \frac{Q}{(r^2+b^2-2rb\cos\theta)^{1/2}} \\ & - \frac{Qa/b}{(r^2+a^4/b^2+2a^2r/b\,\cos\theta)^{1/2}} + \frac{Qa/b}{(r^2+a^4/b^2-2a^2r/b\,\cos\theta)^{1/2}}\,.\end{aligned} \tag{2.57}$$

Expanding as a power series in $1/b$, we find

$$\phi = -\frac{2Q}{b^2}\Big(r - \frac{a^3}{r^2}\Big)\,\cos\theta - \frac{Q}{b^4}\Big(r^3 - \frac{a^7}{r^4}\Big)\,(5\cos^2\theta - 3)\cos\theta + \cdots\,. \tag{2.58}$$

The first term, with the $-2Q/b^2$ prefactor, remains finite and non-zero in the limit when we send b and Q to infinity while holding $E_0 = 2Q/b^2$ fixed. The second term has a prefactor $-Q/b^4$, which is equal to $E_0/(2b^2$, and so this goes to zero as b goes to infinity. Similarly, all the higher terms represented by the ellipses will vanish as b is sent to infinity. In this limit, we therefore find that

$$\phi = -E_0\Big(r - \frac{a^3}{r^2}\Big)\,\cos\theta\,. \tag{2.59}$$

The first term in (2.59) can be written using Cartesian coordinates as $\phi = -E_0\,z$, and so it just describes the purely uniform electric field $\vec{E} = -\vec{\nabla}\phi = (0, 0, E_0)$ that would occur in the absence of the grounded sphere. The second term describes an electric dipole contribution to the potential, arising from the two pairs of charges plus images.

[18] A more mathematically precise statement is an electric field that is asymptotically uniform at large distance.

If we did not perform the limiting procedure, but simply kept b and Q finite, the higher-order terms would remain in (2.58). These would correspond to higher terms in an expansion of the field in multipole moments, with the second displayed term in (2.58) being associated with an octopole. We shall discuss these kinds of expansions in a systematic way later in the course.

2.5.2 Dirichlet Green function for spherical boundary

We can use the results in section 2.5 to construct the Dirichlet Green function for the boundary-value problem where the potential is specified on the surface of a sphere. We just need to set $q = 1$ and $\vec{r}_1 = \vec{r}\,'$ in (2.52), leading to

$$G_D(\vec{r}, \vec{r}\,') = \frac{1}{|\vec{r} - \vec{r}\,'|} - \frac{a/r'}{|\vec{r} - (a^2/r'^2)\,\vec{r}\,'|}\,. \tag{2.60}$$

As in the previous planar example, here the first term is a solution of the inhomogeneous equation $\nabla^2|\vec{r} - \vec{r}\,'|^{-1} = -4\pi\delta^3(\vec{r} - \vec{r}\,')$, while the second term satisfies the homogeneous equation $\nabla^2 F(\vec{r}, \vec{r}\,') = 0$ in the region of interest (outside the sphere), and is added in order to ensure that $G_D(\vec{r}, \vec{r}\,')$ satisfies the Dirichlet boundary condition on the sphere of radius a.

If we introduce γ as the angle between $\vec{r}$ and $\vec{r}\,'$, then (2.60) can be written as

$$G_D(\vec{r}, \vec{r}\,') = \frac{1}{\sqrt{r^2 + r'^2 - 2rr'\,\cos\gamma}} - \frac{1}{\sqrt{a^2 + r^2\,r'^2/a^2 - 2rr'\,\cos\gamma}}\,. \tag{2.61}$$

Written in this form, it is manifest that $G_D(\vec{r}, \vec{r}\,')$ is symmetric under the exchange of $\vec{r}$ and $\vec{r}\,'$. It is also manifest that $G_D(\vec{r}, \vec{r}\,')$ vanishes, as it should, if $\vec{r}$ or $\vec{r}\,'$ lies on the surface of the sphere.

To use this expression in the general boundary-value integral in (2.29), we need to calculate the normal derivative with respect to $\vec{r}\,'$, evaluated on the sphere at $r' = a$. Bearing in mind that the *outward* normal from the volume V (external to the sphere) is directed *inwards* towards the centre of the sphere, we therefore need

$$\frac{\partial G_D(\vec{r}, \vec{r}\,')}{\partial n'}\Big|_{r'=a} = -\frac{\partial G_D(\vec{r}, \vec{r}\,')}{\partial r'}\Big|_{r'=a} = -\frac{r^2 - a^2}{a\,[r^2 + a^2 - 2ar\cos\gamma]^{3/2}}\,. \tag{2.62}$$

Substituting into (2.29) (and taking the charge density $\rho = 0$ for simplicity), we obtain

$$\begin{aligned}
\phi(\vec{r}) &= \frac{a(r^2 - a^2)}{4\pi} \int \frac{\phi(a, \theta', \varphi')}{[r^2 + a^2 - 2ar\cos\gamma]^{3/2}}\, d\Omega'\,, \\
&= \frac{a(r^2 - a^2)}{4\pi} \int_0^\pi \sin\theta'\, d\theta' \int_0^{2\pi} d\varphi'\, \frac{\phi(a, \theta', \varphi')}{[r^2 + a^2 - 2ar\cos\gamma]^{3/2}}\,,
\end{aligned} \tag{2.63}$$

where we are using spherical polar coordinates (r', θ', φ'). The area element on the sphere of radius a is written as $\vec{n} \cdot d\vec{S}' = a^2 d\Omega'$, where $d\Omega' = \sin\theta'\, d\theta' d\varphi'$ is the area element on the unit sphere (i.e. the solid angle element). The expression (2.63) gives the result for the potential everywhere outside a spherical surface of radius a, on which the potential is specified to be $\phi(a, \theta', \varphi')$.

Note that the integration in (2.63) is actually rather complicated, even if $\phi(a, \theta', \varphi')$ itself is a simple function, because of the $\cos\gamma$ appearing in the denominator. Using spherical polar coordinates, the Cartesian components of $\vec{r}$ and $\vec{r}'$ are

$$\vec{r} = (r\sin\theta\cos\varphi, r\sin\theta\sin\varphi, r\cos\theta)\,, \qquad \vec{r}' = (r'\sin\theta'\cos\varphi', r'\sin\theta'\sin\varphi', r'\cos\theta')\,, \tag{2.64}$$

and so $\cos\gamma$, which is defined by $\vec{r}\cdot\vec{r}' = rr'\cos\gamma$, is given by

$$\cos\gamma = \sin\theta\sin\theta'\cos(\varphi - \varphi') + \cos\theta\cos\theta'\,. \tag{2.65}$$

Consider, as an example, the case where one hemisphere of the boundary surface is held at a constant potential V, while the other hemisphere is held at potential $-V$. Since we are using standard spherical polar coordinates, it is natural to orient things so that the two hemispheres correspond to the parts of the sphere with $z > 0$ and $z < 0$ respectively. In other words, we have

$$\begin{aligned} \phi(a,\theta,\varphi) &= +V \qquad \text{for} \quad 0 \le \theta < \frac{\pi}{2}\,, \\ \phi(a,\theta,\varphi) &= -V \qquad \text{for} \quad \frac{\pi}{2} < \theta \le \pi\,. \end{aligned} \tag{2.66}$$

Equation (2.63) therefore gives

$$\begin{aligned} \phi(\vec{r}) &= \frac{aV\,(r^2-a^2)}{4\pi}\int_0^{2\pi} d\varphi' \Big[\int_0^{\pi/2} \frac{\sin\theta'}{(r^2+a^2-2ar\cos\gamma)^{3/2}}\, d\theta' \\ &\qquad - \int_{\pi/2}^{\pi} \frac{\sin\theta'}{(r^2+a^2-2ar\cos\gamma)^{3/2}}\, d\theta'\Big]\,, \end{aligned} \tag{2.67}$$

where $\cos\gamma$ is given by (2.65). By making the change of variables $\theta' \to \pi - \theta'$ and $\varphi' \to \varphi' + \pi$ in the second integral, this can be written as

$$\phi(\vec{r}) = \frac{aV\,(r^2-a^2)}{4\pi}\int_0^{2\pi} d\varphi' \int_0^{\pi/2} \Big[\frac{\sin\theta'}{(r^2+a^2-2ar\cos\gamma)^{3/2}} - \frac{\sin\theta'}{(r^2+a^2+2ar\cos\gamma)^{3/2}}\Big]\, d\theta' \tag{2.68}$$

Unfortunately, the integrations are too complicated to admit a useful explicit closed-form result.[19]

[19] This illustrates an important point, that although we may say that the boundary-value problem for the spherical boundary has been "solved" once we obtained an explicit closed-form result for the Green function, it does not necessarily mean that we can present an explicit closed-form expression for the solution.

We can easily integrate (2.68) if we ask for the potential ϕ only in the special case where we are on the positive z axis, i.e. for $\theta = 0$. It then follows from (2.65) that $\cos\gamma = \cos\theta'$, and then elementary integration of (2.68) gives, for $z > a$,

$$\phi(z) = V\left(1 - \frac{z^2 - a^2}{z\sqrt{z^2 + a^2}}\right) . \tag{2.69}$$

In the absence of a closed-form expression for the general off-axis potential, one could resort to making a power-series expansion of the integrand in (2.68) in powers of $\cos\gamma$, and then performing the integrations term by term. This is a somewhat clumsy approach, and so instead we shall postpone further discussion of this example until a little later in the course, when we shall have developed an approach which will allow us to obtain the power series expression for the off-axis potential very easily.

3 Solution of Laplace's Equation in Cartesian Coordinates

3.1 Introduction

The boundary-value problem in electrostatics is formulated as the problem of solving Poisson's equation $\nabla^2\phi = -4\phi\rho$ in a volume V bounded by a surface S on which appropriate boundary conditions are imposed. Quite commonly, we are interested in the situation where $\rho = 0$ in V, so that the potential ϕ in V is governed entirely by the boundary conditions that it, or its normal derivative, satisfy on S.

The geometry of the boundary surface S typically dictates what type of coordinate system is best adapted to the problem. For example, if S is formed by one or more planar surfaces, then Cartesian coordinates are likely to be the most convenient choice. If, on the other hand, the boundary S is spherical, then spherical polar coordinates will probably be the best choice. For a boundary of cylindrical shape, cylindrical polar coordinates will be most convenient.

All three of these coordinate systems share the special property that when using them the Laplacian operator ∇^2 is *separable*. This would not be true for some arbitrary choice of coordinate system. The defining property of a separable coordinate system is that Laplace's equation, which is itself a second-order partial differential equation, can be factored into a system of second-order *ordinary* differential equations. This is of enormous benefit when one tries to construct solutions.

We shall describe in detail the process of separation of variables in the three cases of Cartesian, spherical polar, and cylindrical polar, coordinates. In each case, the solution

of the factored ordinary differential equations requires an understanding of certain classes of special functions. In the Cartesian case, the relevant special functions are just the familiar sine and cosine trigonometric functions, or their hyperbolic cousins sinh and cosh (or, equivalently, real exponential functions). In the case of spherical polar coordinates, the Legendre and associated Legendre functions arise, whilst in the case of cylindrical polar coordinates it is Bessel functions that arise. We shall also look briefly at a fourth example, involving the use of oblate spheroidal coordinates, where again one can separate variables and in fact it allows one to obtain an elegant solution to a nice electrostatics problem.

We begin in this section with a review of the separation of variables for the Laplace equation in Cartesian coordinates.

3.2 Separation of variables in Cartesian coordinates

The Laplace equation in Cartesian coordinates is simply

$$\frac{\partial^2\phi}{\partial x^2} + \frac{\partial^2\phi}{\partial y^2} + \frac{\partial^2\phi}{\partial z^2} = 0\,. \tag{3.1}$$

The separation of variables is achieved by first looking for solutions where ϕ has the factorised form[20]

$$\phi(x, y, z) = X(x)\, Y(y)\, Z(z)\,. \tag{3.2}$$

Substituting this into (3.1), and dividing out by ϕ, yields

$$\frac{1}{X(x)}\,\frac{d^2X(x)}{dx^2} + \frac{1}{Y(y)}\,\frac{d^2Y(y)}{dy^2} + \frac{1}{Z(z)}\,\frac{d^2Z(z)}{dz^2} = 0\,. \tag{3.3}$$

The first term is independent of y and z, the second is independent of x and z, and the third is independent of x and y. It therefore follows that each term must separately be a constant, with the three constants summing to zero. Therefore either two of the constants are positive with the third negative, or two are negative with the third positive. Let us for now take the constants in the first two terms to be negative, and the last to be positive, so that we may write

$$\frac{d^2X}{dx^2} + \alpha^2 X = 0\,, \qquad \frac{d^2Y}{dy^2} + \beta^2 Y = 0\,, \qquad \frac{d^2Z}{dz^2} - \gamma^2 Z = 0\,, \tag{3.4}$$

with α, β and γ being real constants subject to

$$\gamma^2 = \alpha^2 + \beta^2\,. \tag{3.5}$$

[20]The general solution of (3.1) does not have this factorised form; the important point, though, is that the general solution can be written as a linear combination of factorised solutions.

The solutions for X, Y and Z will therefore be of the forms

$$X = A_1 \sin\alpha x + B_1 \cos\alpha x\,, \qquad Y = A_2 \sin\beta y + B_2 \cos\beta y\,, \qquad Z = A_3 \sinh\gamma z + B_3 \cosh\gamma z\,. \tag{3.6}$$

Equivalently, the solutions for X and Y can be taken to be linear combinations of $e^{\pm i\alpha x}$ and $e^{\pm i\beta y}$ respectively, while Z can be written in terms of $e^{\pm\gamma z}$.

Note that in the above separation of variables we introduced the three constants α, β and γ, and these are subject to the constaint $\gamma^2 = \alpha^2 + \beta^2$. Thus there are two independent *separation constants*, which we may take to be, for example, α and β. In fact the process of separating variables in three dimensions will *always* involve the introduction of two independent separation constants. The analogous separation of variables in n dimensions will always involve the introduction of $(n-1)$ independent separation constants.

The general solution to (3.1) can now be written as a sum over all the factorised solutions of the form (3.2) that we have now constructed. Since α and β are at this stage arbitrary constants, we could write the general solution in the form of a sum over all possible factorised solutions,

$$\phi(x,y,z) = \int_{-\infty}^{\infty} d\alpha \int_{-\infty}^{\infty} d\beta e^{i(\alpha x+\beta y)} \left(a(\alpha,\beta)\, e^{\sqrt{\alpha^2+\beta^2}\, z} + b(\alpha,\beta)\, e^{-\sqrt{\alpha^2+\beta^2}\, z} \right), \tag{3.7}$$

where $a(\alpha,\beta)$ and $b(\alpha,\beta)$ are arbitrary functions of the separation constants α and β.

The general solution to (3.1) is expressed as an integral over a continuum of the basic factorised solutions, as in (3.7). This is not usually, of itself, a particularly useful thing to do. However, as soon as we also impose boundary conditions on the solution, the continuous integrals will be reduced to a discrete sum over factorised solutions.

Example: A rectangular hollow box:

Suppose, for example, we wish to solve Laplace's equation inside a hollow rectangular box, with sides of length a, b and c in the x, y and z directions respectively. We may set up the axes so that the origin is at one corner of the box, so that the faces are located at $x = 0$ and $x = a$; at $y = 0$ and $y = b$; and at $z = 0$ and $z = c$. We must then impose boundary conditions on the surace that forms the boundary of the interior of the box, i.e. on the six faces of box. We shall take a simple example where we impose Dirichlet boundary conditions of the following form: Suppose that the faces are all held at zero potential, except for the face at $z = c$, on which the potential is specified to be

$$\phi(x,y,c) = V(x,y)\,, \tag{3.8}$$

for some specified voltage profile function $V(x,y)$.

Since the potential vanishes at $x = 0$ for all y and z, it follows that we must arrange for $X(x)$ to vanish at $x = 0$. Since the general solution for $X(x)$ is

$$X(x) = A_1 \sin \alpha x + B_1 \cos \alpha x \,, \tag{3.9}$$

this means we must have $B_1 = 0$, and so $X(x) = A_1 \sin \alpha x$. The potential also vanishes at $x = a$ for all y and z, and this means that we must have $X(a) = 0$. This implies that α must be restricted to take only a discrete (but infinite) set of values,

$$\alpha = \frac{m\pi}{a} \,, \tag{3.10}$$

where m is any integer. Without loss of generality we may assume that m is a *positive* integer, since the negative values will just reproduce the same set of functions (multiplied by -1), and the case $m = 0$ just gives $X = 0$..

In the same way, the vanishing of ϕ at $y = 0$ and $y = b$ implies that $Y(y)$ must be proportional to $\sin \beta y$, and that β must be of the form

$$\beta = \frac{n\pi}{b} \,, \tag{3.11}$$

where n is any positive integer.

The vanishing of ϕ at $z = 0$ implies that $B_3 = 0$ and so $Z(z)$ is proportional to $\sinh \gamma z$. Since γ is given in terms of α and β by (3.5), it follows that the general solution for ϕ that satisfies all the boundary conditions except the one on the remaining face at $z = c$ can be written as

$$\phi(x, y, z) = \sum_{m\geq 1} \sum_{n\geq 1} A_{mn} \, \sin \frac{m\pi x}{a} \, \sin \frac{n\pi y}{b} \, \sinh \left(\pi z \sqrt{\frac{m^2}{a^2} + \frac{n^2}{b^2}} \right) \,, \tag{3.12}$$

where A_{mn} are arbitrary constants.

The constants A_{mn} are determined by matching ϕ to the given boundary condition (3.8) at $z = c$. This amounts to constructing a two-dimensional Fourier series expansion for the function $V(x, y)$. Recall that we had obtained the expression (3.12) for $\phi(x, y, z)$ everywhere inside the box, expressed as a double summation. It remains for us to determine the expansion coefficients A_{mn}, by matching $\phi(x, y, z)$ to the given boundary potential $V(x, y)$ at $z = c$. In other words, we must find A_{mn} such that

$$V(x, y) = \sum_{m\geq 1} \sum_{n\geq 1} A_{mn} \, \sin \frac{m\pi x}{a} \, \sin \frac{n\pi y}{b} \, \sinh \left(\pi c \sqrt{\frac{m^2}{a^2} + \frac{n^2}{b^2}} \right) \,, \tag{3.13}$$

This amounts to solving for the coefficients a_{mn} such that

$$V(x, y) = \sum_{m\geq 1} \sum_{n\geq 1} a_{mn} \sin \frac{m\pi x}{a} \, \sin \frac{n\pi y}{b} \,, \tag{3.14}$$

where we have defined a_{mn} by

$$A_{mn} = \frac{a_{mn}}{\sinh\left(\pi c\sqrt{\frac{m^2}{a^2}+\frac{n^2}{b^2}}\right)}\,. \tag{3.15}$$

To determine the coefficients a_{mn} in (3.14), we recognise that this is an example of a two-dimensional expansion in terms of the complete sets of functions $\sin\frac{m\pi x}{a}$ and $\sin\frac{n\pi y}{b}$, and so to read off a_{mn} we just need to multiply by $\sin\frac{p\pi x}{a}\,\sin\frac{q\pi y}{b}$ and integrate. We just need to note that

$$\begin{aligned}\int_0^a \sin\frac{m\pi x}{a}\sin\frac{p\pi x}{a}dx &= \tfrac{1}{2}\int_0^a\Big(\cos\frac{(m-p)\pi x}{a}-(\cos\frac{(m+p)\pi x}{a}\Big)dx\,, \qquad (3.16)\\ &= \Big[\frac{a}{2(m-p)\pi}\,\sin\frac{(m-p)\pi x}{a}-\frac{a}{2(m+p)\pi}\,\sin\frac{(m+p)\pi x}{a}\Big]_0^a = 0\end{aligned}$$

when $m\neq p$, whilst when $m=p$ we have

$$\int_0^a \sin^2\frac{m\pi x}{a}dx = \int_0^a\Big(\tfrac{1}{2}-\tfrac{1}{2}\cos\frac{2m\pi x}{a}\Big)dx = \tfrac{1}{2}a-\frac{a}{4m\pi}\Big[\sin\frac{2m\pi x}{a}\Big]_0^a = \tfrac{1}{2}a\,. \tag{3.17}$$

Thus we have

$$\boxed{\int_0^a \sin\frac{m\pi x}{a}\sin\frac{p\pi x}{a}dx = \tfrac{1}{2}a\,\delta_{mp}\,,} \tag{3.18}$$

where the Kronecker delta δ_{mp} equals 1 if $m=p$ and equals 0 if $m\neq p$.

Multiplying (3.14) by $\sin\frac{p\pi x}{a}\,\sin\frac{q\pi y}{b}$ and integrating $\int_0^a dx\,\int_0^b dy$, we therefore find

$$\begin{aligned}\int_0^a dx\int_0^b dy\,V(v,y)\,\sin\frac{p\pi x}{a}\,\sin\frac{q\pi y}{b} &= \frac{ab}{4}\sum_{m\geq 1}\sum_{n\geq 1}a_{mn}\,\delta_{mp}\,\delta_{nq}\,,\\ &= \frac{ab}{4}\,a_{pq}\,, \qquad (3.19)\end{aligned}$$

where we have made use of (3.18) (and the analogous result for the integal over y). Using this, we deduce that

$$a_{mn} = \frac{4}{ab}\int_0^a dx\int_0^b dy\,V(x,y)\sin\frac{m\pi x}{a}\,\sin\frac{n\pi y}{b}\,, \tag{3.20}$$

and hence, using (3.15), we have solved for the coefficients A_{mn} in the series expansion (3.12) that gives the solution for the potential everywhere inside the rectangular box.

3.3 Two-dimensional problems and complex methods

A special case that is interesting to examine in greater detail is a situation where, using Cartesian coordinates, everything is independent of the z coordinate. (More generally,

where everything is independent of any one specific direction; but we may as well rotate our Cartesian coordinate system so that this direction is the z direction.) An example would be where one considers a cylindrical conductor of infinite length along the z direction, and asks for the potential everywhere outside it.

Thus we may consider the general situation where there is a translational symmetry along the z direction, implying that the potential $\phi(x, y, z)$ is actually just $\phi(x, y)$. Thus solving Laplace's equation $\nabla^2\phi = 0$ just reduces to solving the two-dimensional Laplace equation

$$\Big(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\Big)\,\phi(x, y) = 0\,. \tag{3.21}$$

Thus the problem of finding the potential outside the cylindrical conductor of infinite length along the z direction becomes instead the problem of finding the potential in two dimensions outside a conducting circle. There are some vey special techniques available for solving (3.21), which work only because we are in two dimensions. They are based upon the use of complex analysis.

Since in this subsection we are exclusively concerned with two-dimensional problems in the (x, y) plane, we will not need the symbol z for its usual purpose of denoting the Cartesian coordinate along the third direction. We can therefore use the symbol z to mean something completely different from what we have meant up until now, namely, we use z to mean the complex variable

$$z = x + \mathrm{i}\, y\,. \tag{3.22}$$

A general complex function f can be represented as $f(x, y)$. A *holomorphic* function is one that depends on x and y only in the special form $f = f(z)$. That is, it depends on x and y only as a function of the single complex variable z, but not $\bar{z}$. We can treat z and its complex conjugate $\bar{z}$ as if they were independent variables, with z given by (3.22) and $\bar{z}$ given by its complex conjugate,

$$\bar{z} = x - \mathrm{i}\, y\,. \tag{3.23}$$

Thus we have

$$x = \frac{z + \bar{z}}{2}\,, \qquad y = \frac{z - \bar{z}}{2\mathrm{i}}\,, \tag{3.24}$$

and so we can calculate the derivatives $\partial_z \equiv \partial/\partial z$ and $\partial_{\bar{z}} \equiv \partial/\partial\bar{z}$ by using the chain rule:

$$\begin{aligned} \partial_z &= \frac{\partial x}{\partial z}\,\partial_x + \frac{\partial y}{\partial z}\,\partial_y = \frac{1}{2}(\partial_x - \mathrm{i}\,\partial_y)\,, \\ \partial_{\bar{z}} &= \frac{\partial x}{\partial \bar{z}}\,\partial_x + \frac{\partial y}{\partial \bar{z}}\,\partial_y = \frac{1}{2}(\partial_x + \mathrm{i}\,\partial_y)\,, \end{aligned} \tag{3.25}$$

where ∂_x means $\partial/\partial x$ and ∂_y means $\partial/\partial y$. Thus the two-dimensional Laplace operator in (3.21) can be written as

$$(\partial_x^2 + \partial_y^2) = 4\partial_z\,\partial_{\bar{z}}\,. \tag{3.26}$$

A complex function $f(x,y)$ is said to be *holomorphic* (or, equivalently, *complex analytic*) if it obeys $\partial_{\bar{z}} f = 0$. If we write f in terms of its real and imaginary parts as

$$f(x,y) = \chi(x,y) + \mathrm{i}\,\sigma(x,y)\,, \tag{3.27}$$

then from the expression for $\partial_{\bar{z}}$ in (3.25), it follows that if f is holomorphic then

$$(\partial_x + \mathrm{i}\,\partial_y)(\chi + \mathrm{i}\,\sigma) = 0\,. \tag{3.28}$$

The real and imaginary parts must vanish separately, and since x, y, χ and σ are all real, we get

$$\partial_x\chi = \partial_y\sigma\,, \qquad \partial_y\chi = -\partial_x\sigma\,. \tag{3.29}$$

These are the Cauchy-Riemann equations.

Notice that if we define $\vec{\nabla}$ in this subsection to mean the two-dimensional gradient operator $\vec{\nabla} = (\partial_x, \partial_y)$, then it follows from (3.29) that

$$(\vec{\nabla}\chi)\cdot(\vec{\nabla}\sigma) = 0\,. \tag{3.30}$$

That is, for a holormorphic function f the lines of $\chi =$constant intersect the lines of $\sigma =$ constant at 90 degrees. Also, crucially, any holomorphic function satisfies $\nabla^2 f = 0$, since, as we see from eqn (3.26), $\nabla^2 f = 4\partial_z\,\partial_{\bar{z}}\,f$, and by definition $\partial_{\bar{z}}\,f = 0$ for a holomorphic function. This means that the real and imaginary parts of f each independently satisfy Laplace's equation:

$$\nabla^2\,\chi(x,y) = 0\,, \qquad \nabla^2\,\sigma(x,y) = 0\,. \tag{3.31}$$

(This can also be seen by using the Cauchy-Riemann equations in the form (3.29.)

This last observation lies at the heart of how we can use the methods of complex analysis in order to obtain solutions to the two-dimensional Laplace equation in some quite useful and non-trivial situations. The key further observation is that if we make a holomorphic change of complex variable, from z to w where

$$w = w(z)\,, \tag{3.32}$$

then if $\Psi(w)$ is a holomorphic function of w, it follows that $\psi(z) \equiv \Psi(w(z))$ will be a holomorphic function of z. Thus if we write the real and imaginary parts of w and z as

$$w = u + \mathrm{i}\,v\,, \qquad z = x + \mathrm{i}\,y\,, \tag{3.33}$$

then if we are able to construct a (real) solution $\hat{\phi}(u,v)$ of Laplace's equation in the two-dimensional space with coordinates (u,v), and if we can figure out the corresponding complex holomorphic function $\Psi(w)$ of which $\hat{\phi}(u,v)$ is the real part, then this allows us to construct the complex holomorphic function $\psi(z) = \Psi(w(z))$ in the two-dimensional space that has coordinates (x,y). The real part $\phi(x,y)$ of $\psi(z)$ will give us the solution of Laplace's equation in the two-dimensional (x,y) space that maps into the solution $\hat{\phi}(u,v)$ in the (u,v) space.

The trick now is to find a clever mapping $w = w(z)$ that maps an easily-solved potential theory problem in the (u,v) space into the more complicated problem that we really want to solve in the (x,y) space. Mappings of this sort are known as *conformal mappings*.

An example at this point is worth a thousand words. Suppose we wish to find the potential in the space between two infinitely long cylindrical conductors, each extending along the z direction (old z, the third Cartesian coordinate!). Thus this is really just a two-dimensional problem of the kind we have been discussing, in which we wish to find the two-dimensional potential between two conducting circles. This would be very easy if the the cylinders were concentric, but we want to consider the more complicated situation where the axes of the two cylinders do not coincide. The trick in this example is to find a mapping $w = w(z)$ that maps two concentric circles in the (u,v) plane into two non-concentric circles in the (x,y) plane.

The is a whole body of knowledge about conformal mappings, and how to map simple two-dimensional shapes into more complicated, but useful, two-dimensional shapes. Drawing from this knowledge, let us consider as an example the conformal mapping

$$w = \frac{z - \frac{1}{2}}{\frac{1}{2}z - 1}\,. \tag{3.34}$$

Using $z = x + \mathrm{i}\,y$ we have therefore have

$$w = \frac{2z-1}{z-2} = \frac{(2x-1) + 2\mathrm{i}\,y}{(x-2) + \mathrm{i}\,y}\,, \tag{3.35}$$

and so

$$|w|^2 = \frac{(2x-1)^2 + 4y^2}{(x-2)^2 + y^2}\,. \tag{3.36}$$

Consider the unit circle, centred on the origin, in the w plane, i.e. the circle $|w| = 1$. Using (3.36) we can work out what shape this describes in the z plane. Simple algebra shows that we shall have

$$x^2 + y^2 = 1\,. \tag{3.37}$$

In other words, the unit circle $|w| = 1$ in the w plane has mapped into the unit circle centred on the origin in the z plane, i.e. $|z| = 1$.

Now look at the circle $|w| = \frac{1}{2}$ in the w plane. Again using (3.36), simple algebra shows that this corresponds to

$$\left(x - \tfrac{2}{5}\right)^2 + y^2 = \tfrac{4}{25}\,. \tag{3.38}$$

This is a circle of radius $\frac{2}{5}$, centred on the point $(x, y) = (\frac{2}{5}, 0)$ in the (x, y) plane.

What we have established, therefore, is that the pair of concentric circles $|w| = 1$ and $|w| = \frac{1}{2}$ in the (u, v) plane has been mapped into the pair of non-concentric circles (3.37) and (3.38) in the (x, y) plane.

Note that there is nothing particularly special about the two radii we chose for the original circles in the (u, v) plane. Any circle in the (u, v) plane will map into a circle in the (x, y) plane. In general, if one considers a circle in the (u, v) plane that is centred on the origin, it will map into a circle in the (x, y) plane will be centred on some point on the x axis, displaced from the origin. One could also choose different values for the constants in the transformation (3.34) relating w and z, and this would alter the specifics of the circle in the (x, y) plane that results from a particular circle in the (u, v) plane. Our specific example is chosen just in order to illustrate the general idea.

Suppose now we consider the easy problem in the (u, v) plane, in which the circle at $|w| = \frac{1}{2}$ is held at potential V and the circle at $|w| = 1$ is held at potential 0. The problem is rotationally symmetric, and so we can easily see that the solution is

$$\hat{\phi}(u, v) = c\,\log(u^2 + v^2)^{1/2}\,, \qquad c = -\frac{V}{\log 2}\,. \tag{3.39}$$

This can be written in terms of w as

$$\hat{\phi}(u, v) = c\,\log|w|\,. \tag{3.40}$$

Now we need to find the holomorphic function $\Phi(w)$ whose real part is given by $\hat{\phi}(u, v)$.

A classic exercise in any complex analysis course is where the lecturer presents you with a real function and tells you it is the real part of a holomorphic function. You are then asked to integrate up the Cauchy-Riemann equations (3.29) in order to find the corresponding imaginary part of the holomorphic function, and then, putting this together with the real part, to write down the holomorphic function the lecturer first thought of.[21] Of course you

[21] Actually there is an easier way to find the holomorphic function $f(z)$ whose real part $\chi(x, y)$ is given, without needing to integrate anything at all. The procedure, which does not seem to be well known, is described in section 5.5 of my online lecture notes *Methods of Theoretical Physics I*, where it is referred to

will only succeed in doing this if the lecturer has played fair and given you a real function that is capable of arising as the real part of a holomorphic function. In other words, the function the lecturer first gives you must itself satisfy Laplace's equation. Note, by the way, that although the discussion in this paragraph, and in footnote 21, is about a holomorphic function $f(z) = \chi(x,y) + \mathrm{i}\sigma(x,y)$, in our specific discussion we would be applying it to a holomorphic function $\Psi(w)$ whose real part is $\hat{\phi}(u,v)$.

In our case we do indeed know that $\hat{\phi}$ satisfies Laplace's equation, so we are guaranteed to be able to carry out the above procedure. Actually, this particular example is so simple that we can see the answer almost by inspection; the holomorphic function is just

$$\Psi(w) = c \log w \,. \tag{3.42}$$

We can check this by verifying that its real part is indeed $\hat{\phi}$ in (3.40):

$$\Re(\Psi) = \tfrac{1}{2}(\Psi + \bar{\Psi}) = \frac{c}{2}\,(\log w + \log \bar{w}) = \frac{c}{2}\,\log(w\bar{w}) = \frac{c}{2}\,\log|w|^2 = c\log|w|\,. \tag{3.43}$$

(It is very easy to derive (3.42) by using the Oppenheim method described in footnote 21. This is an example where the constant a must be taken to be non-zero, since $\hat{\phi}(u,v) = c\log|w|$ is non-analytic at $w=0$.)

Now, we use the conformal mapping (3.34) in order to map the problem to the z plane, i.e. the (x,y) plane. We thus have the holomorphic function $\psi(z) = \Psi(w(z))$ and so

$$\psi(z) = c\,\log\frac{z-\frac{1}{2}}{\frac{1}{2}z-1}\,. \tag{3.44}$$

The solution to the potential between the two non-concentric circles is then given by the real part of $\psi(z)$, namely

$$\begin{aligned}\phi(x,y) &= c\,\log\Big|\frac{z-\frac{1}{2}}{\frac{1}{2}z-1}\Big| = \frac{c}{2}\,\log\Big|\frac{z-\frac{1}{2}}{\frac{1}{2}z-1}\Big|^2\,,\\ &= \frac{c}{2}\,\log\Big[\frac{(2x-1)^2+4y^2}{(x-2)^2+y^2}\Big]\,,\end{aligned} \tag{3.45}$$

---

as the Oppenheim method: If $\chi(x,y)$ is the real part of $f(z)$, then $f(z)$ itself is given by

$$f(z) = 2\chi\big(\frac{z+a}{2}, \frac{z-a}{2\mathrm{i}}\big) + b\,, \tag{3.41}$$

where a and b are constants. The constant a can be chosen almost arbitrarily; the only requirement is that $\chi(x,y)$ should be analytic at the point $z \equiv x + \mathrm{i}y = a$. (So unless $\chi(x,y)$ is non-analytic at $x=y=0$, one can simply choose $a=0$.) The real part of b is determined from the known properties of $\chi(x,y)$. The imaginary part of b is undeterminable; the same issue arises if one solves for the imaginary part $\sigma(x,y)$ by the usual procedure of integrating the Cauchy-Riemann equations, since $\sigma(x,y)$ can only be determined up to a constant of integration. The reason why (3.41) works is very simple; see my Methods lecture notes.

where, from eqns (3.39), $c = -V/\log 2$. That is, this is the potential between a circle of radius $\frac{2}{5}$, centred on $(x, y) = (\frac{2}{5}, 0)$ and held at potential V, and a circle of radius 1, centred on the origin and held at potential 0.

The calculation above illustrated just one relatively simple example of how one can use a conformal mapping to turn a tricky two-dimensional potential theory problem into an easily-solvable one, by mapping the complicated geometry of the original problem into a much simpler geometry. It is a technique that can be quite useful in a variety of instances. Anyone interested can easily delve further into the subject for themselves.

4 Separation of variables in spherical polar coordinates

The spherical polar coordinates (r, θ, φ) are related to Cartesian coordinates (x, y, z) by

$$x = r\,\sin\theta\,\cos\phi\,, \qquad y = r\,\sin\theta\,\sin\phi\,, \qquad z = r\,\cos\theta\,. \tag{4.1}$$

In terms of (r, θ, φ), Laplace's equation (3.1) becomes[22]

$$\nabla^2\,\phi = \frac{1}{r^2}\,\frac{\partial}{\partial r}\Big(r^2\,\frac{\partial\phi}{\partial r}\Big) + \frac{1}{r^2}\,\nabla^2_{(\theta,\varphi)}\,\phi = 0\,, \tag{4.2}$$

where $\nabla^2_{(\theta,\varphi)}$ is the two-dimensional Laplace operator on the surface of the unit-radius sphere,

$$\nabla^2_{(\theta,\varphi)} \equiv \frac{1}{\sin\theta}\,\frac{\partial}{\partial\theta}\Big(\sin\theta\,\frac{\partial}{\partial\theta}\Big) + \frac{1}{\sin^2\theta}\,\frac{\partial^2}{\partial\varphi^2}\,. \tag{4.3}$$

It is sometimes useful to note that the radial part of the Laplacian in (4.2) can be rewritten so that the Laplacian becomes

$$\nabla^2\phi\, = \frac{1}{r}\,\frac{\partial^2(r\phi)}{\partial r^2} + \frac{1}{r^2}\,\nabla^2_{(\theta,\varphi)}\,\phi = 0\,, \tag{4.4}$$

Laplace's equation can be separated by first looking for factorised solutions for $\phi(r, \theta, \varphi)$, of the form

$$\phi(r, \theta, \varphi) = \frac{1}{r}\,R(r)\,Y(\theta, \varphi)\,. \tag{4.5}$$

Substituting into (4.4), dividing by ϕ, and multiplying by r^2, we get

$$\frac{r^2}{R}\,\frac{d^2R}{dr^2} + \frac{1}{Y}\,\nabla^2_{(\theta,\varphi)}Y = 0\,. \tag{4.6}$$

[22] This can be seen by a brute-force calculation using the chain rule to convert from derivatives with respect to x, y and z to derivatives with respect to r, θ and φ. Thus, $\partial/\partial x = (\partial r/\partial x)\partial/\partial r + (\partial\theta/\partial x)\partial/\partial\theta + (\partial\varphi/\partial x)\partial/\partial\varphi$, etc. There are also more elegant ways to do the calculation if one is familiar with general-coordinate tensor analysis.

The last term in (4.6) depends only on θ and φ, while the first term depends only on r, and so since the first term is equal to minus the last term, it means that each of the terms must be a constant, so

$$\nabla^2_{(\theta,\varphi)}\, Y = -\lambda\, Y\,, \qquad \frac{d^2R}{dr^2} = \frac{\lambda}{r^2}\, R\,, \qquad \lambda = \text{ constant}\,. \tag{4.7}$$

The key point now is that one can show that the harmonics $Y(\theta,\varphi)$ on the sphere are well-behaved only if the separation constant λ takes a certain discrete infinity of non-negative values. The most elegant way to show this is by making use of the symmetry properties of the sphere, but since this takes us away from the main goals of the course, we shall not follow that approach here.[23] Instead, we shall follow the more "traditional," if more pedestrian, approach of examining the conditions under which singular behaviour of the eigenfunction solutions of the differential equation can be avoided.

To study the eigenvalue problem $\nabla^2_{(\theta,\phi)}\, Y = -\lambda\, Y$ in detail, we make a further separation of variables by first looking for $Y(\theta,\varphi)$ of the factorised form $Y(\theta,\varphi) = \Theta(\theta)\,\Phi(\varphi)$. Substituting this in, and multiplying by $\sin^2\theta\, Y^{-1}$, we get

$$\frac{1}{\Theta}\,\sin\theta\,\frac{d}{d\theta}\Big(\sin\theta\,\frac{d\Theta}{d\theta}\Big) + \lambda\,\sin^2\theta + \frac{1}{\Phi}\,\frac{d^2\Phi}{d\varphi^2} = 0\,. \tag{4.8}$$

By now-familiar arguments, the last term depends only on φ, while the first two depend only on θ. Consistency for all θ and φ therefore implies that the last term must be a constant, which we shall call $-m^2$, and so we have

$$\frac{d^2\Phi}{d\varphi^2} + m^2\,\Phi \;=\; 0\,, \tag{4.9}$$

$$\sin\theta\,\frac{d}{d\theta}\Big(\sin\theta\,\frac{d\Theta}{d\theta}\Big) + (\lambda\,\sin^2\theta - m^2)\,\Theta \;=\; 0\,, \qquad m^2 = \text{ constant}. \tag{4.10}$$

[23] The essential point is that the surface of the unit sphere can be defined as $x^2+y^2+z^2=1$, and this is invariant under transformations of the form

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} \longrightarrow M \begin{pmatrix} x \\ y \\ z \end{pmatrix}\,,$$

where M is any constant 3×3 orthogonal matrix, satisfying $M^T M = \mathbb{1}$. This shows that the sphere is invariant under the orthogonal group $O(3)$, and hence the eigenfunctions Y must fall into representations under $O(3)$. This group is in fact the rotation group in three dimensions (together with reflections). The calculation of the allowed values for λ, and the forms of the associated eigenfunctions Y, then follow from group-theoretic considerations. Anticipating the result that we shall see by other means, the eigenvalues λ take the form $\lambda_\ell = \ell(\ell+1)$, where ℓ is any non-negative integer. The eigenfunctions are classified by ℓ and a second integer m, with $-\ell \le m \le \ell$, and are the well-known spherical harmonics $Y_{\ell m}(\theta,\phi)$. The fact that λ depends on ℓ but not m means that the eigenvalue $\lambda_\ell = \ell(\ell+1)$ has a degeneracy $(2\ell+1)$.

The solution to the Φ equation is $\Phi \sim e^{\pm \mathrm{i}\, m\, \varphi}$. The constant m^2 could, *a priori*, be positive or negative, but we must recall that the coordinate φ is periodic on the sphere, with period 2π. The periodicity implies that the eigenfunctions Φ should be periodic too, and hence it must be that m^2 is non-negative. In order that we have $\Phi(\varphi + 2\pi) = \Phi(\varphi)$ it must furthermore be the case that m is an integer.

As a side remark, note that the statement about periodicity in the preceding paragraph assumes that we are wanting to solve $\nabla^2\phi = 0$ in a region that includes the whole 2π of rotation around the z axis. If we were instead solving the equation in a wedge-shaped region, with φ restricted to the interval $0 \le \varphi \le \alpha$, say, then we might want to impose boundary conditions such as $\phi(r, \theta, 0) = 0$ and $\phi(r, \theta, \alpha) = 0$. In such a case, we would need the azimuthal functions to be of the form $\Phi(\varphi) \sim \sin(n\pi\varphi/\alpha)$ where n is an integer, and so then the separation constant m would be of the form $m = n\pi/\alpha$. We shall not pursue such possibilities in what follows, and it will be understood that m is an integer.

4.1 Series solution of the Legendre equation

To analyse the eigenvalue equation (4.10) for Θ, it is advantageous to define a new independent variable x, related to θ by $x = \cos\theta$. (Do not confuse this variable x with the Cartesian coordinate x!) At the same time, let us now use y instead of Θ as our symbol for the dependent variable. Equation (4.10) therefore becomes

$$\frac{d}{dx}\Big((1-x^2)\,\frac{dy}{dx}\Big) + \Big(\lambda - \frac{m^2}{1-x^2}\Big)\, y = 0\,. \tag{4.11}$$

This equation is called the *Associated Legendre Equation*, and it will become necessary to study its properties, and solutions, in some detail in order to be able to construct solutions of Laplace's equation in spherical polar coordinates. In fact, as we shall see, it is convenient first to study the simpler equation when $m = 0$, which corresponds to the case where the harmonics $Y(\theta, \varphi)$ on the sphere are independent of the azimuthal angle φ. The equation (4.11) in the case $m = 0$ is called the *Legendre Equation*.

Taking $m = 0$ for now, the associated Legendre equation (4.11) reduces to the Legendre equation

$$[(1-x^2)\,y']' + \lambda\, y = 0\,., \tag{4.12}$$

which we can also write as

$$(1-x^2)\,y'' - 2x\,y' + \lambda\, y = 0\,. \tag{4.13}$$

Note that here we are denoting a derivative with respect to x by a prime, so that dy/dx is written as y', and so on. We can construct the solutions to (4.13) by applying a procedure known as the *Frobenius Method*, in which $y(x)$ is obtained as a power series in x.

The general theory of how to construct series solutions to ordinary differential equations is quite involved. However, in the present case things are very simple, because the point $x = 0$ around which we wish to expand the power series is a so-called *ordinary point* of the differential equation. (i.e. if the equation is written in the form $y'' + P(x)y' + Q(x)y = 0$ by dividing out by $(1 - x^2)$, the functions $P = -2x/(1 - x^2)$ and $Q = \lambda/(1 - x^2)$ are regular and analytic around $x = 0$.) This means that the two independent solutions to (4.13) can both be expanded in Taylor series around $x = 0$.[24]

Thus, we begin by writing the series expansion

$$y = \sum_{n\geq 0} a_n\, x^n\,. \tag{4.14}$$

Clearly we shall have

$$y' = \sum_{n\geq 0} n\, a_n\, x^{n-1}\,, \qquad y'' = \sum_{n\geq 0} n\,(n-1)\, a_n\, x^{n-2}\,. \tag{4.15}$$

Substituting into equation (4.13), we find

$$\sum_{n\geq 0} n\,(n-1)\, a_n\, x^{n-2} + \sum_{n\geq 0} \big(\lambda - n\,(n+1)\big)\, a_n\, x^n = 0\,. \tag{4.16}$$

Since we want to equate terms order by order in x, it is useful to shift the summation variable by 2 in the first term, by writing $n = m + 2$;

$$\sum_{n\geq 0} n\,(n-1)\, a_n\, x^{n-2} = \sum_{m\geq -2} (m+2)(m+1)\, a_{m+2}\, x^m = \sum_{m\geq 0} (m+2)(m+1)\, a_{m+2}\, x^m\,. \tag{4.17}$$

(The last step, where we have dropped the $m = -2$ and $m = -1$ terms in the summation, clearly follows from the fact that the $(m + 2)(m + 1)$ factor gives zero for these two values of m.) Finally, relabelling m as n again, we get from (4.16)

$$\sum_{n\geq 0} \Big[(n+2)(n+1)\, a_{n+2} + \big(\lambda - n\,(n+1)\big)\, a_n\Big]\, x^n = 0\,. \tag{4.18}$$

[24] In more general cases where one is expanding around a singular point of the equation (say at $x = b$), one needs to consider a series of the form $y(x) = \sum_{n\geq 0} a_n (x - b)^{n+\sigma}$, where σ is a constant that may be non-integer. It is rather common to see people automatically considering this more general type of expansion, even when they are expanding around an ordinary point of the equation. Although this is not actually wrong, it is rather inconvenient, since it represents an over-parameterisation of the problem and therefore it obscures the essential simplicity of the procedure. Possibly they do it because they are unaware of the theorem that both solutions of a second-order ordinary differential equation are analytic in the neighbourhood of an ordinary point.

Since this must hold for all values of x, it follows that the coefficient of each power of x must vanish separately, giving

$$(n+2)(n+1)\,a_{n+2} + \big(\lambda - n\,(n+1)\big)\,a_n = 0 \tag{4.19}$$

for all $n \geq 0$. Thus we have the recursion relation

$$\boxed{a_{n+2} = \frac{n\,(n+1) - \lambda}{(n+1)(n+2)}\,a_n\,.} \tag{4.20}$$

We see from (4.20) that all the coefficients a_n with $n \geq 2$ can be solved for, in terms of a_0 and a_1.[25] In fact, since the recursion relation involves a step of 2 (i.e. it gives a_{n+2} in terms of a_n), all the a_n for *even* n can be solved for in terms of a_0, while all the a_n for *odd* n can be solved for in terms of a_1. Since the equation is linear, we can take the even-n series and the odd-n series as the two linearly independent solutions of the Legendre equation, which we can call $y_{\rm even}(x)$ and $y_{\rm odd}(x)$:

$$\begin{aligned} y_{\rm even}(x) &= a_0 + a_2\,x^2 + a_4\,x^4 + \cdots\,, \\ y_{\rm odd}(x) &= a_1\,x + a_3\,x^3 + a_5\,x^5 + \cdots\,. \end{aligned} \tag{4.21}$$

The first solution involves only the even a_n, and thus has only even powers of x, whilst the second involves only the odd a_n, and has only odd powers of x. We can conveniently consider the two solutions separately, by taking either $a_1 = 0$, to discuss $y_{\rm even}$, or else taking $a_0 = 0$, to discuss $y_{\rm odd}$.

Starting with $y_{\rm even}$, we can take $a_0 \neq 0$ and $a_1 = 0$, so we therefore have from (4.20) that $a_2 = -\frac{1}{2}\lambda\,a_0$, $a_3 = 0$, $a_4 = \frac{1}{12}(6-\lambda)\,a_2$, $a_5 = 0$, etc. In the expression for a_4, we can substitute the expression already found for a_2, and so on. Thus we shall get

$$\begin{aligned} a_2 &= -\tfrac{1}{2}\lambda\,a_0\,, \qquad a_4 = -\tfrac{1}{24}\lambda\,(6-\lambda)\,a_0\,, \qquad \ldots \\ a_3 &= a_5 = a_7 = \cdots = 0\,. \end{aligned} \tag{4.22}$$

The series solution in this case is therefore given by

$$y_{\rm even} = a_0\left(1 - \tfrac{1}{2}\lambda\,x^2 - \tfrac{1}{24}\lambda\,(6-\lambda)\,x^4 + \cdots\right). \tag{4.23}$$

To discuss the solution $y_{\rm odd}$ instead, we can take $a_0 = 0$ and $a_1 \neq 0$. The recursion relation (4.20) now gives $a_2 = 0$, $a_3 = \frac{1}{6}(2-\lambda)\,a_1$, $a_4 = 0$, $a_5 = \frac{1}{20}(12-\lambda)\,a_3$, $a_5 = 0$, etc, and so we find

$$\begin{aligned} a_3 &= \tfrac{1}{6}(2-\lambda)\,a_1\,, \qquad a_5 = \tfrac{1}{120}(2-\lambda)\,(12-\lambda)\,a_1\,, \qquad \ldots \\ a_2 &= a_4 = a_6 = \cdots = 0\,. \end{aligned} \tag{4.24}$$

[25] Later, we shall give the explicit solution to the recursion relation (4.20).

The series solution in this case therefore has the form

$$y_{\rm odd} = a_1 \left(x + \tfrac{1}{6}(2-\lambda)\, x^3 + \tfrac{1}{120}(2-\lambda)\,(12-\lambda)\, x^5 + \cdots \right). \tag{4.25}$$

To summarise, we have produced two linearly independent solutions to our differential equation (4.13), which are given by (4.23) and (4.25). The fact that they are linearly independent is obvious, since the first is an even function of x whilst the second is an odd function.

So far in the discussion, the separation constant λ has been allowed to be completely arbitrary. As we shall now see, it must be restricted to take a discrete infinite set of values in order to have solutions that are non-singular in the coordinate range $-1 \le x \le 1$. (Recall that $x = \cos\theta$, and so x ranges from 1 down to -1 as θ ranges from $\theta = 0$ (at the north pole of the sphere) to $\theta = \pi$ (at the south pole of the sphere.) Although the solutions of the Legendre equation that we have constructed are valid for any value of λ, we shall see that λ must be restricted to a discrete set of values if one insists that the power series expansions should converge for all x in the interval $-1 \le x \le 1$.

The convergence of an infinite series can be tested by applying the *ratio test*. The statement of this test is that the series converges if the ratio of successive terms in the series is of magnitude less than 1, in the limit as one looks further and further down the series. If, on the other hand, the ratio in this limit is greater than 1, then the series diverges. If the ratio equals 1, then the test is inconclusive.

Since the two independent series we obtained correspond to the even powers of x and the odd power of x in (4.14), the ratio of successive terms in either series will be

$$\begin{aligned} R_n &= \frac{a_{n+2}\, x^{n+2}}{a_n x^n} = \frac{a_{n+2}\, x^2}{a_n}\,, \\ &= \frac{n(n+1)-\lambda}{(n+1)(n+2)}\, x^2\,. \end{aligned} \tag{4.26}$$

(We used the recursion relation (4.20) in getting to the second line.) In the limit when $n \to \infty$, holding λ fixed, we therefore have

$$R_\infty = x^2\,. \tag{4.27}$$

Thus the series (4.14) converges if $|x| < 1$ and diverges if $|x| > 1$. The test is inconclusive for $x = \pm 1$.

Recalling that $x = \cos\theta$ here, we see that the values $x = \pm 1$ are in fact attained in the physical problems we are studying, since these values correspond to $\theta = 0$ and $\theta = \pi$ (the

north and south poles of the sphere). It is therefore important to establish how the series expansion behaves at $x = \pm 1$.

We shall not present a detailed discussion of the convergence here, but we simply state the result, which is that in fact, for generic values of λ, the series (4.14) diverges at $x = \pm 1$; i.e. both solutions (even and odd) diverge. (We shall illustrate this in one example, below.) Thus, in order to obtain regular solutions for the Legendre equation, we must instead arrange, by judicious choice of the values for λ, to make the series terminate.

Looking at the recursion relation (4.20), it is evident that we can arrange for a termination if λ is chosen to be given by

$$\lambda = \ell(\ell+1)\,, \qquad \ell = \text{integer} \geq 0\,. \tag{4.28}$$

Then, we shall have that $a_{\ell+2}$ is equal to zero, and then all the higher coefficients of the form $a_{\ell+4}$, $a_{\ell+6}$, etc., will also vanish. Thus, we obtain a polynomial solution, of degree ℓ, when λ satisfies (4.28). Obviously, since it is a sum of a finite numer of terms, the polynomial solution is non-singular for all x in the interval $-1 \leq x \leq 1$.

Note that if ℓ is an even integer then it is the *even* series that terminates to give a finite polynomial solution, but the odd series does not terminate. Conversely, if ℓ is an odd integer then it is the odd series that terminates, while the even series does not. Thus we only ever get one terminating polynomial solution, for each integer ℓ.

As an example to illustrate the divergent behaviour if the series does not terminate, consider the odd series $y_{\rm odd}(x)$, with $\ell = 0$, i.e. $\lambda = 0$. From (4.20) we then have

$$a_{n+2} = \frac{n\,a_n}{n+2}\,, \tag{4.29}$$

(with n odd), which has the solution $a_n = a_1/n$. Thus the series (4.14) becomes

$$y = a_1\,(x + \tfrac{1}{3}x^3 + \tfrac{1}{5}x^5 + \tfrac{1}{7}x^7 + \cdots) = a_1 \sum_{p\geq 0} \frac{x^{2p+1}}{2p+1}\,, \tag{4.30}$$

which can be recognised as the power-series expansion of

$$y = \tfrac{1}{2}a_1 \log\Big(\frac{1+x}{1-x}\Big)\,. \tag{4.31}$$

This function clearly diverges logarithmically at $x = \pm 1$. For all other values of λ that lead to non-terminating series, one similarly finds a logarithmic divergence at $x = \pm 1$.

To summarise, we have established that if λ is given by (4.28), there exists one polynomial solution to the Legendre equation, and it is in particular regular for all $-1 \leq x \leq 1$. This solution is called the *Legendre Polynomial* $P_\ell(x)$, satisfying

$$(1-x^2)\,\frac{d^2P_\ell(x)}{dx^2} - 2x\,\frac{dP_\ell(x)}{dx} + \ell(\ell+1)\,P_\ell(x) = 0\,. \tag{4.32}$$

By convention, the Legendre polynomial $P_\ell(x)$ is normalised so that

$$P_\ell(1) = 1\,. \tag{4.33}$$

The first few are therefore given by

$$\begin{aligned} &P_0(x) = 1\,, \qquad P_1(x) = x\,, \qquad P_2(x) = \tfrac{1}{2}(3x^2-1)\,,\\ &P_3(x) = \tfrac{1}{2}(5x^3-3x)\,, \qquad P_4(x) = \tfrac{1}{8}(35x^4-30x^2+3)\,. \end{aligned} \tag{4.34}$$

With λ given by (4.28), and with the corresponding Legendre polynomial normalised according to (4.33), it is not hard to solve the recursion relation (4.20) explicitly, giving the result that

$$\boxed{P_\ell(x) = \sum_{k=0}^{[\ell/2]} \frac{(-1)^k\,(2\ell-2k)!}{2^\ell\,k!\,(\ell-k)!\,(\ell-2k)!}\,x^{\ell-2k}\,,} \tag{4.35}$$

where $[\ell/2]$ is the integer part of $\ell/2$.[26]

The expression (4.35) for the ℓ'th Legendre polynomial is somewhat unwieldy, and it is often useful to have alternative ways of writing $P_\ell(x)$. We shall give two such alternative expressions in the next two subsections.

4.2 Rodrigues' formula

First, noting that $\frac{d}{dx}\,x^p = px^{p-1}$, $\frac{d^2}{dx^2}\,x^p = p(p-1)x^{p-2}$, etc., and so $\frac{d^\ell}{dx^\ell}\,x^p = p!\,x^{p-\ell}/(p-\ell)!$, we observe that (4.35) can be written as

$$P_\ell(x) = \frac{d^\ell}{dx^\ell}\,\sum_{k=0}^{[\ell/2]} \frac{(-1)^k}{2^\ell\,k!\,(\ell-k)!}\,x^{2\ell-2k}\,. \tag{4.36}$$

The summation can then be extended up to $k=\ell$, since the extra terms that are added in the summation will involve powers of x of degree less than ℓ, and these will be killed off by the d^ℓ/dx^ℓ. Inserting a factor of $\ell!$ inside the sum, and dividing it out again in the prefactor, then gives

$$P_\ell(x) = \frac{1}{2^\ell\,\ell!}\,\frac{d^\ell}{dx^\ell}\,\sum_{k=0}^{\ell} \frac{(-1)^k\,\ell!}{k!\,(\ell-k)!}\,x^{2\ell-2k}\,, \tag{4.37}$$

[26] To see this, read off the coefficients a_n in (4.14) by equating the coefficients of each power of x with those in (4.35), and then show that these expressions for a_n indeed satisfy the recursion relation (4.20). This proves that (4.35) indeed satisfies the Legendre equation (4.32). The only slightly tricky point is establishing that the normalisation in (4.35) is indeed such that $P_\ell(x)$ satisfies (4.33). We give a simple proof of this in the next section.

and we can recognise the sum as the binomial expansion of $(x^2-1)^\ell$. Thus we arrive at *Rodrigues' formula* for $P_\ell(x)$:

$$\boxed{P_\ell(x) = \frac{1}{2^\ell\,\ell!}\,\frac{d^\ell}{dx^\ell}\,(x^2-1)^\ell\,.} \tag{4.38}$$

Having shown that the expessions (4.35) and (4.38) agree, we can now easily confirm that $P_\ell(x)$ so defined does indeed satisfy the normalisation (4.33). To do this, write (4.38) as

$$P_\ell(x) = \frac{1}{2^\ell\,\ell!}\,\frac{d^\ell}{dx^\ell}\,\Big[(x-1)^\ell\,(x+1)^\ell\Big]\,, \tag{4.39}$$

and note that when we differentiate the product $(x-1)^\ell\,(x+1)^\ell$ a total of ℓ times, the *only* term that survives after then setting $x=1$ is the term where all ℓ derivatives land on $(x-1)^\ell$. Since $\frac{d^\ell}{dx^\ell}\,(x-1)^\ell = \ell!$, we see that indeed

$$P_\ell(1) = \frac{1}{2^\ell\,\ell!}\,\ell!\,(1+1)^\ell = 1\,. \tag{4.40}$$

4.3 The generating function

Another very useful way of representing the Legendre polynomials is by means of a *Generating Function*. The claim is that the Legendre polynomials $P_\ell(x)$ satisfy

$$\boxed{G(x,t) \equiv (1-2x\,t+t^2)^{-1/2} = \sum_{\ell\ge 0} t^\ell\,P_\ell(x)\,,} \tag{4.41}$$

where, for convergence of the series, we must have $|t|<1$. We can see how this is working by looking at the first few terms in the power-series expansion of the left-hand side in powers of t, giving

$$G(x,t) = 1 + x\,t + \tfrac{1}{2}(3x^2-1)\,t^2 + \tfrac{1}{2}(5x^3-3x)\,t^3 + \tfrac{1}{8}(35x^4-30x^2+3)\,t^4 + \cdots\,. \tag{4.42}$$

Equating this with the right-hand side of (4.41), and comparing the coefficients of each power of t, we read off

$$P_0(x) = 1\,,\quad P_1(x) = x\,,\quad P_2(x) = \tfrac{1}{2}(3x^2-1)\,,\quad P_3(x) = \tfrac{1}{2}(5x^3-3x) \tag{4.43}$$

and so on, which is precisely what we were finding in (4.34).

To prove that (4.41) correctly generates *all* the Legendre polynomials, we note that if $P_\ell(x)$ satisfies the Legendre equation (4.32) for all ℓ, then multiplying by t^ℓ and summing over ℓ implies that H defined by

$$H \equiv \sum_{\ell\ge 0} t^\ell\,[(1-x^2)\,P_\ell'' - 2x\,P_\ell' + \ell(\ell+1)\,P_\ell] \tag{4.44}$$

should vanish. In fact, H will vanish if and only if $P_\ell(x)$ satisfies the Legendre equation (4.32) for all integers $\ell \geq 0$. Now, looking at (4.41) we can see that H can be written as

$$H = (1-x^2)\,\frac{\partial^2 G(x,t)}{\partial x^2} - 2x\,\frac{\partial G(g,t)}{\partial x} + t\,\frac{\partial^2 (t\,G(x,t))}{\partial t^2}\,. \tag{4.45}$$

(The three terms here correlate exactly with the three terms on the right-hand side of (4.44).) It is now just a simple exercise to calculate the G derivatives in (4.45), using the definition of G in (4.41), to show that indeed we have $H = 0$, which proves that the functions $P_\ell(x)$ defined by (4.41) satisfy the Legendre equation. They are clearly polynomials, because the power-series expansion of the left-hand side of (4.41) in powers of t will clearly have x-dependent coefficients that are polynomial in x, as we saw in the first few terms of the expansion, in (4.42).

Finally, we must check the normalisation, i.e. that $P_\ell(1) = 1$. This is very easy; we just set $x = 1$ in (4.41), to get

$$(1-2t+t^2)^{-1/2} = \sum_{\ell\geq 0} t^\ell\, P_\ell(1)\,. \tag{4.46}$$

But the left-hand side is just $(1-t)^{-1}$, which has the binomial expansion

$$\frac{1}{1-t} = 1 + t + t^2 + t^3 + t^4 + \cdots = \sum_{\ell\geq 0} t^\ell\,, \tag{4.47}$$

and so by comparing with the right-hand side in (4.46) we immediately get $P_\ell(1) = 1$.

4.4 Expansion in Legendre polynomials

Recall that our goal is to construct solutions of Laplace's equation (4.2) written in spherical polar coordinates, and that we have established that for azimuthally-symmetric solutions the relevant functions in the θ direction are expressed in terms of Legendre polynomials. We have seen that the factorised azimuthally-symmetric solutions for the potential will be of the form $\phi(r,\theta) = r^{-1}\,R_\ell(r)\,P_\ell(\cos\theta)$, where P_ℓ are the Legendre polynomials and $R_\ell(r)$ will be the solutions of the radial equation in (4.7), now with the separation constant λ determined to be equal to $\ell(\ell+1)$. The general azimuthally-symmetric solution will then involve a sum over all such factorised solutions, with ℓ-dependent constant coefficients that will be determined by the boundary conditions.

Constructing a general azimuthally-symmetric solution will therefore require that we expand a general function of θ as a sum over Legendre polynomials $P_\ell(\cos\theta)$. In terms of

$x = \cos\theta$, the first task then is to expand a general function $f(x)$ in the form

$$\boxed{f(x) = \sum_{\ell \geq 0} a_\ell \, P_\ell(x) \,.} \tag{4.48}$$

We can establish the following properties of the Legendre polynomials. Firstly,

$$\int_{-1}^{1} dx P_\ell(x) P_n(x) = 0 \,, \qquad \text{if} \quad \ell \neq n \,. \tag{4.49}$$

This can be seen by taking the Legendre equation (4.32) and multiplying it by $P_n(x)$, then subtracting the same expression but with ℓ and n interchanged, and finally, integrating over the interval $-1 \leq x \leq 1$. This gives

$$\begin{aligned} &\int_{-1}^{1} dx \Big[P_n(x) \frac{d}{dx}\Big((1-x^2)\frac{dP_\ell(x)}{dx}\Big) - P_\ell(x)\frac{d}{dx}\Big((1-x^2)\frac{dP_n(x)}{dx}\Big) \\ &[\ell(\ell+1) - n(n+1)]\, P_\ell(x) P_n(x)\Big] = 0 \,. \end{aligned} \tag{4.50}$$

Integrating the top line by parts gives

$$\begin{aligned} &\Big[P_n(x)\,(1-x^2)\,\frac{dP_\ell(x)}{dx} - P_\ell(x)\,(1-x^2)\,\frac{dP_n(x)}{dx}\Big]_{-1}^{1} \\ &+[\ell(\ell+1) - n(n+1)] \int_{-1}^{1} dx\, P_\ell(x) P_n(x)\Big] = 0 \,. \end{aligned} \tag{4.51}$$

The first line vanishes, because of the $(1-x^2)$ factor, leaving

$$[\ell(\ell+1) - n(n+1)] \int_{-1}^{1} dx\, P_\ell(x) P_n(x) = 0 \,. \tag{4.52}$$

Thus if $\ell \neq n$ (and recalling that ℓ and n are both non-negative integers), we see that the orthogonality condition (4.49) must hold.

We also need to know what the integral in (4.49) gives when $\ell = n$. Here is a slightly unorthodox derivation, which is actually quite elegant. Using the generating function (4.41) twice over, once for $\sum_\ell t^\ell P_\ell(x)$ and once for $\sum_n t^n P_n(x)$, we can deduce that

$$\int_{-1}^{1} dx (1-2xt+t^2)^{-1/2} (1-2xt+t^2)^{-1/2} = \sum_{\ell \geq 0} \sum_{n \geq 0} t^\ell t^n \int_{-1}^{1} dx P_\ell(x) P_n(x) \,, \tag{4.53}$$

and so

$$\int_{-1}^{1} \frac{dx}{1-2xt+t^2} = \sum_{\ell \geq 0} \sum_{n \geq 0} t^\ell t^n \int_{-1}^{1} dx P_\ell(x) P_n(x) \,, \tag{4.54}$$

Using the already-established orthogonality result (4.49), we therefore find

$$\int_{-1}^{1} \frac{dx}{1-2xt+t^2} = \sum_{\ell \geq 0} t^{2\ell} \int_{-1}^{1} dx \big(P_\ell(x)\big)^2 \,. \tag{4.55}$$

Performing the integral on the left-hand side gives

$$\frac{1}{t}\,\log\left(\frac{1+t}{1-t}\right) = \sum_{\ell\geq 0} t^{2\ell}\,\int_{-1}^{1} dx\big(P_\ell(x)\big)^2\,, \tag{4.56}$$

and expanding the logarithm in a Taylor series implies

$$\sum_{\ell\geq 0}\frac{2t^{2\ell}}{2\ell+1} = \sum_{\ell\geq 0} t^{2\ell}\,\int_{-1}^{1} dx\big(P_\ell(x)\big)^2\,. \tag{4.57}$$

Equating the coefficients of each power of t then gives[27]

$$\int_{-1}^{1} dx\big(P_\ell(x)\big)^2 = \frac{2}{2\ell+1}\,. \tag{4.60}$$

In summary, we have shown that

$$\boxed{\int_{-1}^{1} dx P_\ell(x)P_n(x) = \frac{2}{2\ell+1}\,\delta_{\ell,n}\,,} \tag{4.61}$$

where $\delta_{\ell,n}$ is the Kronecker delta symbol, which equals 1 if ℓ and n are equal, and is 0 otherwise.

With this result, we can now determine the coefficients a_ℓ in the generalised Fourier expansion (4.48). Multiplying this equation by $P_n(x)$, integrating over the interval $-1 \leq x \leq 1$, and using (4.61), we find

$$\int_{-1}^{1} dx f(x)P_n(x) = \sum_{\ell\geq 0} a_\ell\,\int_{-1}^{1} dx P_\ell(x)P_n(x) = \sum_{\ell\geq 0}\frac{2}{2\ell+1}\,\delta_{\ell n}\,a_\ell = \frac{2}{2n+1}\,a_n\,, \tag{4.62}$$

and hence the coefficients in (4.48) are given by

$$\boxed{a_\ell = \tfrac{1}{2}(2\ell+1)\,\int_{-1}^{1} dx f(x)P_\ell(x)\,.} \tag{4.63}$$

[27] With a little more work, one can consider the result where we use a different auxiliary variable s, instead of t, in the generating function for $P_n(x)$:

$$(1-2xs+s^2)^{-1/2} = \sum_{n\geq 0} s^n\,P_n(x)\,.$$

We then have

$$\int_{-1}^{1} dx(1-2xt+t^2)^{-1/2}(1-2xs+s^2)^{-1/2} = \sum_{\ell\geq 0}\sum_{n\geq 0} t^\ell s^n\,\int_{-1}^{1} dx P_\ell(x)P_n(x)\,. \tag{4.58}$$

This gives

$$\frac{1}{\sqrt{st}}\,\Big[2\log(1+\sqrt{st}) - \log(1-st)\Big] = \sum_{p\geq 0}\frac{2}{2p+1}\,(st)^p\,, \tag{4.59}$$

and so matching the powers of s and t with those on the right-hand side of (4.58), one can deduce immediately the orthogonality (4.49) *and* the result (4.60).

4.5 Azimuthally-symmetric solutions of Laplace's equation

With these preliminaries, we can now return to the original problem, of finding solutions to Laplace's equation in spherical polar coordinates. For now, we are restricting attention to the case where the problem has azimuthal symmetry, meaning that it is independent of the azimuthal coordinate φ. This means that we take $m = 0$ in (4.8), and so the original potential ϕ has been separated by finding factorised solutions of the form

$$\phi(r,\theta) = \frac{1}{r}\, R(r)\, \Theta(\theta)\,, \tag{4.64}$$

with R and Θ satisfying

$$\frac{d^2R}{dr^2} = \frac{\lambda}{r^2}\, R\,, \qquad \frac{1}{\sin\theta}\,\frac{d}{d\theta}\Big(\sin\theta\,\frac{d\Theta}{d\theta}\Big) + \lambda\,\Theta = 0\,. \tag{4.65}$$

(See equations (4.7) and (4.8).) As we now know, the equation for Θ becomes the Legendre equation (4.11) after making the change of variable $\cos\theta = x$, and its regular solutions are the Legendre polynomials $P_\ell(x)$, occurring when $\lambda = \ell(\ell+1)$.

It is easy to see, by looking for trial solutions of the form $R(r) = r^\alpha$, that the two linearly independent solutions of the radial equation in (4.65) are given by

$$R = r^{\ell+1} \qquad \text{and} \qquad R = r^{-\ell}\,. \tag{4.66}$$

Thus by summing over all possible factorised solutions of the form (4.64), we arrive at the general azimuthally-symmetric solution

$$\boxed{\phi(r,\theta) = \sum_{\ell\geq 0}\Big(A_\ell\, r^\ell + B_\ell\, r^{-\ell-1}\Big)\, P_\ell(\cos\theta)\,,} \tag{4.67}$$

where A_ℓ and B_ℓ are arbitrary constants. These will be determined by the boundary conditions that specify a particular physical configuration.

Exterior Problem:

Suppose, for example, we are considering the *exterior* problem, in which the potential is specified on a spherical surface at $r = a$, and one wishes to calculate the potential at all points outside this surface. We may assume that the potential goes to zero at infinity, and so we must have

$$A_\ell = 0\,. \tag{4.68}$$

The B_ℓ coefficients are determined in the manner we established in the previous subsection. Note that after transforming back to the variable θ from the variable $x = \cos\theta$, the integral

(4.61) becomes

$$\int_0^\pi d\theta \sin\theta\, P_\ell(\cos\theta) P_n(\cos\theta) = \frac{2}{2\ell+1}\,\delta_{\ell,n}\,, \tag{4.69}$$

and so the coefficients B_ℓ in (4.67) are given in terms of the boundary values $\phi(a,\theta)$ of the potential by

$$B_\ell = \tfrac{1}{2}(2\ell+1)\, a^{\ell+1} \int_0^\pi d\theta\, \sin\theta\, \phi(a,\theta)\, P_\ell(\cos\theta)\,. \tag{4.70}$$

Interior Problem:

A related case arises for the *interior* problem, where the potential is specified on a closed spherical surface $r = a$ and one wishes to solve for the potential everywhere inside this surface. Assuming there are no charges inside the sphere (i.e. it is just a vacuum inside), then the coefficients B_ℓ in the general expansion (4.67) must now be zero, since we require that the potential be regular at $r = 0$. The A_ℓ coefficients will now be given by

$$A_\ell = \tfrac{1}{2}(2\ell+1)\, a^{-\ell} \int_0^\pi d\theta\, \sin\theta\, \phi(a,\theta)\, P_\ell(\cos\theta)\,. \tag{4.71}$$

Region Between Two Spherical Surfaces:

A third related example arises if one wishes to solve for ϕ in the free-space region between two concentric spherical surfaces with $r = a$ and $r = b$. Let us assume that $b > a$. Now, both the A_ℓ and B_ℓ coefficients will in general be non-vanishing. The boundary values for ϕ must be specified on both the surfaces, and so one again has the correct total number of equations to solve for all of the coefficients. Thus one will have

$$\begin{aligned} A_\ell a^\ell + B_\ell a^{-\ell-1} &= \tfrac{1}{2}(2\ell+1) \int_0^\pi d\theta\, \sin\theta\, \phi(a,\theta)\, P_\ell(\cos\theta)\,, \\ A_\ell b^\ell + B_\ell b^{-\ell-1} &= \tfrac{1}{2}(2\ell+1) \int_0^\pi d\theta\, \sin\theta\, \phi(b,\theta)\, P_\ell(\cos\theta)\,, \end{aligned} \tag{4.72}$$

which can be solved for all A_ℓ and B_ℓ.

Example with Neumann boundary condition:

Of course one can easily consider other classes of problem too. For example, one may consider Neumann cases where it is the normal derivative of ϕ, rather than ϕ itself, that is specified on the spherical boundary or boundaries. Consider, for example, the exterior problem in this case. From (4.67), again assuming that ϕ goes to zero at infinity and so $A_\ell = 0$, one has

$$\frac{\partial\phi(r,\theta)}{\partial r}\Big|_{r=a} = -\sum_{\ell\geq 0} (\ell+1) B_\ell a^{-\ell-2}\, P_\ell(\cos\theta)\,. \tag{4.73}$$

Multiplying by $P_n(\cos\theta)$ and integrating, we therefore find

$$B_\ell = -\frac{(2\ell+1)}{2(\ell+1)}\, a^{\ell+2} \int_0^\pi d\theta\, \sin\theta \left(\frac{\partial\phi(r,\theta)}{\partial r}\Big|_{r=a}\right) P_\ell(\cos\theta)\,. \tag{4.74}$$

Analogous results can be obtained for the interior problem and for the case of the region between two concentric spheres.

4.6 Some useful techniques for azimuthally-symmetric problems

The procedure for solving Laplace's equation described in the previous subsection is straightforward, but evaluating the integrals in order to read off the expansion coefficients in (4.48) can sometimes be a little involved. There is a very useful "trick" which can often be employed to obtain the answer in a more elegant way, with considerably less effort. We describe this, and a couple of further techniques, below.

Solution by off-axis extrapolation:

In an electrostatics problem with azimuthal symmetry, it is often relatively easy to obtain the expression for the potential along the axis of symmetry by elementary means. Thus, if we make the natural choice and align the z axis along the axis of azimuthal symmetry, this means that we may have a situation where the potential $\phi(z)$, on the z axis, is easily calculated. In terms of spherical polar coordinates, points along the z axis correspond to $\theta = 0$ (for the positive z axis), or $\theta = \pi$ (for the negative z axis).

Consider an "exterior" azimuthally-symmetric electrostatics problem, for which we wish to find $\phi(r,\theta)$ in an exterior region, say $r \geq a$, subject to the assumption that ϕ goes to zero at infinity. From (4.48), the A_ℓ coefficients will be zero, and the general such solution to Laplace's equation will take the form

$$\phi(r,\theta) = \sum_{\ell\geq 0} \frac{B_\ell}{r^{\ell+1}}\, P_\ell(\cos\theta)\,. \tag{4.75}$$

Suppose now that we already know what the solution for ϕ is on the z axis; call this $\phi(z)$. For simplicity we shall first consider first the positive z axis, which corresponds to $\theta = 0$. Recalling that in the transformation from Cartesian to spherical polar coordinates we have $z = r\cos\theta$, we note that $\theta = 0$ and $r = z$ on the positive z axis, and therefore the potential on the positive z axis, following from (4.75), will be

$$z > 0: \qquad \phi(z) = \phi(z,0) = \sum_{\ell\geq 0} \frac{B_\ell}{z^{\ell+1}}\, P_\ell(1) = \sum_{\ell\geq 0} \frac{B_\ell}{z^{\ell+1}}\,, \tag{4.76}$$

where, in getting to the final expression, we have used the normalisation property $P_\ell(1) = 1$ of the Legendre polynomials.

If we have already, by some other means, calculated the potential $\phi(z)$ on the positive z axis then we only need to take this expression, expand it as a power series in inverse powers of z, and equate the coefficients of each inverse power of z to the corresponding coefficients B_ℓ in (4.76). Having thereby determined the B_ℓ coefficients we can then substitute them into (4.77) in order to obtain an expression for the potential $\phi(r, \theta)$ in the entire upper half space corresponding to $0 \le \theta < \frac{1}{2}\pi$. Thus, we have performed an extropolation of a result for the potential on the positive z axis into the entire upper half space.

There may in fact be circumstances where the expansion coefficients are different in the regions above and below the $z = 0$ plane (i.e. in the regions $0 \le \theta < \frac{1}{2}\pi$ and $\frac{1}{2}\pi < \theta \le \pi$). To allow for this possibility, we shall write

$$\phi(r,\theta) = \sum_{\ell\ge 0} \frac{B_\ell}{r^{\ell+1}}\, P_\ell(\cos\theta)\,, \qquad 0 \le \theta < \tfrac{1}{2}\pi\,, \tag{4.77}$$

$$\phi(r,\theta) = \sum_{\ell\ge 0} \frac{\widetilde{B}_\ell}{r^{\ell+1}}\, P_\ell(\cos\theta)\,, \qquad \tfrac{1}{2}\pi < \theta \le \pi\,. \tag{4.78}$$

We can then carry out an analogous extrapolation in the lower half space. On the negative z axis we have $\theta = \pi$ and so $r = -z = |z|$ (r is of course always, by definition, non-negative). Thus from (4.78) we shall have[28]

$$z < 0: \qquad \phi(z) = \phi(|z|,\pi) = \sum_{\ell\ge 0} \frac{\widetilde{B}_\ell}{|z|^{\ell+1}}\, P_\ell(-1) = \sum_{\ell\ge 0} \frac{\widetilde{B}_\ell}{|z|^{\ell+1}} (-1)^\ell = -\sum_{\ell\ge 0} \frac{\widetilde{B}_\ell}{z^{\ell+1}}\,. \tag{4.79}$$

Note that we have made use of the fact that $P_\ell(-x) = (-1)^\ell\, P_\ell(x)$, and so $P_\ell(-1) = (-1)^\ell$. Thus if we already know by some other means the expression for the potential on the negative z axis, we just have to expand it as a power series in inverse powers of z and then, matching it with (4.79), read off the coefficients $\widetilde{B}_\ell$. Substituting into (4.78) then gives an expression for the potential in the entire lower half space.

The crucial thing that allows this procedure of off-axis extrapolation to work is that the number of arbitrary coefficients in the expansion (4.77) (or (4.78)), namely one coefficient per inverse power of r, is exactly the same as the number of coefficients in the expansion of a potential on the z axis in inverse powers of z, namely, one coefficient per inverse power.

[28] For a first reading of this material, it is probably advantageous to forget about the subtleties of negative z versus positive z, and just focus on the previous discussion for positive z only. Otherwise, there is a risk that the essential simplicity of the procedure becomes obscured in the minutiae of the minus signs.

Thus the knowledge of these coefficients in the expansion of the potential on the z axis fully determines the coefficients in (4.77) or (4.78).[29]

An analogous discussion can be given for the small-r region, where we match expansions of the form

$$\phi(r,\theta) = \sum_{\ell\geq 0} A_\ell\, r^\ell\, P_\ell(\cos\theta)\,, \qquad 0\leq\theta<\tfrac{1}{2}\pi\,, \tag{4.80}$$

$$\phi(r,\theta) = \sum_{\ell\geq 0} \widetilde{A}_\ell\, r^\ell\, P_\ell(\cos\theta)\,, \qquad \tfrac{1}{2}\pi<\theta\leq\pi \tag{4.81}$$

to a small-z expansion of the on-axis potential. Again, one needs in general to allow for the possibility that the coefficients A_ℓ determined from the expansion on the positive z axis may be different from the coefficients $\tilde{A}_\ell$ determined from the expansion on the negative z axis.

Note that in practice, the potentials one is working with often have a very simple symmetry between the positive-z region and the negative-z region. For example, one may have $\phi(z)=\phi(-z)$, or else one may have $\phi(z)=-\phi(-z)$. In either of these cases, an easy way to sort out the relation beween the positive-z expansion and the negative-z expansion is to work out first the extrapolation from $\phi(z)$ to $\phi(r,\theta)$ for positive z, and then check to see whether the resulting Legendre polynomial expansion for $\phi(r,\theta)$ gives the right result on the negative z axis. If it agrees with the known expression for $\phi(z)$ on the negative z axis then all is well. If the signs of the coefficients are incorrect, then one has to write a different expansion for $\phi(r,\theta)$ in the negative-z region, with the signs changed appropriately.

As an example, consider the problem that we discussed in section (2.5.2), when illustrating the use of the Dirichlet Green function for the sphere. The potential on the upper hemishpere of a sphere of radius a was taken to be the constant $+V$, and on the lower hemisphere the potential was $-V$. The exact expression for the potential in the region $r>a$ was obtained in the form of the integral (2.68), but this cannot be evaluated explicitly in any useful form. We observed at the time that the expression for ϕ on the (positive) z axis took the form (2.69). We should really be a little more careful now, and consider the expression on the negative z axis also. If we do this, it is straightfoward to see that the general on-axis expression for ϕ, valid both for $z>a$ and for $z<-a$, is

$$\phi(z) = V\,\left(\mathrm{sign}(z) - \frac{z^2-a^2}{z\sqrt{z^2+a^2}}\right)\,, \tag{4.82}$$

[29] An off-axis extrapolation like this would not work in a situation where there is no azimuthal symmetry, as we shall see later.

where sign(z) equals 1 if $z > 0$, whilst it equals -1 if $z < 0$. Note that (4.82) has the property $\phi(z) = -\phi(-z)$, i.e. that it changes sign under $z \to -z$, as it should since the boundary data on the sphere at $r = a$ changes sign under $z \to -z$. The potential (4.82) can be written as

$$\phi(z) = \text{sign}(z)\, V \left[1 - \left(1 - \frac{a^2}{z^2}\right)\left(1 + \frac{a^2}{z^2}\right)^{-1/2}\right]. \tag{4.83}$$

It is a simple matter to expand the inverse square root using the binomial theorem, and thereby obtain the expansion for $\phi(z)$ as an infinite series in powers of $1/z$. The first few terms are

$$\phi(z) = \text{sign}(z)\left[\frac{3V}{2}\,\frac{a^2}{z^2} - \frac{7V}{8}\,\frac{a^4}{z^4} + \frac{11V}{16}\,\frac{a^6}{z^6} - \frac{75V}{128}\,\frac{a^8}{z^8} + \cdots\right], \tag{4.84}$$

and the general result is

$$\begin{aligned}\phi(z) &= \text{sign}(z)\, V \left[1 - \left(1 - \frac{a^2}{z^2}\right)\sum_{n\ge 0}\frac{(-1)^n\,(2n)!}{2^{2n}\,(n!)^2}\,\left(\frac{a}{z}\right)^{2n}\right] \\ &= \text{sign}(z)\, V \sum_{n\ge 1}\frac{(-1)^{n+1}\,(4n-1)(2n-2)!}{2^{2n-1}\,n!\,(n-1)!}\,\left(\frac{a}{z}\right)^{2n}.\end{aligned} \tag{4.85}$$

Reading off the coefficients in the expansion (4.84) or (4.85), and matching them against the expansion coefficients B_ℓ and $\widetilde{B}_\ell$ in (4.76) and (4.79), we can therefore immediately conclude that $B_{2n} = \widetilde{B}_{2n} = 0$, that the odd coefficients are non-zero, with $B_{2n+1} = \widetilde{B}_{2n+1}$, and that the general off-axis series for the potential is given by

$$\begin{aligned}\phi(r,\theta) &= V \sum_{n\ge 1}\frac{(-1)^{n+1}\,(4n-1)(2n-2)!}{2^{2n-1}\,n!\,(n-1)!}\,\left(\frac{a}{r}\right)^{2n} P_{2n-1}(\cos\theta) \\ &= \frac{3V}{2}\,\frac{a^2}{r^2}\,P_1(\cos\theta) - \frac{7V}{8}\,\frac{a^4}{r^4}\,P_3(\cos\theta) + \frac{11V}{16}\,\frac{a^6}{r^6}\,P_5(\cos\theta) - \frac{75V}{128}\,\frac{a^8}{r^8}\,P_7(\cos\theta) + \cdots,\end{aligned} \tag{4.86}$$

valid in the region $r > a$. As a check, we may observe that the potential indeed changes sign under reflection in the $z = 0$ plane, i.e. under $\theta \to \pi - \theta$, because all the $P_\ell(\cos\theta)$ have this property when ℓ is odd. Thus in this case, the result we obtain by off-axis extrapolation in the upper half space is already also valid in the lower half space.

The technique of off-axis extrapolation is clearly very powerful. Indeed, it can really be viewed as the default approach that one should try first, when solving electrostatics problems where there is azimuthal symmetry.

Solution by inversion:

It is sometimes the case that one already has the result for the series expansion in, say, the exterior region $r > a$, and one now wishes to find the series expansion in the interior

region $r < a$. This can, of course, be done by going back and re-solving the problem from scratch, in the region $r < a$, where now it will be the A_ℓ coefficients in (4.48), rather than the B_ℓ coefficients, that are non-zero. Sometimes, an easier method is to use the following procedure, known as *Inversion*.

Suppose we have already found that the solution in the exterior region is given by

$$\phi(r,\theta) = \sum_{\ell\geq 0} \frac{B_\ell}{r^{\ell+1}}\, P_\ell(\cos\theta)\,, \tag{4.87}$$

where the coefficients B_ℓ are known. On general grounds, we know that the solution in the interior region will be of the general form

$$\phi(r,\theta) = \sum_{\ell\geq 0} A_\ell\, r^\ell\, P_\ell(\cos\theta)\,. \tag{4.88}$$

(We are assuming there are no free charges in the interior region.)

Under the assumption that there are no charges that could give rise to singularities at the $r = a$ interface, it follows that we can use either of the expressions (4.87) or (4.88) at $r = a$ itself, and so equating the two, we find

$$\sum_{\ell\geq 0} \frac{B_\ell}{a^{\ell+1}}\, P_\ell(\cos\theta) = \sum_{\ell\geq 0} A_\ell\, a^\ell\, P_\ell(\cos\theta)\,. \tag{4.89}$$

Since this equation must hold for all θ, and since the Legendre polynomials are linearly independent, it follows that the coefficients of each P_ℓ separately must be equal on the two sides of the equation. Therefore we can deduce that

$$A_\ell = \frac{B_\ell}{a^{2\ell+1}}\,, \tag{4.90}$$

and hence the series expansion (4.88) in the interior region is given by

$$\phi(r,\theta) = \sum_{\ell\geq 0} \frac{B_\ell\, r^\ell}{a^{2\ell+1}}\, P_\ell(\cos\theta)\,, \tag{4.91}$$

in terms of the already-determined coefficients B_ℓ of the exterior region.

On could, of course, equally well apply the inversion procedure the other way around, and instead find the expansion in the exterior region if the expansion in the interior region is already known.

The example we discussed previously, of the potential outside the two hemispheres held at potentials V and $-V$, provides a nice illustration of the method of inversion. The general solution for the potential outside the two hemispheres was found to be given by (4.86). Using

the inversion relation (4.90), it therefore follows that all we need to do to obtain the solution inside the two hemispheres is to apply the replacement rule

$$\left(\frac{a}{r}\right)^{\ell+1} P_\ell(\cos\theta) \longrightarrow \left(\frac{r}{a}\right)^{\ell} P_\ell(\cos\theta \tag{4.92}$$

to the term in $P_\ell(\cos\theta)$ in (4.86), for each value of ℓ. This therefore gives the interior solution

$$\begin{aligned}\phi(r,\theta) &= V\sum_{n\geq 1}\frac{(-1)^{n+1}\,(4n-1)(2n-2)!}{2^{2n-1}\,n!\,(n-1)!}\left(\frac{r}{a}\right)^{2n-1} P_{2n-1}(\cos\theta) \qquad (4.93)\\ &= \frac{3V}{2}\,\frac{r}{a}\,P_1(\cos\theta) - \frac{7V}{8}\,\frac{r^3}{a^3}\,P_3(\cos\theta) + \frac{11V}{16}\,\frac{r^5}{a^5}\,P_5(\cos\theta) - \frac{75V}{128}\,\frac{r^7}{a^7}\,P_7(\cos\theta) + \cdots .\end{aligned}$$

The method of inversion is clearly a very convenient way of solving the interior problem, once the exterior problem has been solved, or *vice versa*. In fact, all that is really going on here is that if one knows the solution for $\phi(r,\theta)$ in the region $a \leq r \leq \infty$ *exterior* to the surface $r = a$, then in particular one knows the potential on the surface at $r = a$. This expression, $\phi(a,\theta)$, can then be taken as the Dirichlet boundary data for solving the *interior* problem in the region $0 \leq r \leq a$.

One must, however, be careful about the circumstances under which the inversion procedure can be applied. It is *essential* that the series expansion that is valid in the region $r > a$ should also be valid actually at $r = a$, and likewise that the solution valid for $r < a$ should be valid at $r = a$. This is an issue of *convergence* of the series, and in turn this is related to the question of the *analyticity* of the solution.

In general, the exterior and interior series will be convergent on the surface $r = a$ itself *as long as there are no singularities anywhere on the $r = a$ surface.* This is true in the example with the two hemisphers that we considered, where the potential is perfectly finite at all points on the $r = a$ surface (it is either $+V$ or $-V$). However, an example where it may *not* be true is if we considered a situation where there was an infinite charge density somewhere on the $r = a$ surface, such as would result from point charges or other kinds of singular charge densities. In fact the singularities need not be so extreme as in this example, and still the method of inversion may fail. For example, a function such as $f(x) = (x - x_0)^{5/2}$ is singular at $x = x_0$, in the sense that it is not analytic there. (The third derivative, and above, do not exist at $x = x_0$, and so $f(x)$ does not admit a Taylor expansion around $x = x_0$.)

In summary, while the method of inversion can sometimes be useful, it must be applied with great care. It is often in practice safer to recalculate from scratch in the inner region

(for example, using the off-axis extrapolation method discussed previously), rather than applying inversion to the solution in the outer region.

4.7 The spherical harmonics

So far in our discussion of solutions of Laplace's equation in spherical polar coordinates, we have focused on situations with azimuthal symmetry, for which the associated Legendre equation (4.11) reduced to the Legendre equation (4.13). Now, we have to restore the integer m, and consider the associated Legendre equation itself.

For convenience, we present again the Associated Legendre Equation:

$$\frac{d}{dx}\Big((1-x^2)\,\frac{dy}{dx}\Big) + \Big(\lambda - \frac{m^2}{1-x^2}\Big)\,y = 0\,. \tag{4.94}$$

Luckily, it turns out that we can construct the relevant solutions of this equation rather simply, in terms of the Legendre polynomials that we have already studied.

To begin, we change variables from $y(x)$ to $w(x)$, where $y(x) = (1-x^2)^{m/2}\,w(x)$, and substitute this into (4.94). After simple algebra we find, after extracting an overall factor of $(1-x^2)^{m/2}$, that w must satisfy

$$(1-x^2)\,w'' - 2(m+1)\,x\,w' + [\lambda - m\,(m+1)]\,w = 0\,. \tag{4.95}$$

(We are using a prime to denote differentiation d/dx here.) Now suppose that we have a solution u of the ordinary Legendre equation:

$$(1-x^2)\,u'' - 2x\,u' + \lambda\,u = 0\,. \tag{4.96}$$

Next, we differentiate this m times. Let us use the notation ∂^m as a shorthand for d^m/dx^m. It is useful to note that we have the following lemma, which is just a consequence of Leibnitz' rule for the differentiation of a product, iterated m times:

$$\begin{aligned}\partial^m(f\,g) \;&=\; f\,(\partial^m g) + m\,(\partial f)\,(\partial^{m-1} g) + \frac{m(m-1)}{2!}\,(\partial^2 f)\,(\partial^{m-2} g)\\ &\quad + \frac{m(m-1)(m-2)}{3!}\,(\partial^3 f)\,(\partial^{m-3} g) + \cdots\,.\end{aligned} \tag{4.97}$$

We only need the first two or three terms in this expression if we apply it to the products in (4.96), and so we easily find that

$$(1-x^2)\,\partial^{m+2} u - 2(m+1)\,x\,\partial^{m+1} u + [\lambda - m(m+1]\,\partial^m\,u = 0\,. \tag{4.98}$$

Comparing this equation with the equation (4.95) for w, we see that $\partial^m\,u$ satisfies the same equation as does w, and so by setting $w = \partial^m u$, we have constructed a solution

of (4.95) in terms of a solution u of the Legendre equation (4.96). Thus we have shown that $y = (1-x^2)^{m/2}\,\partial^m u$ satisfies the associated Legendre equation (4.94). The upshot, therefore, is that if we define

$$P_\ell^m(x) \equiv (-1)^m\,(1-x^2)^{m/2}\,\frac{d^m}{dx^m}\,P_\ell(x)\,, \tag{4.99}$$

where $P_\ell(x)$ is a Legendre polynomial, then $P_\ell^m(x)$ will be a solution of the Associated Legendre Equation with $\lambda = \ell\,(\ell+1)$:

$$\frac{d}{dx}\Big((1-x^2)\,\frac{dP_\ell^m}{dx}\Big) + \Big(\ell\,(\ell+1) - \frac{m^2}{1-x^2}\Big)\,P_\ell^m = 0\,. \tag{4.100}$$

(The $(-1)^m$ in eqn (4.99) is included to fit with standard conventions for the definition of $P_\ell^m(x)$.) Since $P_\ell(x)$ is regular everywhere including $x = \pm 1$, it is clear that $P_\ell^m(x)$, viewed as a function of θ, will be too. (Recall that $x = \cos\theta$, so $(1-x^2)^{m/2} = (\sin\theta)^m$.) It is understood here that (for now) we are taking the integer m to be non-negative. It is clear that we can restrict m such that $m \le \ell$ too, since if m exceeds ℓ then the m-fold derivative of the ℓ'th Legendre polynomial (which itself is of degree ℓ) will give zero.

Recall next that we have Rodrigues' formula (4.38), which gives us an expression for $P_\ell(x)$. Substituting this into (4.99), we get the *Generalised Rodrigues Formula*

$$\boxed{P_\ell^m(x) = \frac{(-1)^m}{2^\ell\,\ell!}\,(1-x^2)^{m/2}\,\frac{d^{\ell+m}}{dx^{\ell+m}}\,(x^2-1)^\ell\,.} \tag{4.101}$$

A nice little miracle now occurs: this formula makes sense for negative values of m too, provided that $m \ge -\ell$. Thus we have a construction of Associated Legendre Functions for all integers m in the interval $-\ell \le m \le \ell$.

Looking at the Associated Legendre Equation (4.100), we note that the equation itself is invariant under sending

$$m \longrightarrow -m\,, \tag{4.102}$$

since m appears only as m^2 in the equation. This means that if we take a solution with a given m, then sending m to $-m$ gives us another solution. What is more, only one solution at fixed ℓ and m^2 can be regular at $x = \pm 1$, since the second solution will have logarithmic singularities there (just like we saw for the Legendre functions). Since $P_\ell^m(x)$ and $P_\ell^{-m}(x)$ given by 4.101 are both regular at $x = \pm 1$ (and therefore neither of them can involve the second solution with logarithmic singularities at $x = \pm 1$), it follows that they must be linearly dependent; i.e. $P_\ell^{-m}(x)$ must be some constant multiple of $P_\ell^m(x)$:

$$P_\ell^{-m}(x) = k\,P_\ell^m(x)\,. \tag{4.103}$$

It is easy to determine what the constant k is, by using (4.101). From (4.103) we get

$$\partial^{\ell-m}(x^2-1)^\ell = k\,(1-x^2)^m\,\partial^{\ell+m}(x^2-1)^\ell\,. \tag{4.104}$$

It is good enough just to look at the highest power of x, since we have already argued that (4.103) must hold, and so all we need to do is to calculate what k is.[30] Thus we get

$$\frac{(2\ell)!}{(\ell+m)!}\,x^{\ell+m} = k\,(-1)^m\,x^{2m}\,\frac{(2\ell)!}{(\ell-m)!}\,x^{\ell-m} \tag{4.105}$$

at the leading order in x, which fixes k and hence establishes that

$$P_\ell^{-m}(x) = (-1)^m\frac{(\ell-m)!}{(\ell+m)!}\,P_\ell^m(x)\,. \tag{4.106}$$

Using this result we can now very easily work out the normalisation integral for the associated Legendre functions $P_\ell^m(x)$. The relevant integral we shall need to evaluate is

$$\int_{-1}^1 dx\,P_\ell^m(x)\,P_n^m(x)\,. \tag{4.107}$$

(It will become clear later why we have set the upper indices m equal here.) Using the same method as we used earlier for the Legendre polynomials,[31] it is easy to show that (4.107) vanishes unless $\ell = n$. For $\ell = n$, we can make use of (4.106) to write the required integral as

$$C_{\ell m} \equiv \int_{-1}^1 dx\,[P_\ell^m(x)]^2 = (-1)^m\frac{(\ell+m)!}{(\ell-m)!}\int_{-1}^1 dx\,P_\ell^m(x)\,P_\ell^{-m}(x)\,. \tag{4.108}$$

Our task is to calculate the constants $C_{\ell m}$.

Using (4.99), and knowing from (4.101) that (4.99) actually makes perfectly good sense for the negative values of m as well as the positive ones,[32] we shall have

$$\begin{aligned}\int_{-1}^1 dx\,P_\ell^m(x)\,P_\ell^{-m}(x) &= \int_{-1}^1 dx\,(-1)^m(1-x^2)^{m/2}\partial^m P_\ell(x)\,(-1)^m(1-x^2)^{-m/2}\partial^{-m}P_\ell(x)\\ &= \int_{-1}^1 dx\,\partial^m P_\ell(x)\partial^{-m}P_\ell(x)\,.\end{aligned} \tag{4.109}$$

We now integrate by parts m times, to push the ∂^m derivatives onto $\partial^{-m}P_\ell(x)$, noting that the boundary terms will cancel for each integration by parts. (Bear in mind the precise

[30] One could, more adventurously, give another proof that $P_\ell^{-m}(x)$ and $P_\ell^m(x)$ are linearly dependent by checking *all* powers of x. We leave this as an exercise for the reader.

[31] i.e. plug P_ℓ^m into the associated Legendre equation, multiply by P_n^m, integrate $\int_{-1}^1 dx$ and then subtract the same with the quantities ℓ and n exchanged.

[32] Remember here that the Legendre polynomial $P_\ell(x)$ is being written here in terms of Rodrigues' formula (4.38), so $\partial^{-m}\,P_\ell(x)$ actually means $1/(2^\ell\,\ell!)\,\partial^{\ell-m}\,(x^2-1)^\ell$. Thus, since we always require $m \geq -\ell$, this means that $\partial^{-m}\,P_\ell(x)$ always involves a non-negative number of derivatives applied to $(x^2-1)^\ell$.

definition of $\partial^{-m} P_\ell(x)$, as explained in footnote 32.) Thus we have

$$\int_{-1}^{1} dx\, P_\ell^m(x)\, P_\ell^{-m}(x) = (-1)^m \int_{-1}^{1} dx \big(P_\ell(x)\big)^2 = \frac{2(-1)^m}{2\ell+1}\,, \tag{4.110}$$

where we have used the previous result (4.60) in getting to the final result. Looking back to (4.108), we have therefore established that

$$\int_{-1}^{1} dx\, P_\ell^m(x)\, P_{\ell'}^m(x) = \frac{2}{(2\ell+1)}\,\frac{(\ell+m)!}{(\ell-m)!}\,\delta_{\ell\ell'}\,. \tag{4.111}$$

Recalling that in the separation of variables discussed in section 4 we had considered factorised solutions of the form $\phi(r,\theta,\varphi) = r^{-1}\,R(r)\,Y(\theta,\varphi)$, and that $Y(\theta,\varphi)$ was itself factorised in the form $\Theta(\theta)\Phi(\varphi)$ with $\Phi(\varphi) \sim e^{\mathrm{i}m\varphi}$ and $\Theta(\theta)$ satisfying the associated Legendre equation, we see that $Y(\theta,\varphi)$ will be of the general form $P_\ell^m(\cos\theta)e^{\mathrm{i}m\varphi}$. To be precise, we shall define

$$Y_{\ell m}(\theta,\varphi) \equiv \sqrt{\frac{(2\ell+1)}{4\pi}}\,\sqrt{\frac{(\ell-m)!}{(\ell+m)!}}\,P_\ell^m(\cos\theta)\,e^{\mathrm{i}\,m\,\varphi}\,, \qquad \ell \geq 0\,, \quad -\ell \leq m \leq \ell\,. \tag{4.112}$$

The *Spherical Harmonics* $Y_{\ell m}(\theta,\varphi)$ satisfy

$$-\nabla^2_{(\theta,\varphi)}\,Y_{\ell m}(\theta,\varphi) = \ell\,(\ell+1)\,Y_{\ell m}(\theta,\varphi)\,. \tag{4.113}$$

These spherical harmonics form the complete set of regular solutions of $\nabla^2_{(\theta,\varphi)}\,Y = -\lambda\,Y$ on the unit sphere. Note from (4.106) that we have

$$Y_{\ell,-m}(\theta,\varphi) = (-1)^m\,\overline{Y}_{\ell m}(\theta,\varphi)\,, \tag{4.114}$$

where the bar denotes complex conjugation.

As we shall see now, the spherical harmonics satisfy the orthogonality properties

$$\int d\Omega\,\overline{Y}_{\ell' m'}(\theta\,\varphi)\,Y_{\ell m}(\theta,\varphi) = \delta_{\ell\ell'}\,\delta_{mm'}\,, \tag{4.115}$$

where

$$d\Omega \equiv \sin\theta\,d\theta\,d\varphi \tag{4.116}$$

is the area element on the unit sphere, and $\int d\Omega\,X$ means

$$\int_0^{2\pi} d\varphi \int_0^{\pi} d\theta\,\sin\theta\,X\,. \tag{4.117}$$

Thus (4.115) says that the integral on the left-hand side is zero unless $\ell' = \ell$ and $m' = m$. The calculations goes as follows. Using the definition in eqn (4.112) for the spherical

harmonics, we have

$$\int d\Omega\, \overline{Y}_{\ell' m'}(\theta\,\varphi)\, Y_{\ell m}(\theta,\varphi) = \sqrt{\frac{(2\ell+1)}{4\pi}}\,\sqrt{\frac{(\ell-m)!}{(\ell+m)!}}\,\sqrt{\frac{(2\ell'+1)}{4\pi}}\,\sqrt{\frac{(\ell'-m')!}{(\ell'+m')!}}\,\times \int_0^\pi d\theta\, \sin\theta\, P_\ell^m(\cos\theta)\, P_{\ell'}^{m'}(\cos\theta)\, \int_0^{2\pi} d\varphi\, e^{\mathrm{i}\,(m-m')\,\varphi}\,. \quad (4.118)$$

Since

$$\int_0^{2\pi} d\varphi\, e^{\mathrm{i}\,(m-m')\,\varphi} = 2\pi\,\delta_{mm'}\,, \quad (4.119)$$

(that is, this integral is zero if the integers m and m' are unequal, and it equals 2π if m and m' are equal), it follows that eqn (4.118) becomes

$$\int d\Omega\, \overline{Y}_{\ell' m'}(\theta\,\varphi)\, Y_{\ell m}(\theta,\varphi) = \sqrt{\frac{(2\ell+1)}{4\pi}}\,\sqrt{\frac{(\ell-m)!}{(\ell+m)!}}\,\sqrt{\frac{(2\ell'+1)}{4\pi}}\,\sqrt{\frac{(\ell'-m)!}{(\ell'+m)!}}\,\times \int_0^\pi d\theta\, \sin\theta\, P_\ell^m(\cos\theta)\, P_{\ell'}^{m}(\cos\theta)\, 2\pi\,\delta_{mm'}\,. \quad (4.120)$$

Then, using the result in eqn (4.111) for the orthogonality of the associated Legendre functions, this gives the final result in (4.115).

Note that the integration over φ, which we performed first, gave the result that the answer would be non-zero only when $m' = m$, which is why we were able to replace $P_{\ell'}^{m'}(\cos\theta)$ by $P_{\ell'}^{m}(\cos\theta)$ in eqn (4.120), and then we could make use of the orthogonality result that we already obtained in eqn (4.111). If we had tried performing the integrations in eqn (4.118) in the other order, first integrating over θ, we would have had a lot more work to do, because we would then have needed to know the results for the integrals

$$\int_0^\pi d\theta\, \sin\theta\, P_\ell^m(\cos\theta)\, P_{\ell'}^{m'}(\cos\theta) \quad (4.121)$$

where m and m' are unequal. This would in fact have been a huge waste of time and effort, because having laboured to evaluate all the integrals (4.121), all the hard-earned results for the cases when $m \neq m'$ would get thrown away at the second stage, when the integral over φ was performed.

It is instructive to look at the first few spherical harmonics explicitly. From (4.112), and

using (4.101) to give the expressions for the P_ℓ^m, we find

$$\begin{aligned}
Y_{0,0}(\theta,\varphi) &= \frac{1}{\sqrt{4\pi}}\,,\\
\\
Y_{1,1}(\theta,\varphi) &= -\sqrt{\frac{3}{8\pi}}\,\sin\theta\,e^{\mathrm{i}\,\varphi}\,,\\
Y_{1,0}(\theta,\varphi) &= \sqrt{\frac{3}{4\pi}}\,\cos\theta\,,\\
Y_{1,-1}(\theta,\varphi) &= \sqrt{\frac{3}{8\pi}}\,\sin\theta\,e^{-\mathrm{i}\,\varphi}\,,\\
\\
Y_{2,2}(\theta,\varphi) &= \sqrt{\frac{15}{32\pi}}\,\sin^2\theta\,e^{2\mathrm{i}\,\varphi}\,,\\
Y_{2,1}(\theta,\varphi) &= -\sqrt{\frac{15}{8\pi}}\,\sin\theta\,\cos\theta\,e^{\mathrm{i}\,\varphi}\,,\\
Y_{2,0}(\theta,\varphi) &= \sqrt{\frac{5}{16\pi}}\,(3\cos^2\theta-1)\,,\\
Y_{2,-1}(\theta,\varphi) &= \sqrt{\frac{15}{8\pi}}\,\sin\theta\,\cos\theta\,e^{-\mathrm{i}\,\varphi}\,,\\
Y_{2,-2}(\theta,\varphi) &= \sqrt{\frac{15}{32\pi}}\,\sin^2\theta\,e^{-2\mathrm{i}\,\varphi}\,.
\end{aligned}\tag{4.122}$$

It is also instructive to rewrite the spherical harmonics in terms of Cartesian, rather than spherical polar, coordinates. Recall that the two coordinate systems are related by

$$x = r\,\sin\theta\,\cos\varphi\,,\quad y = r\,\sin\theta\,\sin\varphi\,,\quad z = r\,\cos\theta\,.\tag{4.123}$$

We can write the expressions for x and y more succinctly in a single complex equation,

$$x + \mathrm{i}\,y = r\,\sin\theta\,e^{\mathrm{i}\,\varphi}\,,\tag{4.124}$$

since we have the well-known result that $e^{\mathrm{i}\,\varphi} = \cos\varphi + \mathrm{i}\,\sin\varphi$. Thus for the spherical

harmonics listed in (4.122) we have

$$\begin{aligned}
Y_{0,0} &= \frac{1}{\sqrt{4\pi}}\,,\\
Y_{1,1} &= -\sqrt{\frac{3}{8\pi}}\,\frac{x+\mathrm{i}\,y}{r}\,,\\
Y_{1,0} &= \sqrt{\frac{3}{4\pi}}\,\frac{z}{r}\,,\\
Y_{1,-1} &= \sqrt{\frac{3}{8\pi}}\,\frac{x-\mathrm{i}\,y}{r}\,,\\
Y_{2,2} &= \sqrt{\frac{15}{32\pi}}\,\frac{(x+\mathrm{i}\,y)^2}{r^2}\,,\\
Y_{2,1} &= -\sqrt{\frac{15}{8\pi}}\,\frac{z\,(x+\mathrm{i}\,y)}{r^2}\,,\\
Y_{2,0} &= \sqrt{\frac{5}{16\pi}}\,\Big(\frac{3z^2}{r^2}-1\Big) = \sqrt{\frac{5}{16\pi}}\,\frac{2z^2-x^2-y^2}{r^2}\,,\\
Y_{2,-1} &= \sqrt{\frac{15}{8\pi}}\,\frac{z\,(x-\mathrm{i}\,y)}{r^2}\,,\\
Y_{2,-2} &= \sqrt{\frac{15}{32\pi}}\,\frac{(x-\mathrm{i}\,y)^2}{r^2}\,.
\end{aligned} \tag{4.125}$$

What we are seeing here is that for each value of ℓ, we are getting a set of functions, labelled by m with $-\ell \le m \le \ell$, that are all of the form of polynomials $\mathcal{Q}_\ell$ of degree ℓ in the Cartesian coordinates (x, y, z), divided by r^ℓ:

$$Y_{\ell m} \sim \frac{\mathcal{Q}_\ell}{r^\ell}\,. \tag{4.126}$$

The larger ℓ is, the larger the number of possible such polynomials. Looking at $\ell = 1$, we have in total three $Y_{1,m}$ functions, which could be reorganised, by taking appropriate linear combinations, as

$$\frac{x}{r}\,, \qquad \frac{y}{r}\,, \qquad \frac{z}{r}\,. \tag{4.127}$$

Thus once we know one of them, the other two just correspond to rotating the coordinate system through 90 degrees about one or another axis. The same is true of all the higher harmonics too. The spherical harmonics thus have built into them the "knowledge" of the rotational symmetries of the sphere. Our procedure for deriving the spherical harmonics was completely non-transparent, in the sense that no explicit use of the rotational symmetries of the sphere was made in the derivation. But at the end of the day, we see that the harmonics we have obtained do indeed exhibit these symmetries. In the language of group theory, one says that the spherical harmonics $Y_{\ell m}$ fall into representations of the rotation group. One of the rather remarkable "miracles" that we encountered during our derivation, namely that

the solutions to the associated Legendre equation could be constructed from solutions of the ordinary Legendre equation, ultimately has its explanation in the fact that the harmonics $Y_{\ell m}$ with $m \neq 0$ are simply related to the $m = 0$ harmonic $Y_{\ell 0}$ by symmetry rotations of the sphere.

4.8 General solution of Laplace's equation without azimuthal symmetry

We have now established that the most general solution of Laplace's equation in spherical polar coordinates[33] can be written as

$$\phi(r,\theta,\varphi) = \sum_{\ell\geq 0} \sum_{m=-\ell}^{\ell} (A_{\ell m}\, r^{\ell} + B_{\ell m}\, r^{-\ell-1})\, Y_{\ell m}(\theta,\varphi)\,. \tag{4.128}$$

The constants $A_{\ell m}$ and $B_{\ell m}$, which depend on both ℓ and m, are as yet arbitrary. Their values are determined by boundary conditions, as in the previous potential-theory examples that we have looked at. Because we are now allowing the azimuthal separation constant m to be non-zero, the class of solutions described by (4.128) includes those that are dependent on the azimuthal angle φ.

In a boundary-value problem where the potential $\phi(r,\theta,\varphi)$ is specified on a spherical boundary surface at $r = a$, one simply uses the orthogonality conditions (4.115) in order to determine the coefficients $A_{\ell m}$ and $B_{\ell m}$ in the general solution (4.128). For example, suppose we are solving for ϕ in the exterior region $r > a$. This means that $A_{\ell m} = 0$, and so, setting $r = a$, multiplying by $\overline{Y}_{\ell',m'}(\theta,\varphi)$ and integrating, we shall have

$$\begin{aligned}
\int \phi(a,\theta,\varphi)\overline{Y}_{\ell',m'}(\theta,\varphi)d\Omega &= \sum_{\ell\geq 0}\sum_{m=-\ell}^{\ell} B_{\ell m}\, a^{-\ell-1} \int Y_{\ell m}(\theta,\varphi)\overline{Y}_{\ell',m'}(\theta,\varphi)d\Omega \\
&= \sum_{\ell\geq 0}\sum_{m=-\ell}^{\ell} B_{\ell m}\, a^{-\ell-1}\, \delta_{\ell,\ell'}\, \delta_{m,m'} \\
&= B_{\ell',m'}\, a^{-\ell'-1}\,,
\end{aligned} \tag{4.129}$$

and so we have

$$B_{\ell m} = a^{\ell+1} \int \phi(a,\theta,\varphi)\overline{Y}_{\ell m}(\theta,\varphi)d\Omega\,. \tag{4.130}$$

It is sometimes useful to consider an expansion of a potential in spherical harmonics even when the problem is not specified in terms of boundary values on a spherical surface. Typically, expansions in spherical harmonics can be useful whenever one wants to express as

[33] That is, the most general solution that is regular on the spherical surfaces at constant r.

potential in the form of a large-r expansion, or a small-r expansion. Consider, for example, our discussion of the Green function for an infinite planar surface, in section 2.4.1. Equation (2.43) gives the expression for $\phi(x,y,z)$ everywhere in the half-space $z>0$ in terms of its boundary value on the plane $z=0$. Suppose we take the boundary potential to be equal to V (a constant) within the square of side 2 whose vertices are located at $x=\pm1$, $y=\pm1$, and take the potential to be zero everywhere outside this square in the $z=0$ plane. Thus, the potential for $z>0$ is given by

$$\phi(x,y,z) = \frac{Vz}{2\pi}\int_{-1}^{1} dx' \int_{-1}^{1} dy' \, \frac{1}{[(x-x')^2+(y-y')^2+z^2]^{3/2}}\,. \tag{4.131}$$

Although this can be integrated explicitly, the result is rather complicated. Instead, let us write x, y and z in terms of spherical polar coordinates, expand the integrand in inverse powers of r, and integrate term by term. (Note that we shall stick with the Cartesian coordinates x' and y' for the integration variables. This is much more convenient, since the region of integration (the square) is much more easily parameterised using Cartesian coordinates.) After some algebra, which is easily done using *Mathematica*, we find

$$\begin{aligned}\phi(r,\theta,\varphi) &= \frac{2cV}{\pi r^2} + \frac{(3c-5c^3)V}{\pi r^4} + \frac{7c(15-70c^2+63c^4-9s^4\cos4\varphi)V}{24\pi r^6} \\ &\quad + \frac{3c(35-315c^2+693c^4-429c^6+11s^4(13c^2-3)\cos4\varphi)V}{16\pi r^8} + \cdots\,,\end{aligned} \tag{4.132}$$

where we have defined $c=\cos\theta$ and $s=\sin\theta$.

Of course, since ϕ satisfies Laplace's equation it must necessarily be possible to express it in a large-r expansion in spherical harmonics, of the form

$$\phi(r,\theta,\varphi) = \sum_{\ell\ge0}\sum_{m=-\ell}^{\ell} B_{\ell m}\, r^{-\ell-1}\, Y_{\ell m}(\theta,\varphi)\,. \tag{4.133}$$

From the definition of $Y_{\ell m}(\theta,\varphi)$, it is not hard to express (4.132) in such a series, giving

$$\begin{aligned}\phi(r,\theta,\varphi) &= \frac{4V}{\sqrt{3\pi}}\frac{1}{r^2}\,Y_{1,0}(\theta,\varphi) - \frac{4V}{\sqrt{7\pi}}\frac{1}{r^4}\,Y_{3,0}(\theta,\varphi) \\ &\quad + \frac{V}{r^6}\left(\frac{14}{3\sqrt{11\pi}}Y_{5,0}(\theta,\varphi) - \sqrt{\frac{14}{55\pi}}\big(Y_{5,4}(\theta,\varphi)+Y_{5,-4}(\theta,\varphi)\big)\right) \\ &\quad + \frac{V}{r^8}\left(-2\sqrt{\frac{3}{5\pi}}Y_{7,0}(\theta,\varphi) + \sqrt{\frac{22}{35\pi}}\big(Y_{7,4}(\theta,\varphi)+Y_{7,-4}(\theta,\varphi)\big)\right) + \cdots\end{aligned} \tag{4.134}$$

Observe that the expression for the potential is azimuthally symmetric at the first couple of orders in inverse powers of $1/r$. That is, only the spherical harmonics of the form $Y_{\ell\,0}$, with $m=0$, arise in the $1/r^2$ and $1/r^4$ terms in the expansion. The non-azimuthal symmetry,

which must of course necessarily be present since the original boundary value specification is not azimuthally-symmetric because it involves a square of non-zero potential at $z = 0$, sets in at orders $1/r^6$ and above.

Observe also that the potential exhibits the discrete four-fold rotational symmetry that we should expect, given that the boundary value specification itself has a discrete four-fold rotational symmetry. In other words, the potential must exhibit the property of quarter-rotational symmetry:

$$\phi(r, \theta, \varphi) = \phi(r, \theta, \varphi + \tfrac{1}{2}\pi)\,. \tag{4.135}$$

Note that this means that the answer *must* involve only those spherical harmonics $Y_{\ell m}(\theta, \varphi)$ for which $Y_{\ell m}(\theta, \varphi) = Y_{\ell m}(\theta, \varphi + \frac{1}{2}\pi)$, and hence $e^{\mathrm{i}m\varphi} = e^{\mathrm{i}m(\varphi+\frac{1}{2}\pi)}$, which implies $e^{\frac{1}{2}\mathrm{i}m\pi} = 1$ and hence

$$m = 4n\,, \qquad n = \text{ integer}\,. \tag{4.136}$$

Since the spherical harmonic $Y_{\ell m}(\theta, \varphi)$ must have $|m| \le \ell$, this means that it would be *impossible* for any azimuthal dependence to appear before order $1/r^5$ at the absolute earliest. In fact, because other symmetries of the problem mean that only *even* inverse powers of r occur here, the azimuthal dependence is deferred until order $1/r^6$.

It is worth emphasising at this point that a new feature that arises when there is no azimuthal symmetry is that the technique of off-axis extrapolation, which was so useful for azimuthally-symmetric problems, no longer works. The reason for this is easily understood. When there is no azimuthal symmetry the general solution takes the form

$$\phi(r, \theta, \varphi) = \sum_{\ell \ge 0} \sum_{m=-\ell}^{\ell} \Bigl(A_{\ell m}\, r^{\ell} + B_{\ell m}\, r^{-\ell-1} \Bigr)\, Y_{\ell m}(\theta, \varphi)\,, \tag{4.137}$$

and so there are $2 \times (2\ell + 1)$ constants, $A_{\ell m}$ and $B_{\ell m}$, for each value of ℓ in the expansion. But, as we saw previously, on the z axis there are only 2 terms in the expansion for $\phi(z)$, for each value of ℓ:

$$\phi(z) = \sum_{\ell \ge 0} (a_\ell\, z^{\ell} + b_\ell\, z^{-\ell-1})\,. \tag{4.138}$$

Thus the information available in an on-axis expansion is insufficient to determine all the coefficients in the general expansion (4.137). Explicitly, if we take the expression (4.137) and restrict it to the z axis (consider just the positive z axis, for simplicity), we have to set $\theta = 0$, whereupon $r = z$. Now the associated Legendre function $P_\ell^m(\cos\theta)$ has a factor $(\sin\theta)^{|m|}$, which vanishes at $\theta = 0$ (or $\theta = \pi$) when $m \neq 0$, and so restricting to the positive

z axis (4.137) becomes

$$\phi(z) = \phi(z, 0, \varphi) = \sum_{\ell \geq 0} \left(A_{\ell 0}\, z^\ell + B_{\ell 0}\, z^{-\ell-1} \right) \sqrt{\frac{2\ell+1}{4\pi}}\,. \tag{4.139}$$

(See eqn (4.112).) Thus one can determine the expansion coefficients $A_{\ell 0}$ and $B_{\ell 0}$ in (4.137) by matching with the solution on the z axis, but the coefficients $A_{\ell m}$ and $B_{\ell m}$ for which $m \neq 0$ cannot be determined by this means.

4.9 Another look at the generating function

We now return to the generating function for the Legendre polynomials, defined in (4.41). There is a nice physical interpretation of this construction, which we shall now describe.

Consider the problem of a point charge of unit strength, sitting on the z axis at a point $z = r'$. We know, since it is an axially-symmetric situation, that the potential at an arbitrary point must be expressible in the form (4.67)

$$\phi(r, \theta) = \sum_{\ell \geq 0} (A_\ell\, r^\ell + B_\ell\, r^{-\ell-1})\, P_\ell(\cos\theta)\,. \tag{4.140}$$

To determine the coefficients, we must first make a choice between considering the expansion either in the region where $r > r'$, or in the region where $r < r'$.

For $r > r'$, the solution should be an infinite series in inverse powers of r, so that it dies off at infinity. Thus for $r > r'$ we must have $A_\ell = 0$. On the other hand, for $r < r'$ the solution should be an infinite series in positive powers of r, so that it remains regular at $r = 0$. For $r < r'$, therefore, we must have $B_\ell = 0$.

The non-zero coefficients B_ℓ or A_ℓ in the two regions can be determined by the method we discussed earlier, of first finding the potential when $\vec{r}$ is on the z axis. This is easy; we shall have

$$\begin{aligned} z > r': \qquad \phi &= \frac{1}{z - r'} = \frac{1}{z}\left(1 - \frac{r'}{z}\right)^{-1} = \sum_{\ell \geq 0} \frac{r'^\ell}{z^{\ell+1}}\,, \\ z < r': \qquad \phi &= \frac{1}{r' - z} = \frac{1}{r'}\left(1 - \frac{z}{r'}\right)^{-1} = \sum_{\ell \geq 0} \frac{z^\ell}{r'^{\ell+1}}\,, \end{aligned} \tag{4.141}$$

The general off-axis solution, where $\vec{r}$ is arbitrary, is therefore given by

$$\phi(r, \theta) = \sum_{\ell \geq 0} \frac{r'^\ell}{r^{\ell+1}}\, P_\ell(\cos\theta)\,, \qquad r > r'\,, \tag{4.142}$$

$$\phi(r, \theta) = \sum_{\ell \geq 0} \frac{r^\ell}{r'^{\ell+1}}\, P_\ell(\cos\theta)\,, \qquad r < r'\,. \tag{4.143}$$

These expressions give the potential at a point (r, θ, φ) due to a unit point charge at the point $z = r'$ on the z axis. The answer is, of course, independent of the azimuthal angle φ because the point charge is on the z axis.

Note that it is sometimes convenient to write the two expressions in (4.142) and (4.143) in the form of a single equation

$$\phi(r, \theta) = \sum_{\ell \geq 0} \frac{r_<^\ell}{r_>^{\ell+1}} P_\ell(\cos\theta)\,, \tag{4.144}$$

where $r_<$ means whichever of r and r' is the smaller, and $r_>$ means whichever of r and r' is the larger.

We can relate these results to the generating function, by observing that we can in fact write down the solution to this problem in closed form. The potential $\phi(r, \theta)$ will just be the inverse of the distance from the point (r, θ) to the point $z = r'$ on the z axis where the unit charge is located. Using the cosine rule, this distance is $(r^2 - 2r\,r'\,\cos\theta + r'^2)^{1/2}$, and so

$$\phi(r, \theta) = \frac{1}{(r^2 - 2r\,r'\,\cos\theta + r'^2)^{1/2}}\,. \tag{4.145}$$

But when $r > r'$, we have just obtained the expansion (4.142) for the potential, and so we have

$$\phi(r, \theta) = \frac{1}{r[1 - 2(r'/r)\,\cos\theta + (r'/r)^2]^{1/2}} = \sum_{\ell \geq 0} \frac{r'^\ell}{r^{\ell+1}} P_\ell(\cos\theta)\,. \tag{4.146}$$

Letting $t = r'/r$, we therefore have

$$(1 - 2t\,\cos\theta + t^2)^{-1/2} = \sum_{\ell \geq 0} t^\ell\, P_\ell(\cos\theta)\,, \tag{4.147}$$

which is nothing but the generating function formula (4.41) with x set equal to $\cos\theta$.

It is straightforward to repeat the above exercise for the region where $r < r'$. This time, we pull out a factor of r' in the denominator of (4.145), and define $t = r/r'$. Equating to (4.143) then again gives rise to the generating function formula (4.41).

The discussions above show how the generating function (4.41) admits a very simple physical interpretation as giving the expansion, in terms of Legendre polynomials, of the potential due to a unit point charge on the z axis.

We may also generalise the discussion, to the case where the unit charge is placed at a general position (r', θ', φ') that need not lie on the z axis. The expression for the potential at (r, θ, φ) is therefore no longer azimuthally symmetric, and so it must be expanded in the general form (4.128), in terms of spherical harmonics.

Of course we could expand the potential simply using Legendre polynomials if we introduce the angle γ between $\vec{r}$ and $\vec{r}\,'$, so that we may write

$$\phi(r,\theta,\varphi) = (r^2 + r'^2 - 2rr'\,\cos\gamma)^{-1/2}\,. \tag{4.148}$$

Using the generating function formula (4.41) we may therefore write

$$\begin{aligned}\frac{1}{|\vec{r}-\vec{r}\,'|} &= \sum_{\ell\ge 0} \frac{r'^\ell}{r^{\ell+1}}\,P_\ell(\cos\gamma)\,, \qquad r > r'\,,\\ \frac{1}{|\vec{r}-\vec{r}\,'|} &= \sum_{\ell\ge 0} \frac{r^\ell}{r'^{\ell+1}}\,P_\ell(\cos\gamma)\,, \qquad r < r'\,.\end{aligned} \tag{4.149}$$

However, as noted earlier the expression for γ in terms of θ, φ, θ' and φ' is actually rather complicated, namely $\cos\gamma = \sin\theta\,\sin\theta'\,\cos(\varphi-\varphi') + \cos\theta\,\cos\theta'$ (see (2.65)).

It will be much more useful to have an expression for $|\vec{r}-\vec{r}\,'|^{-1}$ that is written directly as a series in spherical harmonics. To do this, we first note from (4.128) that in the region $r > r'$, and viewing $|\vec{r}-\vec{r}\,'|^{-1}$ as the potential at $\vec{r}$ due to a unit point charge located at $\vec{r}\,'$, we shall have an expansion of the form

$$\frac{1}{|\vec{r}-\vec{r}\,'|} = \sum_{\ell\ge 0}\sum_{m=-\ell}^{\ell} \frac{1}{r^{\ell+1}}\,B_{\ell m}(r',\theta',\varphi')\,Y_{\ell m}(\theta,\varphi)\,, \tag{4.150}$$

where we have indicated explicitly that the expansion coefficients $B_{\ell m}$ will depend upon the chosen location (r',θ',φ') of the unit charge. On the other hand, in the region $r < r'$ we shall have an expansion of the form

$$\frac{1}{|\vec{r}-\vec{r}\,'|} = \sum_{\ell\ge 0}\sum_{m=-\ell}^{\ell} r^\ell\,A_{\ell m}(r',\theta',\varphi')\,Y_{\ell m}(\theta,\varphi)\,. \tag{4.151}$$

How do we determine the expansion coefficients $A_{\ell m}(r',\theta',\varphi')$ and $B_{\ell m}(r',\theta',\varphi')$? The first thing to notice is that the function $|\vec{r}-\vec{r}\,'|^{-1}$ itself is symmetrical under the exchange of $\vec{r}$ and $\vec{r}\,'$. This means that the way in which r', θ' and φ' enter into the expansions (4.150) and (4.151) should be symmetrically related to the way in which r, θ and φ enter. In saying this, due allowance must be made for the fact that if $r > r'$, then, of course $r' < r$, so the way in which r' appears in (4.150) should be very like the way that r appears in (4.151), and *vice versa*. In fact, roughly speaking, we can expect in symmetry grounds that $B_{\ell m}(r',\theta',\varphi')$ should be something like r'^ℓ multiplied by one of the $Y_{\ell m'}(\theta',\varphi')$ spherical harmonics, and that $A_{\ell m}(r',\theta',\varphi')$ should be something like $r'^{-\ell-1}$ multiplied by some $Y_{\ell m'}(\theta',\varphi')$ spherical harmonic.

At this point, we shall write down the answer, and then prove that it is correct. The result is that

$$\frac{1}{|\vec{r}-\vec{r}\,'|} = \sum_{\ell\geq 0}\sum_{m=-\ell}^{\ell}\frac{4\pi}{2\ell+1}\,\frac{r'^{\ell}}{r^{\ell+1}}\,\overline{Y}_{\ell m}(\theta',\varphi')\,Y_{\ell m}(\theta,\varphi)\,, \qquad r>r'\,, \tag{4.152}$$

$$\frac{1}{|\vec{r}-\vec{r}\,'|} = \sum_{\ell\geq 0}\sum_{m=-\ell}^{\ell}\frac{4\pi}{2\ell+1}\,\frac{r^{\ell}}{r'^{\ell+1}}\,\overline{Y}_{\ell m}(\theta',\varphi')\,Y_{\ell m}(\theta,\varphi)\,, \qquad r<r'\,, \tag{4.153}$$

These expressions can also be written as the single equation

$$\boxed{\frac{1}{|\vec{r}-\vec{r}\,'|} = \sum_{\ell\geq 0}\sum_{m=-\ell}^{\ell}\frac{4\pi}{2\ell+1}\,\frac{r_<^{\ell}}{r_>^{\ell+1}}\,\overline{Y}_{\ell m}(\theta',\varphi')\,Y_{\ell m}(\theta,\varphi)\,.} \tag{4.154}$$

where, as before, $r_<$ means the lesser of r and r', while $r_>$ means the greater of r and r'.

To prove that (4.154) is true, we shall follow a strategy that makes use of the uniqueness theorem. We know that $|\vec{r}-\vec{r}\,'|^{-1}$ obeys the Poisson equation

$$\nabla^2\,\frac{1}{|\vec{r}-\vec{r}\,'|} = -4\pi\,\delta^3(\vec{r}-\vec{r}\,')\,, \tag{4.155}$$

and that $|\vec{r}-\vec{r}\,'|^{-1}$ goes to zero as $\vec{r}$ goes to infinity. Invoking the uniqueness theorem, we know that $|\vec{r}-\vec{r}\,'|^{-1}$ is the unique function that obeys this equation with this boundary condition. Our strategy for proving (4.154) will be to show that the expression on the right-hand side also obeys this same Poisson equation, with the same delta-function source, and that it goes to zero at infinity. By the uniqueness theorem, it will therefore follow that the right-hand side of (4.154) must in fact be equal to the function $|\vec{r}-\vec{r}\,'|^{-1}$.

We shall first need to establish a couple of preliminary results. First, we note that since the spherical harmonics form, by construction, a complete set of functions on the unit sphere, we can expand an arbitrary smooth function $f(\theta,\varphi)$ defined on the sphere in the form of a generalised Fourier expansion

$$f(\theta,\varphi) = \sum_{\ell\geq 0}\sum_{m=-\ell}^{\ell} b_{\ell m}\,Y_{\ell m}(\theta,\varphi)\,. \tag{4.156}$$

Using the orthonormality relation (4.115) for the spherical harmonics, we find

$$b_{\ell m} = \int d\Omega\,\overline{Y}_{\ell m}(\theta,\varphi)\,f(\theta,\varphi)\,, \tag{4.157}$$

and hence substituting this back into (4.156), after first changing variables from (θ,φ) to (θ',φ') in (4.157) to avoid a clash with the (θ,φ) variables in (4.156), we obtain

$$f(\theta,\varphi) = \int d\Omega' f(\theta',\varphi')\,\left(\sum_{\ell\geq 0}\sum_{m=-\ell}^{\ell}\overline{Y}_{\ell m}(\theta',\varphi')Y_{\ell m}(\theta,\varphi)\right)\,. \tag{4.158}$$

The quantity enclosed in the large parentheses is evidently behaving just like a 2-dimensional Dirac delta function, and we shall have

$$\sum_{\ell\geq 0}\sum_{m=-\ell}^{\ell} \overline{Y}_{\ell m}(\theta',\varphi')Y_{\ell m}(\theta,\varphi) = \frac{\delta(\theta-\theta')\delta(\varphi-\varphi')}{\sin\theta'}\,. \tag{4.159}$$

(The $\sin\theta'$ in the denominator is needed to cancel the $\sin\theta'$ factor in the area element $d\Omega' = \sin\theta'\,d\theta' d\varphi'$.) The expression (4.159) is called the *Completeness Relation* for the spherical harmonics. It signifies the fact that the spherical harmonics comprise a complete set of functions on the sphere, in terms of which one can always expand any smooth function, as in (4.156).

Secondly, it is useful to introduce the *Heaviside function* $\vartheta(x)$, often called the step function or theta-function, whose definition is

$$\begin{aligned} \vartheta(x) &= 0\,, \qquad x<0\,,\\ \vartheta(x) &= 1\,, \qquad x>0\,, \end{aligned} \tag{4.160}$$

The discontinuous jump as x passes through zero implies that there is an infinite spike in the derivative of $\vartheta(x)$ at $x=0$, and in fact

$$\vartheta'(x) = \delta(x)\,. \tag{4.161}$$

This can be verified by integrating $\vartheta'(x)$ over the interval $x_1 \leq x \leq x_2$:

$$\int_{x_1}^{x_2} \vartheta'(x)dx = \Big[\vartheta(x)\Big]_{x_1}^{x_2}. \tag{4.162}$$

From the definition (4.160), we see that this is equal to 1 if $x=0$ lies in the integration range, but it is instead 0 if $x=0$ lies outside the integration range. The integral of $\delta(x)$ has exactly the same features.

Using the Heaviside function, the two expressions (4.152) and (4.153) for $|\vec{r}-\vec{r}\,'|^{-1}$ can be combined into the single formula:

$$\frac{1}{|\vec{r}-\vec{r}\,'|} = \sum_{\ell\geq 0}\sum_{m=-\ell}^{\ell} \frac{4\pi}{2\ell+1}\left(\vartheta(r'-r)\,\frac{r^\ell}{r'^{\ell+1}} + \vartheta(r-r')\,\frac{r'^\ell}{r^{\ell+1}}\right)\overline{Y}_{\ell m}(\theta',\varphi')\,Y_{\ell m}(\theta,\varphi)\,. \tag{4.163}$$

We are now in a position to verify that (4.163) is indeed correct, by verifying that it does indeed satisfy the Poisson equation (4.155). As discussed above, this, together with the fact

that the right-hand side of (4.163) obviously satisfies the proper boundary condition that it goes to zero as r goes to infinity, pins down the Green function $|\vec{r}-\vec{r}\,'|^{-1}$ uniquely.

Viewed as a function of r, θ and φ, the right-hand side of (4.163) *by construction* is annihilated by the Laplacian in the bulk, i.e. away form the crossover at $r = r'$. This is obvious from the fact that in each region $r > r'$ and $r < r'$, it is a particular instance of an infinite series of the form (4.128), which by construction is a solution of Laplace's equation. Therefore, when we act with ∇^2 on the right-hand side of (4.163), all the terms involving θ and φ derivatives, together with those from the r derivatives that *do not act on the Heaviside functions*, will add to zero. We therefore need only to retain those terms in which the r derivatives act on the Heaviside functions.

To save some writing, let us temporarily define

$$W_{\ell m} \equiv \frac{4\pi}{2\ell+1}\, \overline{Y}_{\ell m}(\theta',\varphi')\, Y_{\ell m}(\theta,\varphi)\,. \tag{4.164}$$

Recalling that the radial part of $\nabla^2\phi$ can be written as $(\nabla^2\phi)_{\rm radial} = r^{-1}\partial^2(r\phi)/\partial r^2$, we shall have

$$\begin{aligned}
&(\nabla^2\phi)_{\rm radial} = \\
&\frac{1}{r}\,\frac{\partial^2}{\partial r^2}\sum_{\ell\ge 0}\sum_{m=-\ell}^{\ell} W_{\ell m}\left(\vartheta(r'-r)\,\frac{r^{\ell+1}}{r'^{\ell+1}} + \vartheta(r-r')\,\frac{r'^{\ell}}{r^{\ell}}\right)\\
&= \frac{1}{r}\,\frac{\partial}{\partial r}\sum_{\ell\ge 0}\sum_{m=-\ell}^{\ell} W_{\ell m}\left(\frac{d\vartheta(r'-r)}{dr}\,\frac{r^{\ell+1}}{r'^{\ell+1}} + (\ell+1)\vartheta(r'-r)\frac{r^{\ell}}{r'^{\ell+1}} + \frac{d\vartheta(r-r')}{dr}\,\frac{r'^{\ell}}{r^{\ell}} - \ell\vartheta(r-r')\frac{r'^{\ell}}{r^{\ell+1}}\right)\\
&= \frac{1}{r}\,\frac{\partial}{\partial r}\sum_{\ell\ge 0}\sum_{m=-\ell}^{\ell} W_{\ell m}\left(-\delta(r-r')\,\frac{r^{\ell+1}}{r'^{\ell+1}} + (\ell+1)\vartheta(r'-r)\frac{r^{\ell}}{r'^{\ell+1}} + \delta(r-r')\,\frac{r'^{\ell}}{r^{\ell}} - \ell\vartheta(r-r')\frac{r'^{\ell}}{r^{\ell+1}}\right)\\
&= \frac{1}{r}\,\frac{\partial}{\partial r}\sum_{\ell\ge 0}\sum_{m=-\ell}^{\ell} W_{\ell m}\left((\ell+1)\vartheta(r'-r)\,\frac{r^{\ell}}{r'^{\ell+1}} - \ell\vartheta(r-r')\,\frac{r'^{\ell}}{r^{\ell+1}}\right)\\
&= \sum_{\ell\ge 0}\sum_{m=-\ell}^{\ell} W_{\ell m}\Big[\ell(\ell+1)\Big(\vartheta(r'-r)\,\frac{r^{\ell-1}}{r'^{\ell+1}} + \vartheta(r-r')\frac{r'^{\ell}}{r^{\ell+2}}\Big)\\
&\qquad\qquad +(\ell+1)\,\frac{d\vartheta(r'-r)}{dr}\,\frac{r^{\ell}}{r'^{\ell+1}} - \ell\,\frac{d\vartheta(r-r')}{dr}\,\frac{r'^{\ell}}{r^{\ell+1}}\Big]\,.
\end{aligned} \tag{4.165}$$

Note that we have used the results that $d\vartheta(r-r')/dr = \delta(r-r')$ and $d\vartheta(r'-r)/dr = -\delta(r-r')$, and that since $\delta(r-r')$ is non-zero only when $r = r'$, then in any function of r and r' that is multiplied by $\delta(r-r')$, we can always set $r' = r$. Thus, for example, $\delta(r-r')r'^{\ell}/r^{\ell} = \delta(r-r')$.

The terms on the first line of the last expression here are just the usual "bulk" terms, with no derivatives on the Heaviside functions, which will cancel againts the terms from the angular derivatives in ∇^2, as we mentioned above. It is the second line of the last expression that we are after. This second line then becomes

$$\begin{aligned}
&= \frac{1}{r}\sum_{\ell\geq 0}\sum_{m=-\ell}^{\ell} W_{\ell m}\left(-(\ell+1)\delta(r-r')\frac{r^\ell}{r'^{\ell+1}} - \ell\delta(r-r')\frac{r'^\ell}{r^{\ell+1}}\right)\\
&= -\frac{1}{r^2}\sum_{\ell\geq 0}\sum_{m=-\ell}^{\ell} W_{\ell m}(2\ell+1)\delta(r-r')\\
&= -\frac{4\pi}{r^2}\,\delta(r-r')\sum_{\ell\geq 0}\sum_{m=-\ell}^{\ell}\overline{Y}_{\ell m}(\theta',\varphi')\,Y_{\ell m}(\theta,\varphi)\\
&= -4\pi\,\frac{\delta(r-r')\delta(\theta-\theta')\delta(\varphi-\varphi')}{r^2\sin\theta}\\
&= -4\pi\delta^3(\vec{r}-\vec{r}\,')\,. \qquad (4.166)
\end{aligned}$$

Note that we used the completeness relation (4.159) in getting from the third to the fourth line in eqn (4.159).

To summarise, we have shown that indeed $|\vec{r}-\vec{r}\,'|^{-1}$ can be written as (4.152) and (4.153) (or, equivalently, as (4.163)), since we have verified that these infinite sums imply that $|\vec{r}-\vec{r}\,'|^{-1}$ does indeed satisfy (4.155), as it should. It is interesting to see how the delta function on the right-hand side of (4.155) arises in this representation of $|\vec{r}-\vec{r}\,'|^{-1}$ in (4.163). The factors in the θ and φ directions come from the completeness relation (4.159), but the factor in the radial direction arises quite differently. It comes from the fact that the expressions (4.152) and (4.153) in the $r > r'$ and $r < r'$ regions, which are combined into one formula in (4.163), are themselves *equal* at $r = r'$, but there is a discontinuity in the gradient. This in turn implies that there is a delta function in the second derivative with respect to r, and it is this that produces the radial delta function factor.

Observe also that by comparing the expansions (4.149) with those in (4.152) and (4.153), we can deduce that

$$P_\ell(\cos\gamma) = \frac{4\pi}{2\ell+1}\sum_{m=-\ell}^{\ell}\overline{Y}_{\ell m}(\theta',\varphi')Y_{\ell m}(\theta,\varphi)\,, \qquad (4.167)$$

where, it will be recalled, γ is given by (2.65).

4.10 Dirichlet Green function expansion

The expression (4.163) obtained in the previous section is useful in its own right, in situations where there is a charge distribution $\rho(\vec{r})$ and one wants to calculate the resulting

contribution to the potential as given by (2.13) in the case that there are no boundaries.

We can also make use of (4.163) in order to obtain an expansion for the Dirichlet Green function for a spherical boundary. Recall that we used the method of images in order to construct the Dirichlet Green function (2.60) for the exterior boundary value problem in which the potential is specified on the spherical surface $r = a$, and one wants to determine $\phi(r, \theta, \varphi)$ for all $r \geq a$. For convenience, we give the expression (2.60) again here:

$$G_D(\vec{r}, \vec{r}\,') = \frac{1}{|\vec{r} - \vec{r}\,'|} - \frac{a/r'}{|\vec{r} - (a^2/r'^2)\,\vec{r}\,'|}\,. \tag{4.168}$$

It then follows straightforwardly from (4.152) and (4.153) that the expansion for $G_D(\vec{r}, \vec{r}\,')$ in (4.168) will be given by

$$\begin{aligned} G_D(\vec{r}, \vec{r}\,') &= \sum_{\ell \geq 0} \sum_{m=-\ell}^{\ell} \frac{4\pi}{2\ell+1}\, \frac{1}{r^{\ell+1}} \left(r'^{\ell} - \frac{a^{2\ell+1}}{r'^{\ell+1}} \right) \overline{Y}_{\ell m}(\theta', \varphi')\, Y_{\ell m}(\theta, \varphi)\,, \qquad r > r'\,, \\ G_D(\vec{r}, \vec{r}\,') &= \sum_{\ell \geq 0} \sum_{m=-\ell}^{\ell} \frac{4\pi}{2\ell+1}\, \frac{1}{r'^{\ell+1}} \left(r^{\ell} - \frac{a^{2\ell+1}}{r^{\ell+1}} \right) \overline{Y}_{\ell m}(\theta', \varphi')\, Y_{\ell m}(\theta, \varphi)\,, \qquad r < r'\,, \end{aligned} \tag{4.169}$$

(4.169)

Recall from our original discussion of the Green function that for the Dirichlet boundary-value problem, the potential inside a volume V bounded by a surface S is given by (2.29)

$$\phi(\vec{r}) = \int_V \rho(\vec{r}\,')\, G_D(\vec{r}, \vec{r}\,') d^3\vec{r}\,' - \frac{1}{4\pi} \int_S \phi(\vec{r}\,')\, \frac{\partial G_D(\vec{r}, \vec{r}\,')}{\partial n'}\, dS'\,. \tag{4.170}$$

We therefore need to calculate the normal derivative of the upper expression in (4.10) at $r' = a$ (approached from $r' > a$, since we are solving the exterior Dirichlet problem). The normal derivative should be directed *out* of the volume V, which in our case is all of space in the region $r > a$, and so the normal derivative that we require is $(-\partial G_D(\vec{r}, \vec{r}\,')/\partial r')|_{r'=a}$. From the upper expression in (4.10), this is therefore given by

$$-\frac{\partial G_D(\vec{r}, \vec{r}\,')}{\partial r'}\Big|_{r'=a} = -\frac{4\pi}{a^2} \sum_{\ell \geq 0} \sum_{m=-\ell}^{\ell} \left(\frac{a}{r}\right)^{\ell+1} \overline{Y}_{\ell m}(\theta', \varphi') Y_{\ell m}(\theta, \varphi)\,. \tag{4.171}$$

Consider, for simplicity, the case where $\rho = 0$ in the entire region $r > a$. It then follows from (4.170) and (4.171) that $\phi(r, \theta, \varphi)$ outside the surface $r = a$ is given in terms of the boundary values $\phi(a, \theta, \varphi)$ on the sphere by

$$\phi(r, \theta, \varphi) = \sum_{\ell \geq 0} \sum_{m=-\ell}^{\ell} \left(\int \phi(a, \theta', \varphi') \overline{Y}_{\ell m}(\theta', \varphi') d\Omega' \right) \left(\frac{a}{r}\right)^{\ell+1} Y_{\ell m}(\theta, \varphi)\,. \tag{4.172}$$

This can be seen to be contained, as it must be, within the general class of solutions (4.128). It corresponds to $A_{\ell m} = 0$ and

$$B_{\ell m} = a^{\ell+1} \int \phi(a, \theta', \varphi') \overline{Y}_{\ell m}(\theta', \varphi') d\Omega' \,. \tag{4.173}$$

This is indeed the same result that we obtained previously, in (4.130), by a different (but of course equivalent) method.

An analogous discussion can be given for the interior problem, where one solves for $\phi(r, \theta, \varphi)$ for $0 < r < a$ in terms of boundary values $\phi(a, \theta, \varphi)$ on the sphere at $r = a$.

4.11 Inversion symmetry revisited

Before closing the discussion of solving Laplace's equation in spherical polar coordinates, we shall return again to the topic of the inversion symmetry that was discussed in section 4.6. Although the discussion there was for solutions with azimuthal symmetry, the notion of applying inversion in order to map one solution into another can be applied also to cases where there is no azimuthal symmetry.

In a general setting, the idea of inversion can be described as follows. Suppose that $\phi(r, \theta, \varphi)$ is any solution of Laplace's equation. We then define

$$\psi(r, \theta, \varphi) \equiv \frac{a}{r} \, \phi\big(\frac{a^2}{r}, \theta, \varphi\big) \,, \tag{4.174}$$

where a is any constant. It is straightforward to see that ψ is also a solution of Laplace's equation. This can be shown either by direct substitution of (4.174) into the Laplacian, or, perhaps more simply, as follows: Since $\phi(r, \theta, \varphi)$ solves Laplace's equation, we know from the earlier discussions that it can be expanded in spherical harmonics in the form

$$\phi(r, \theta, \varphi) = \sum_{\ell \geq 0} \sum_{m=-\ell}^{\ell} \Big(A_{\ell m} \, r^{\ell} + B_{\ell m} \, r^{-\ell-1} \Big) \, Y_{\ell m}(\theta, \varphi) \,. \tag{4.175}$$

From (4.174) we therefore have

$$\begin{aligned} \psi(r, \theta, \varphi) &= \frac{a}{r} \sum_{\ell \geq 0} \sum_{m=-\ell}^{\ell} \Big(A_{\ell m} a^{2\ell} \, r^{-\ell} + B_{\ell m} \, a^{-2\ell-2} \, r^{\ell+1} \Big) \, Y_{\ell m}(\theta, \varphi) \,, \\ &= \sum_{\ell \geq 0} \sum_{m=-\ell}^{\ell} \Big(A_{\ell m} \, a^{2\ell+1} \, r^{-\ell-1} + B_{\ell m} \, a^{-2\ell-1} \, r^{\ell} \Big) \, Y_{\ell m}(\theta, \varphi) \,. \end{aligned} \tag{4.176}$$

Thus we see that

$$\psi(r, \theta, \varphi) = \sum_{\ell \geq 0} \sum_{m=-\ell}^{\ell} \Big(\widetilde{A}_{\ell m} \, r^{\ell} + \widetilde{B}_{\ell m} \, r^{-\ell-1} \Big) \, Y_{\ell m}(\theta, \varphi) \,, \tag{4.177}$$

where

$$\widetilde{A}_{\ell m} = \frac{B_{\ell m}}{a^{2\ell+1}}\,, \qquad \widetilde{B}_{\ell m} = A_{\ell m}\, a^{2\ell+1}\,. \tag{4.178}$$

Since we know that the right-hand side of (4.175) solves Laplace's equation for *any* values of the constants $A_{\ell m}$ and $B_{\ell m}$, it follows that since ψ in (4.177) is also of this general form, it is necessarily a solution of Laplace's equation. The relations (4.178) are just the natural generalisation to the non-azimuthally symmetric case of relations such as (4.90) that we saw previously when discussing inversion. And indeed, it follows from (4.174) that

$$\psi(a,\theta,\varphi) = \phi(a,\theta,\varphi)\,, \tag{4.179}$$

meaning that ϕ and ψ are equal at the "inversion radius" a. Furthermore, the transformation (4.174) maps a solution ϕ of Laplace's equation in the region $r > a$ into a solution ψ in the region $r < a$, and *vice versa*.

The upshot of the above discussion is that in a case where there exists an inversion symmetry, the mapping (4.174) will map the exterior solution into the interior solution, and *vice versa*.

Let us now specialise to the case where there is azimuthal symmetry. In such cases, as we have seen, the expression for the potential at general points (r,θ) can be derived from the expression on the z axis, by the method of off-axis extrapolation. In these situations, one can actually restate the criterion for the existence of an inversion symmetry in a given problem as a criterion purely on the potential evaluated on the z axis.

Suppose we can solve for the potential in a particular problem on the z axis, finding $\phi_>(z)$ in the region $|z| > a$ and $\phi_<(z)$ in the region $|z| < a$. If there is an inversion symmetry through the radius a then it would follow, from (4.174), and the discussion above, that we should have[34]

$$\phi_<(z) = \frac{a}{z}\,\phi_>\big(\frac{a^2}{z}\big)\,. \tag{4.180}$$

If this relation holds on the z axis then the problem admits an inversion symmetry. If (4.180) does not hold, then there is no inversion symmetry.

As an example where the inversion symmetry does work, consider the problem discussed in section 2.4, where the on-axis solution was obtained for the potential outside a spherical surface at radius a, whose upper hemisphere is held at potential V and whose lower hemisphere is held at potential $-V$. Again, we shall simplify the discussion by considering just

[34]For simplicity, to avoid burdening the discussion with the rather trivial but tiresome complications of considering the negative z axis as well, we shall restrict ourselves in the discussion presented here to the positive z axis. The reader can easily extend this discussion to include the negative z axis if desired.

the region where z is positive, to avoid the trivial but tiresome complications of considering negative z as well. The potential, which we now denote by $\phi_>(z)$, is given by (2.69):

$$\phi_>(z) = V\,\left(1 - \frac{z^2 - a^2}{z\sqrt{z^2 + a^2}}\right)\,. \tag{4.181}$$

Using (4.180), we therefore find that

$$\begin{aligned}\phi_<(z) &= \frac{a}{z}\,V\,\Big[1 - \frac{a^4/z^2 - a^2}{(a^2/z)\,\sqrt{a^4/z^2 + a^2}}\Big]\,,\\ &= V\,\Big[\frac{a}{z} - \frac{a^2 - z^2}{z\,\sqrt{a^2 + z^2}}\Big]\,.\end{aligned} \tag{4.182}$$

One can readily check, using the appropriate Green function for the interior problem, that this is indeed the correct expression for the potential on the z axis in the region $|z| < a$ that is *inside* the spherical shell.

An example where the inversion symmetry fails is the problem of finding the potential due to a thin circular conducting disk of radius a held at potential V.[35] This was explored at length in one of the homeworks. One can indeed easily verify in this case that the on-axis expansions for the potentials $\phi_>(z)$ and $\phi_<(z)$ in the two regions are *not* related by the mapping (4.180).

5 Separation of Variables in Cylindrical Polar Coordinates

Another common situation that arises when considering boundary-value problems in electrostatics is when there is a cylindrical symmetry, in which case cylindrical polar coordinates are typically the most convenient choice. We shall take these to be (ρ, φ, z), where

$$x = \rho\cos\varphi\,, \qquad y = \rho\sin\varphi\,, \qquad z = z\,. \tag{5.1}$$

In other words, we still use z as the coordinate along the Cartesian z axis, but in the (x, y) plane we use polar coordinates (ρ, φ). A straightforward calculation shows that Laplace's equation (3.1) becomes

$$\frac{1}{\rho}\,\frac{\partial}{\partial\rho}\Big(\rho\frac{\partial\phi}{\partial\rho}\Big) + \frac{1}{\rho^2}\,\frac{\partial^2\phi}{\partial\varphi^2} + \frac{\partial^2\phi}{\partial z^2} = 0\,. \tag{5.2}$$

We can separate variables by writing $\phi(\rho, \varphi, z) = R(\rho)\,\Phi(\phi)\,Z(z)$, which leads, after dividing out by ϕ, to

$$\frac{1}{\rho\,R}\,\frac{d}{d\rho}\Big(\rho\frac{dR}{d\rho}\Big) + \frac{1}{\rho^2\,\Phi}\,\frac{d^2\Phi}{d\varphi^2} + \frac{1}{Z}\,\frac{d^2Z}{dz^2} = 0\,. \tag{5.3}$$

[35] The failure of the inversion symmetry in this example is the reason why some purported solutions to this problem that one can find online are incorrect.

The first two terms depend on ρ and φ but not z, whilst the last term depends on z but not ρ and φ. Thus the last term must be a constant, which we shall call k^2, and then

$$\frac{1}{\rho\,R}\,\frac{d}{d\rho}\Big(\rho\frac{dR}{d\rho}\Big)+\frac{1}{\rho^2\,\Phi}\,\frac{d^2\Phi}{d\varphi^2}+k^2=0\,. \tag{5.4}$$

Multiplying by ρ^2, we obtain

$$\frac{\rho}{R}\,\frac{d}{d\rho}\Big(\rho\frac{dR}{d\rho}\Big)+k^2\,\rho^2+\frac{1}{\Phi}\,\frac{d^2\Phi}{d\varphi^2}=0\,. \tag{5.5}$$

The first two terms depend on ρ but not φ, whilst the last term depends on φ but not ρ. We deduce that the last term is a constant, which we shal call $-\nu^2$. The separation process is now complete, and we have

$$\frac{d^2Z}{dz^2}-k^2\,Z=0\,,\qquad \frac{d^2\Phi}{d\varphi^2}+\nu^2\,\Phi=0\,, \tag{5.6}$$

$$\frac{d^2R}{d\rho^2}+\frac{1}{\rho}\,\frac{dR}{d\rho}+\Big(k^2-\frac{\nu^2}{\rho^2}\Big)\,R=0\,, \tag{5.7}$$

where k^2 and ν^2 are separation constants.

The Z equation and Φ equation in (5.6) are easily solved, giving

$$Z(z)\sim e^{\pm kz}\,,\qquad \Phi(\varphi)\sim e^{\pm\mathrm{i}\nu\varphi}\,. \tag{5.8}$$

Usually, we shall be seeking solutions where the potential should be periodic in the polar angle φ in the (x,y) plane, so that $\Phi(\varphi+2\pi)=\Phi(\varphi)$, and hence

$$\nu=\text{ integer}\,. \tag{5.9}$$

However, it will still be useful to think of ν as being a more general real constant for now. The reason for this will emerge when we investigate the solutions of the radial equation (5.7).

Rescaling the radial coordinate by defining $x=k\,\rho$, and renaming R as y (these functions x and y are not to be confused with the original Cartesian coordinates x and y!), the radial equation (5.7) takes the form

$$x^2\,y''(x)+x\,y'(x)+(x^2-\nu^2)\,y=0\,, \tag{5.10}$$

where y' means dy/dx. This is known as *Bessel's Equation*, and we can construct solutions in the form of power-series expansions, by applying Frobenius's method as we did when discussing the Legendre equation.

5.1 Solutions of Bessel's equation

It may be recalled that when seeking power-series solutions of Legendre's equation in the form of expansions around $x = 0$, it was sufficient to consider Taylor expansions in non-negative integer powers of x, since $x = 0$ was an *ordinary point* of the Legendre equation. By contrast, the point $x = 0$ is a *singular point* of the Bessel equation. This can be seen by dividing out (5.10) by x^2 so that the y'' term has unit coefficient, and then noting that the coefficients of y' and y become singular at $x = 0$. Technically, the nature of the behaviour at $x = 0$ implies that it is a *regular singular point*, and the upshot is that we should now seek solutions of the form

$$y(x) = \sum_{n\geq 0} a_n\, x^{n+\sigma}\,, \tag{5.11}$$

where σ is a constant. Substituting into (5.10), we obtain

$$\sum_{n\geq 0}[(n+\sigma)^2 - \nu^2]\, a_n\, x^{n+\sigma} + \sum_{n\geq 0} a_n\, x^{n+\sigma+2} = 0\,. \tag{5.12}$$

Since this must hold for all x, we can now equate to zero the coefficient of each power of x. To do this, in the first sum we make the replacement $n \to n+2$, so that (5.12) is re-expressed as[36]

$$\begin{aligned}\sum_{n\geq 0}\Big\{[(n+\sigma+2)^2 - \nu^2]\, a_{n+2} + a_n\Big\}\, x^{n+2}\\ +(\sigma^2-\nu^2)\, a_0 + [(\sigma+1)^2-\nu^2]\, a_1\, x = 0\,,\end{aligned} \tag{5.13}$$

where we have also extracted out an overall factor of x^σ. Since the coefficient of each power of x in eqn (5.13) must separately vanish, we see that

$$a_{n+2} = \frac{a_n}{\nu^2 - (n+\sigma+2)^2}\,, \tag{5.14}$$

for $n \geq 0$. In addition we have, from the two "extra" terms,

$$(\sigma^2-\nu^2)\, a_0 = 0 \qquad \text{and} \qquad [(\sigma+1)^2-\nu^2]\, a_1 = 0\,. \tag{5.15}$$

We begin with the first equation in (5.15). This is called the *Indicial Equation*. Notice that we can insist, without any loss of generality, that $a_0 \neq 0$. The reason for this is as follows. Suppose a_0 *were* equal to zero. The series (5.11) would then begin with the a_1

[36] Recall that "sending $n \to n+2$ in the first sum" means first setting $n = m+2$, so that the summation over m runs from -2 up to $+\infty$. Then, we write this as the sum from $m = 0$ to $+\infty$ together with the "extra" two terms $m = -2$ and $m = -1$ added on in addition. Finally, we relabel the m summation variable as n.

term, so it would be a series whose powers of x were $(x^{\sigma+1}, x^{\sigma+2}, x^{\sigma+3}, \ldots)$. But since at the stage when we write (5.11) σ is a completely arbitrary constant, not yet determined, we could as well relabel it by writing $\sigma = \sigma' - 1$. We would then have a series whose powers of x are $(x^{\sigma'}, x^{\sigma'+1}, x^{\sigma'+2}, \ldots)$. But this is exactly what we would have had if the a_0 term were in fact non-zero, after relabelling σ' as σ. So insisting that a_0 be non-zero loses no generality at all.

Proceeding, we then have the indical equation $\sigma^2 - \nu^2 = 0$, i.e.

$$\sigma = \pm\nu\,. \tag{5.16}$$

Now we look at the second equation in (5.15). Since we already know from the indicial equation that $\sigma^2 = \nu^2$, we can rewrite the second equation as

$$(2\sigma + 1)\, a_1 = 0\,. \tag{5.17}$$

Thus either $a_1 = 0$ or else $\sigma = -\frac{1}{2}$. But since we already know from the indicial equation that $\sigma = \pm\nu$, it follows that except in the very special cases where $\nu = \pm\frac{1}{2}$, which has to be analysed separately, we must have that $a_1 = 0$. Let us assume that $\nu \neq \pm\frac{1}{2}$, for simplicity. In fact, we shall assume for now that ν takes a generic value, which is not equal to any integer or half integer.

Finally, in the recursion relation (5.14), we substitute the two possible values for σ, i.e. $\sigma = \nu$ or $\sigma = -\nu$. In each of these two cases, the recursion relation then gives us expressions for all the a_n with $n \geq 2$, in terms of a_0 (which is non-zero), and a_1 (which is zero since we are assuming $\nu \neq \pm\frac{1}{2}$).

We can check the radius of convergence of the series solutions, by applying the ratio test. The ratio of successive terms (bearing in mind that $a_1 = 0$, which means all the odd a_n are zero) is given by

$$\frac{a_{n+2}\, x^{n+2+\sigma}}{a_n\, x^{n+\sigma}} = \frac{a_{n+2}\, x^2}{a_n} = \frac{x^2}{\nu^2 - (n+\sigma+2)^2}\,, \tag{5.18}$$

where $\sigma = \pm\nu$. In either case, at large n we see that the absolute value of the ratio tends to x^2/n^2, and thus the ratio becomes zero for any finite x, no matter how large. Thus the radius of convergence is infinite.

To summarise, we have obtained two solutions to the Bessel equation, $y_1(x)$ and $y_2(x)$, given by

$$y_1(x) = x^\nu \sum_{n\geq 0} a_n^+\, x^n\,, \qquad y_2(x) = x^{-\nu} \sum_{n\geq 0} a_n^-\, x^n\,, \tag{5.19}$$

where $a_0^\pm$ is arbitrary, $a_1^\pm = 0$, and from the recursion relation (5.14) we have

$$a^+_{n+2} = \frac{a^+_n}{\nu^2 - (n+\nu+2)^2}\,, \qquad a^-_{n+2} = \frac{a^-_n}{\nu^2 - (n-\nu+2)^2}\,. \tag{5.20}$$

It is straightforward to see that these two recursion relations can be solved, to give the two solutions, called $J_\nu(x)$ and $J_{-\nu}(x)$:

$$J_\nu(x) = \Big(\frac{x}{2}\Big)^\nu \sum_{p\ge 0} \frac{(-1)^p}{p!\,\Gamma(p+\nu+1)} \Big(\frac{x}{2}\Big)^{2p}, \quad J_{-\nu}(x) = \Big(\frac{x}{2}\Big)^{-\nu} \sum_{p\ge 0} \frac{(-1)^p}{p!\,\Gamma(p-\nu+1)} \Big(\frac{x}{2}\Big)^{2p}. \tag{5.21}$$

Here Γ is the Gamma function, which can be defined for $\Re(z) > 0$ by

$$\Gamma(z) = \int_0^\infty e^{-t}\, t^{z-1} dt\,. \tag{5.22}$$

If z is an integer, it is related to the factorial function by $\Gamma(n+1) = n!$. In general, one show from (5.22) by performing an integration by parts that it satisfies the relation

$$z\Gamma(z) = \Gamma(z+1)\,. \tag{5.23}$$

This relation can be used in order to make an analytic extension of $\Gamma(z)$ to the whole complex plane. In particular, one can show that $\Gamma(z)$ is non-singular in the entire finite complex plane, apart from simple poles at all the non-positive integers. Near $z = -N$, where $N = 0,\, 1,\, 2, \cdots$, one can see that $\Gamma(z)$ behaves like

$$\Gamma(z) = \frac{(-1)^N}{N!\,(z+N)} + \text{finite terms}\,. \tag{5.24}$$

The two *Bessel functions* $J_\nu(x)$ and $J_{-\nu}(x)$ are linearly independent, if ν takes a generic real value. This is obvious from the fact that the leading term in $J_\nu(x)$ is proportional to x^ν, while the leading term in $J_{-\nu}(x)$ is proportional to $x^{-\nu}$.

However, if ν is an integer (which, it should be recalled, is generally going to be the case in physical situations where we are solving Laplace's equation for the electrostatic potential), the two Bessel functions become proportional to one another. For example, if $\nu = 1$ we find from (5.21) that

$$J_1(x) = \tfrac{1}{2}x - \tfrac{1}{16}x^3 + \tfrac{1}{384}x^5 + \cdots\,, \qquad J_{-1}(x) = -\tfrac{1}{2}x + \tfrac{1}{16}x^3 - \tfrac{1}{384}x^5 + \cdots\,, \tag{5.25}$$

and in fact $J_1(x) = -J_{-1}(x)$. (One might have thought from the expression for $J_{-1}(x)$ in (5.21) that its leading term would be proportional to $1/x$, but the coefficient of the $p = 0$ term in the summation is $1/\Gamma(0)$, and $\Gamma(0)$ is in fact infinite (see (5.23), with $z = 0$, bearing

in mind that $\Gamma(1) = 0! = 1$).) It is not hard to see from the series expansions (5.21) that when $\nu = n$ is an integer, we shall have

$$J_{-n}(x) = (-1)^n J_n(x)\,. \tag{5.26}$$

Since when ν is an integer the two generically-independent solutions J_ν and $J_{-\nu}$ to Bessel's equation become linearly dependent, it follows that the actual "second solution" cannot be of the originally-assumed form (5.11) when ν is an integer. In fact, what is missing is that the actual second solution acquires a dependence on $\log x$ when ν is an integer.[37]

The second solution can be constructed by applying a limiting procedure. Essentially, we take the linear combination of $J_\nu(x)$ and $J_{-\nu}(x)$ that vanishes when ν approaches an integer, and divide it by a ν-dependent factor that also vanishes as ν approaches an integer. This ratio of "0/0" is actually finite and non-zero, and provides us with the second solution. Thus, we define

$$Y_\nu(x) = \frac{J_\nu(x)\cos\nu\pi - J_{-\nu}(x)}{\sin\nu\pi}\,. \tag{5.27}$$

We may take $J_\nu(x)$ and $Y_\nu(x)$ to be the two linearly-independent solutions of Bessel's equation (5.10) for arbitrary ν, integer or non-integer. (When ν is not an integer, Y_ν is a nonsingular linear combination of J_ν and $J_{-\nu}$, and so it is linearly independent of J_ν.) The function $J_\nu(x)$ is sometimes called the *Bessel function of the first kind*, and $Y_\nu(x)$ is called the *Bessel function of the second kind*.

It is evident from the series expansion (5.21) that at small x, the Bessel function $J_\nu(x)$ behaves like

$$J_\nu(x) = \frac{1}{\Gamma(\nu+1)}\left(\frac{x}{2}\right)^\nu + \mathcal{O}(x^{\nu+2})\,. \tag{5.28}$$

By means of a more intricate analysis, which requires first finding a suitable integral representation for $J_\nu(x)$, one can show that at large x it takes roughly the form of a cosine function, with a slowly decaying amplitude:

$$J_\nu(x) \sim \sqrt{\frac{2}{\pi x}}\,\cos\left(x - \frac{\nu\pi}{2} - \frac{\pi}{4}\right)\,, \qquad x \longrightarrow \infty\,. \tag{5.29}$$

A standard contour integral representation for the Bessel function that is employed in order to establish this is

$$J_\nu(z) = \frac{1}{2\pi\mathrm{i}}\int_C t^{-\nu-1}\, e^{\frac{1}{2}(t-t^{-1})}\, dt\,, \tag{5.30}$$

[37] The way in which this happens can be seen by considering, as a toy example, $f(x) = (x^\nu - x^{-\nu})\,\nu^{-1}$. By writing $x^{\pm\nu} = e^{\pm\nu\log x}$ and expanding $e^{\pm\nu\log x} = 1 \pm \nu\log x + \mathcal{O}(\nu^2)$ when ν is small, we see that in limit $\nu \to 0$ we get $f(x) \to 2\log x$.

where the contour in the complex t plane is taken to start at $t = -\infty$ just below the real t axis, swings around the point $t = 0$ and then head back to $t = -\infty$ just above the real t axis.

The $Y_\nu(x)$ Bessel function has a similar decaying oscillatory behaviour at large x:

$$Y_\nu(x) \sim \sqrt{\frac{2}{\pi x}}\, \sin\Big(x - \frac{\nu\pi}{2} - \frac{\pi}{4}\Big)\,, \qquad x \longrightarrow \infty\,. \tag{5.31}$$

If ν is not an integer its small-x behaviour is just that which is implied by the two series expansions in (5.21), combined according to the definition (5.27). As mentioned previously, $Y_\nu(x)$ involves dependence on $\log x$ if ν is an integer. The small-x behaviour for the first couple of $Y_n(x)$ functions for integer n are:

$$\begin{aligned} Y_0(x) &= \frac{2}{\pi}\Big(\log\tfrac{1}{2}x + \gamma\Big)\Big(1 - \tfrac{1}{4}x^2 + \tfrac{1}{64}x^4 + \cdots\Big) + \frac{1}{2\pi}\Big(x^2 - \tfrac{3}{32}x^4 + \cdots\Big)\,, \\ Y_1(x) &= \frac{1}{\pi}\Big(\log\tfrac{1}{2}x + \gamma\Big)\Big(x - \tfrac{1}{8}x^3 + \tfrac{1}{192}x^5 + \cdots\Big) - \frac{2}{\pi}\Big(\frac{1}{x} + \tfrac{1}{4}x - \tfrac{5}{64}x^3\cdots\Big)\,, \end{aligned} \tag{5.32}$$

where $\gamma = \lim_{n\to\infty}(-\log n + \sum_{k=1}^{n} k^{-1}) \approx 0.5772157$ is the Euler-Mascheroni constant. In general, for non-integer $\nu > 0$, the leading-order small-x behaviour of $Y_\nu(x)$ is of the form

$$Y_\nu(x) \sim -\frac{\Gamma(\nu)}{\pi}\,\left(\frac{2}{x}\right)^\nu\,. \tag{5.33}$$

The three figures below contain plots of the $J_0(x)$, $J_1(x)$ and $J_5(x)$ Bessel functions.

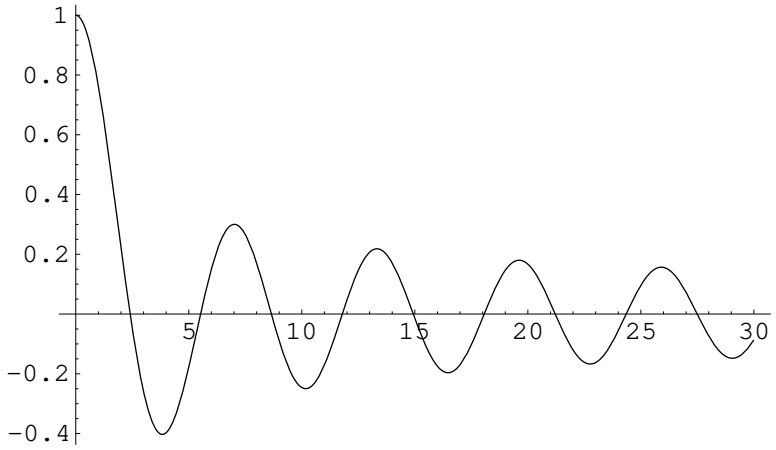

Figure 1: The $J_0(x)$ Bessel Function

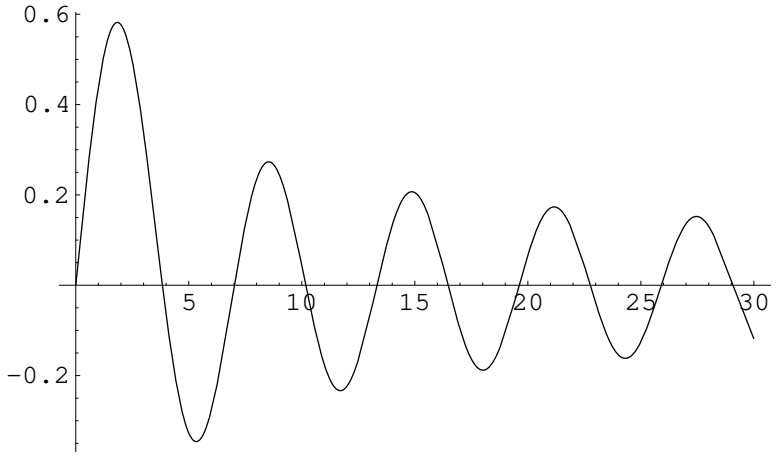

Figure 2: The $J_1(x)$ Bessel Function

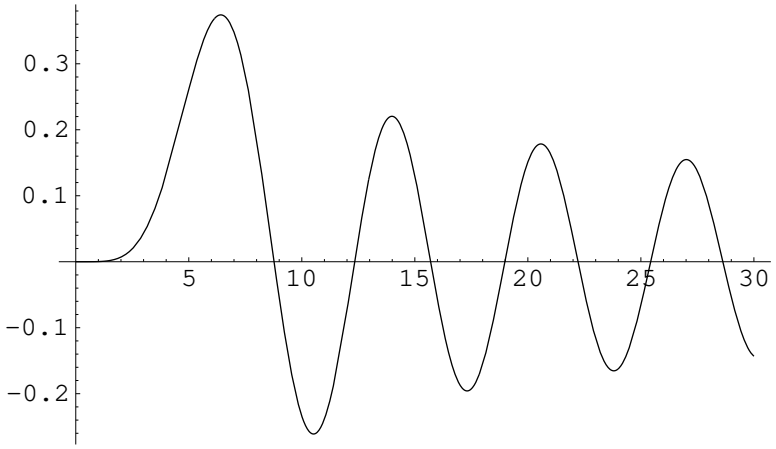

Figure 3: The $J_5(x)$ Bessel Function

5.2 Properties of the Bessel functions

From the asymptotic forms (5.29) and (5.31) of the J_ν and Y_ν Bessel functions, it can be seen to be natural to define also the complex combinations

$$H_\nu^{(1)}(x) = J_\nu(x) + \mathrm{i}Y_\nu(x)\,, \qquad H_\nu^{(2)}(x) = J_\nu(x) - \mathrm{i}Y_\nu(x)\,. \tag{5.34}$$

These are known as the *Hankel functions*, or "Bessel functions of the third kind." They bear the same relation to J_v and Y_ν as $e^{\pm\mathrm{i}\theta}$ does to $\cos\theta$ and $\sin\theta$.

All the Bessel functions can be shown to satisfy the recurrence relations

$$W_{\nu-1}(x) + W_{\nu+1}(x) = \frac{2\nu}{x} W_\nu(x)\,, \tag{5.35}$$

$$W_{\nu-1}(x) - W_{\nu+1}(x) = 2\frac{dW_\nu(x)}{dx}\,, \tag{5.36}$$

where W_ν is taken to be any one of J_ν, Y_ν, $H_\nu^{(1)}$ or $H_\nu^{(2)}$. These relations can be proven directly, using the expression in (5.21) for the series expansion of $J_\nu(x)$.

We have seen from the plots of the J_ν Bessel functions, and from their asymptotic behaviour, that $J_\nu(x)$ has a discrete infinite set of zeros, at points on the x axis that asymptotically approach an equal spacing. Let us say that the m'th zero of $J_\nu(x)$ occurs at

$$x = \alpha_{\nu m}\,, \qquad \text{so} \qquad J_\nu(\alpha_{\nu m}) = 0\,. \tag{5.37}$$

Thus $x = \alpha_{\nu\,1}$ is the location of the first zero, $x = \alpha_{\nu\,2}$ is the location of the second, and so on, as x increases from 0. They occur at definite values of $\alpha_{\nu m}$, though it is not easy to give explicit expressions for $\alpha_{\nu m}$.

Recall that the Bessel equation arose from our separation of variables in cylindrical polar coordinates (ρ, φ, z), and that the independent variable x in the Bessel equation (5.10) was related to the radial coordinate ρ by $x = k\rho$, where k was one of the separation constants in (5.6) and (5.7). In the same way as we saw previously when solving boundary value problems in Cartesian or spherical polar coordinates, here we shall also need to determine the coefficients in a sum over all the factorised solutions by establishing results for the appropriate generalised Fourier expansions. In particular, as we shall see in examples later, we shall need to consider a generalised Fourier expansion of a function $f(\rho)$ that is defined in the interval $0 \le \rho \le a$ in terms of a Fourier-Bessel series of the form

$$f(\rho) = \sum_{n\ge 1} a_n\, J_\nu(\frac{\alpha_{\nu n}}{a}\,\rho)\,, \tag{5.38}$$

where it will be recalled that $\alpha_{\nu n}$ is the n'th positive zero of the J_ν Bessel function. Note that this series expansion, based on the Bessel function $J_\nu(x)$, is analogous to a standard Fourier expansion, based on the use of the sine function $\sin x$, of a function $h(x)$ defined in the interval $0 \le x \le a$ in the form

$$h(x) = \sum_{n\ge 1} b_n\, \sin(\frac{n\pi}{a}\,x)\,. \tag{5.39}$$

To determine the coefficients a_n in (5.38), we need to establish appropriate orthogonality and normalisation relations for the Bessel functions. First, we show that the functions $\rho^{1/2}\, J_\nu(\alpha_{\nu n}\rho/a)$, for a fixed $\nu \ge 0$ and with n taking all the positive integer values, form an orthogonal set in the interval $0 \le \rho \le a$. It follows from Bessel's equation (5.10) that $J_\nu(\alpha_{\nu m}\rho/a)$ satisfies

$$\frac{d}{d\rho}\left(\rho\,\frac{dJ_\nu(\alpha_{\nu m}\,\rho/a)}{d\rho}\right) + \Big(\frac{\alpha_{\nu m}^2\,\rho}{a^2} - \frac{\nu^2}{\rho}\Big) J_\nu(\alpha_{\nu m}\,\rho/a) = 0\,. \tag{5.40}$$

We prove orthogonality by multiplying by $J_\nu(\alpha_{\nu n}\rho/a)$, and then subtracting off the equation where the rôles of m and n are exchanged, to give

$$\begin{aligned} &J_\nu(\alpha_{\nu n}\,\rho/a)\,\frac{d}{d\rho}\Big(\rho\,\frac{d}{d\rho}J_\nu(\alpha_{\nu m}\,\rho/a)\Big) - J_\nu(\alpha_{\nu m}\,\rho/a)\,\frac{d}{d\rho}\Big(\rho\,\frac{d}{d\rho}J_\nu(\alpha_{\nu n}\,\rho/a)\Big) \\ &\qquad = \frac{\alpha_{\nu n}^2 - \alpha_{\nu m}^2}{a^2}\,\rho\,J_\nu(\alpha_{\nu m}\,\rho/a)\,J_\nu(\alpha_{\nu n}\,\rho/a)\,. \end{aligned} \tag{5.41}$$

Next, we integrate this from $\rho = 0$ to $\rho = a$. On the left-hand side we integrate by parts, finding that there is now a cancellation of the resulting two integrands, leaving only the "boundary terms." Thus we have

$$\begin{aligned} &\Big[\rho\,J_\nu(\alpha_{\nu n}\,\rho/a)\,\frac{d}{d\rho}J_\nu(\alpha_{\nu m}\,\rho/a)\Big]_0^a - \Big[\rho\,J_\nu(\alpha_{\nu m}\,\rho/a)\,\frac{d}{d\rho}J_\nu(\alpha_{\nu n}\,\rho/a)\Big]_0^a \\ &\qquad = \frac{\alpha_{\nu n}^2 - \alpha_{\nu m}^2}{a^2}\,\int_0^a J_\nu(\alpha_{\nu m}\,\rho/a)\,J_\nu(\alpha_{\nu n}\,\rho/a)\,\rho\,d\rho\,. \end{aligned} \tag{5.42}$$

Recalling that near $\rho = 0$, $J_\nu(\alpha_{\nu n}\,\rho/a)$ is proportional to ρ^ν, we see that with our assumption that $\nu \geq 0$ the lower limits on the left-hand side of (5.42) will give zero. Furthermore, the upper limits will also give zero, since by construction $J_\nu(\alpha_{\nu m}) = 0$. Thus we arrive at the conclusion that for $m \neq n$ (which implies $\alpha_{\nu m} \neq \alpha_{\nu n}$), we shall have

$$\int_0^a J_\nu(\alpha_{\nu m}\,\rho/a)\,J_\nu(\alpha_{\nu n}\,\rho/a)\,\rho\,d\rho = 0\,. \tag{5.43}$$

Having established orthogonality when $m \neq n$, it remains to determine the normalisation of the integral that we get when instead we take $m = n$. To do this, let $x = \alpha_{\nu m}\,\rho/a$, so that

$$\int_0^a [J_\nu(\alpha_{\nu n}\,\rho/a)]^2\,\rho\,d\rho = \frac{a^2}{\alpha_{\nu n}^2}\,\int_0^{\alpha_{\nu n}} [J_\nu(x)]^2\,x\,dx\,. \tag{5.44}$$

To evaluate the integral on the right-hand side, we integrate by parts, by writing $[J_\nu(x)]^2\,x = \frac{1}{2}d/dx(x^2\,[J_\nu(x)]^2) - \frac{1}{2}x^2\,d/dx([J_\nu(x)]^2)$, so that

$$\begin{aligned} \int_0^{\alpha_{\nu n}} [J_\nu(x)]^2\,x\,dx &= \Big[\tfrac{1}{2}x^2\,J_\nu^2\Big]_0^{\alpha_{\nu n}} - \int_0^{\alpha_{\nu n}} x^2\,J_\nu\,J_\nu'\,dx\,, \\ &= -\int_0^{\alpha_{\nu n}} x^2\,J_\nu\,J_\nu'\,dx\,. \end{aligned} \tag{5.45}$$

Note that the boundary term vanishes at both endpoints. Now use the Bessel equation (5.10) to write $x^2\,J_\nu$ as $\nu^2\,J_\nu - x\,J_\nu' - x^2\,J_\nu''$, so that we get

$$\begin{aligned} \int_0^{\alpha_{\nu n}} [J_\nu(x)]^2\,x\,dx &= -\int_0^{\alpha_{\nu n}} \Big(\nu^2\,J_\nu\,J_\nu' - x\,{J_\nu'}^2 - x^2\,J_\nu'\,J_\nu''\Big)\,dx\,, \\ &= -\int_0^{\alpha_{\nu n}} \Big(\tfrac{1}{2}\nu^2\,(J_\nu^2)' - \tfrac{1}{2}(x^2\,{J_\nu'}^2)'\Big)\,dx \\ &= \tfrac{1}{2}\Big[-\nu^2\,J_\nu^2 + x^2\,{J_\nu'}^2\Big]_0^{\alpha_{\nu n}}\,. \end{aligned} \tag{5.46}$$

The first term in the final line vanishes at both our endpoints (recall that $\alpha_{\nu n}$ are precisely the values of argument for which $J_\nu(\alpha_{\nu n}) = 0$). For the second term, we subtract (5.35) from (5.36) to give

$$J'_\nu(x) = \frac{\nu}{x}\, J_\nu(x) - J_{\nu+1}(x)\,. \tag{5.47}$$

Thus, with our assumption that $\nu \geq 0$ we see that $x^2\, {J'_\nu}^2$ will vanish at $x = 0$. Also, from (5.47) we see that $J'_\nu(\alpha_{\nu n}) = -J_{\nu+1}(\alpha_{\nu n})$, and so

$$\int_0^{\alpha_{\nu n}} [J_\nu(x)]^2\, x\, dx = \tfrac{1}{2}\alpha_{\nu n}^2\, [J_{\nu+1}(\alpha_{\nu n})]^2\,, \tag{5.48}$$

implying finally that

$$\int_0^a J_\nu(\alpha_{\nu m}\, \rho/a)\, J_\nu(\alpha_{\nu n}\, \rho/a)\, \rho\, d\rho = \tfrac{1}{2}a^2\, [J_{\nu+1}(\alpha_{\nu n})]^2\, \delta_{mn}\,. \tag{5.49}$$

Armed with these results, we now return to the expansion (5.38) of the function $f(\rho)$. Multiplying by $\rho J_\nu(\alpha_{\nu m}\rho/a)$, integrating over ρ, and using the orthogonality relation (5.49), we can solve for the coefficients a_n, finding

$$\boxed{a_n = \frac{2}{a^2[J_{\nu+1}(\alpha_{\nu n})]^2} \int_0^a d\rho\, \rho f(\rho) J_\nu\Big(\alpha_{\nu n}\frac{\rho}{a}\Big)\,.} \tag{5.50}$$

Going back to our separation of variables in cylindrical polar coordinates, where ϕ was written in the factorised form $\phi(\rho, \varphi, z) = R(\rho)\Phi(\varphi)Z(z)$, we see that the general solution of Laplace's equation can be written as a sum over all factorised solutions, each with its own constant coefficient, in the form

$$\phi(\rho, \varphi, z) = \sum_m \int dk \Big(A_m(k) J_m(k\rho) + B_m(k) Y_m(k\rho)\Big)\, e^{\mathrm{i}m\varphi}\, e^{kz}\,, \tag{5.51}$$

where $A_m(k)$ and $B_m(k)$ are arbitrary coefficients. (We have assumed here that $\phi(\rho, \varphi, z)$ is periodic in φ, but that boundary conditions that would restrict k to taking a discrete set of values have not yet been imposed.) In a situation where, for example, the potential was zero on a cylinder at $\rho = a$ (and regular at $\rho = 0$), the continuous integral over k would be replaced by a discrete sum of the form

$$\phi(\rho, \varphi, z) = \sum_{m,n} J_m(\alpha_{mn}\frac{\rho}{a})\, e^{\mathrm{i}m\varphi}\, (A_{mn}\, e^{\alpha_{mn}z/a} + \widetilde{A}_{mn} e^{-\alpha_{mn}z/a})\,. \tag{5.52}$$

5.3 A boundary-value problem in cylindrical polar coordinates

We can now apply some of the technology of Bessel functions to the solution of electrostatics problems in cylindrical polar coordinates. Consider the following example. A hollow

conducting cylinder of height h and radius a is held at zero potential. A flat conductor closes off the cylinder at $z = 0$, and is also at zero potential. The top face, at $z = h$, is held at some specified potential

$$\phi(\rho, \varphi, h) = V(\rho, \varphi)\,. \tag{5.53}$$

The problem is to determine the potential everywhere inside the cavity. We can exclude the Y_ν Bessel functions in this problem, since there should be no logarithmic singularities on the axis of the cylinder. Thus only the J_ν Bessel functions can arise.

From (5.6) we see that the z dependence and φ dependence of the separation functions $Z(z)$ and $\Phi(\varphi)$ will be

$$\begin{aligned} Z(z) &\sim A \sinh kz + B \cosh kz\,, \\ \Phi(\varphi) &\sim C \cos\nu\varphi + D \sin\nu\varphi\,. \end{aligned} \tag{5.54}$$

The vanishing of the potential on the plate at $z = 0$ means that for $Z(z)$, we shall have only the $\sinh kz$ solution. The periodicity in φ means that ν must be an integer.

The general solution of Laplace's equation for this problem will be

$$\phi(\rho, \varphi, z) = \sum_{m=0}^{\infty}\sum_{n=1}^{\infty} J_m(\alpha_{mn}\,\rho/a)\,(a_{mn}\,\sin m\varphi + b_{mn}\,\cos m\varphi)\,\sinh(\alpha_{mn}\,z/a)\,. \tag{5.55}$$

The expansion coefficients a_{mn} and b_{mn} are determined by matching this solution to the specified boundary condition (5.53) at $z = h$. Thus we have

$$V(\rho, \varphi) = \sum_{m=0}^{\infty}\sum_{n=1}^{\infty} J_m(\alpha_{mn}\,\rho/a)\,(a_{mn}\,\sin m\varphi + b_{mn}\,\cos m\varphi)\,\sinh(\alpha_{mn}\,h/a)\,. \tag{5.56}$$

The orthogonality relation (5.49) for the Bessel functions, together with the standard orthogonality for the trigonometric functions, means that all we need to do is to multiply (5.56) by $\rho J_p(\alpha_{pq}\,\rho/a)\,\sin p\varphi$ or $\rho J_p(\alpha_{pq}\,\rho/a)\,\cos p\varphi$ and integrate over ρ and φ in order to read off the integrals that determine the individual coefficients a_{pq} and b_{pq}. It is easy to see that the result is

$$a_{pq} = \frac{2}{\pi\,a^2\,\sinh(\alpha_{pq}h/a)\,J_{p+1}(\alpha_{pq})^2}\int_0^{2\pi} d\varphi \int_0^a \rho\,d\rho\,V(\rho,\varphi)\,J_p(\alpha_{pq}\,\rho/a)\,\sin p\varphi\,, \tag{5.57}$$

$$b_{pq} = \frac{2}{\pi\,a^2\,\sinh(\alpha_{pq}h/a)\,J_{p+1}(\alpha_{pq})^2}\int_0^{2\pi} d\phi \int_0^a \rho\,d\rho V(\rho,\varphi)\,J_p(\alpha_{pq}\,\rho/a)\,\cos p\varphi\,, \quad p > 0$$

$$b_{0q} = \frac{1}{\pi\,a^2\,\sinh(\alpha_{0q}h/a)\,J_1(\alpha_{0q})^2}\int_0^{2\pi} d\varphi \int_0^a \rho\,d\rho V(\rho,\varphi)\,J_0(\alpha_{0q}\,\rho/a)\,. \tag{5.58}$$

(The reason for the change of the numerator in the prefactor in the expression for b_{0q} is because of the nature of the orthogonality relations for the cosine functions. Namely, that $\int_0^{2\pi}\cos m\varphi\cos p\varphi d\varphi = \pi\delta_{pm}$ for $p \neq 0$, but it equals 2π if $p = m = 0$.)

5.4 Modified Bessel functions

In our separation of variables in cylindrical polar coordinates, there was a separation constant k^2 that arose in the equations (5.6) and (5.7). If this constant had been chosen to be of the opposite sign, then the separated equations would instead have been

$$\frac{d^2Z}{dz^2} + k^2\, Z = 0\,, \qquad \frac{d^2\Phi}{d\varphi^2} + \nu^2\, \Phi = 0\,, \tag{5.59}$$

$$\frac{d^2R}{d\rho^2} + \frac{1}{\rho}\,\frac{dR}{d\rho} - \Big(k^2 + \frac{\nu^2}{\rho^2}\Big)\, R = 0\,, \tag{5.60}$$

The solutions for Z would then be of the form

$$Z \sim e^{\pm \mathrm{i} k z}\,, \tag{5.61}$$

Writing $x = k\rho$, $y = R$ as before we now obtain the *Modified Bessel Equation*

$$x^2\, y''(x) + x\, y'(x) - (x^2 + \nu^2)\, y = 0\,, \tag{5.62}$$

It can seen that (5.62) can be obtained from the original Bessel equation (5.10) by sending $x \longrightarrow \mathrm{i}\, x$. Thus if $w(x)$ is a solution of the original Bessel equation then $w(\mathrm{i}x)$ will be a solution of the modified Bessel equation. Since in general we are considering solutions where ν is not necessarily an integer, we should be a little careful about how to handle the i, writing it as $e^{\mathrm{i}\pi/2}$ since it will be raised to a fractional power. The two linearly-independent solutions of the modified Bessel equation can be taken to be $I_\nu(x)$ and $K_\nu(x)$, defined by

$$I_\nu(x) \equiv e^{-\frac{1}{2}\pi\nu\mathrm{i}}\, J_\nu(x e^{\frac{1}{2}\pi\mathrm{i}})\,, \tag{5.63}$$

$$K_\nu(x) = \tfrac{1}{2}\pi e^{\frac{1}{2}(\nu+1)\pi\mathrm{i}}\, H^{(1)}_\nu(x e^{\frac{1}{2}\pi\mathrm{i}})\,. \tag{5.64}$$

Note that from these definitions we have that when ν is an integer, $\nu = n$,

$$I_{-n}(x) = I_n(x)\,, \qquad K_{-n}(x) = K_n(x)\,. \tag{5.65}$$

The function $I_\nu(x)$ has a simple power-series expansion

$$I_\nu(x) = \sum_{r\geq 0} \frac{1}{r!\,\Gamma(\nu + r + 1)}\, \Big(\frac{x}{2}\Big)^{\nu+2r}\,. \tag{5.66}$$

This is convergent for any finite x. Note that the coefficients in the series are all positive, and so $I_\nu(x) > 0$ for all positive x. At small x, we have

$$I_\nu(x) = \frac{1}{\Gamma(\nu+1)}\, \Big(\frac{x}{2}\Big)^{\nu} + \mathcal{O}(x^{\nu+2})\,. \tag{5.67}$$

At large x, it can be shown that $I_\nu(x)$ has the asymptotic form

$$I_\nu(x) \sim \frac{1}{\sqrt{2\pi x}}\, e^x \Big(1 + \mathcal{O}(x^{-1}\Big)\,. \tag{5.68}$$

The $K_\nu(x)$ Bessel functions for the first couple of integer values for ν have the small-x expansions

$$\begin{aligned} K_0(x) &= -(\log \tfrac{1}{2}x + \gamma)(1 + \tfrac{1}{4}x^2 + \tfrac{1}{64}x^4 + \cdots) + \tfrac{1}{4}x^2 + \tfrac{3}{128}x^4 + \cdots\,, \\ K_1(x) &= \tfrac{1}{2}(\log \tfrac{1}{2}x + \gamma)(x + \tfrac{1}{8}x^3 + \tfrac{1}{192}x^5 + \cdots) + \frac{1}{x} - \tfrac{1}{4}x - \tfrac{5}{64}x^3 + \cdots\,. \end{aligned} \tag{5.69}$$

In general, when $\nu > 0$ is not an integer, the leading-order behaviour is

$$K_\nu(x) \sim \tfrac{1}{2}\Gamma(\nu)\, \Big(\frac{x}{2}\Big)^{-\nu}\,. \tag{5.70}$$

At large x, $K_\nu(x)$ has the asymptotic form

$$K_\nu(x) \sim \sqrt{\frac{\pi}{2x}}\, e^{-x} \Big(1 + \mathcal{O}(x^{-1}\Big)\,, \tag{5.71}$$

so it falls off as $x \to \infty$.

From the asymptotic forms of the I_n and K_n modified Bessel functions, we see that $I_n(x)$ is well-behaved at small x but divergent at large x, while $K_n(x)$ is well-behaved at large x but singular at small x. Thus, when we are using these functions in constructing solutions of Laplace's equation, we shall need only the I_n functions in a small-ρ solution if there is non singularity on z axis. Likewise, in a large-ρ solution we shall need only the K_n functions, provided that the potential should be falling off at infinity.

5.5 Green function in cylindrical polar coordinates

Recall that in section 4.10 we obtained an expansion for $|\vec{r} - \vec{r}\,'|^{-1}$ (i.e. the Green function in free space) in spherical polar coordinates, in terms of the spherical harmonics. From this, we were then able to obtain an expansion for the Green function for the Dirichlet boundary-value problem on a sphere.

The essential idea involved in obtaining the expansion for $|\vec{r} - \vec{r}\,'|^{-1}$ in terms of spherical harmonics was to consider the most general solution of Laplace's equation in each of the regions $r < r'$ and $r > r'$, with the radial functions $r^{-1}R(r)$ in each of the two regions being taken to be the ones appropriate to the regularity requirements in that region. Thus, with the radial functions being r^ℓ and $r^{-(\ell+1)}$, we took the former choice for the region $r < r'$, and the latter for the region $r > r'$. The discontinuity between the two solutions of Laplace's equation as the radial coordinate made the transition from $r < r'$ to $r > r'$

was responsible for producing the delta function factor $\delta(r-r')$ when the Laplacian of the expressions was carefully evaluated. The $\delta(\theta-\theta')$ and $\delta(\varphi-\varphi')$ factors, on the other hand, arose through the completeness relation (4.159).

The same idea can be applied to obtain an expansion for $|\vec{r}-\vec{r}\,'|^{-1}$ in cylindrical polar coordinates. This time, the radial functions $R(\rho)$ are either $J_\nu(k\rho)$ and $Y_\nu(k\rho)$ or else $I_\nu(k\rho)$ and $K_\nu(k\rho)$, depending upon whether the separation constant k^2 enters the equations as in (5.6) and (5.7) or else in (5.59) and (5.60). We shall make the (5.59) and (5.60) sign choice here, and so the radial functions are $I_\nu(k\rho)$ and $K_\nu(k\rho)$, and the functions $Z(z)$ and $\Phi(\varphi)$ are of the forms

$$\begin{aligned} Z(z) &\sim \cos kz \quad \text{or} \quad \sin kz\,, \\ \Phi(\varphi) &\sim \cos\nu\varphi \quad \text{or} \quad \sin\nu\varphi\,. \end{aligned} \tag{5.72}$$

Single-valuedness of $\Phi(\varphi)$ under $\varphi \to \varphi + 2\pi$ implies that we must take $\nu = m$ =integer.

The relevant completeness relations that will generate the delta function factors $\delta(z-z')$ and $\delta(\varphi-\varphi')$ are the standard ones from Fourier analysis:

$$\begin{aligned} \delta(z-z') &= \frac{1}{2\pi}\int_{-\infty}^{\infty} dk e^{\mathrm{i}k(z-z')} = \frac{1}{\pi}\int_0^\infty dk\cos k(z-z')\,, \\ \delta(\varphi-\varphi') &= \frac{1}{2\pi}\sum_{m=-\infty}^{\infty} e^{\mathrm{i}m(\varphi-\varphi')}\,. \end{aligned} \tag{5.73}$$

The Green function can now be expanded in the form

$$G(\vec{r},\vec{r}\,') = \frac{2}{\pi}\sum_{m=-\infty}^{\infty}\int_0^\infty dk\, R_{(k,m)}(\rho,\rho')\, e^{\mathrm{i}m(\varphi-\varphi')}\,\cos k(z-z')\,, \tag{5.74}$$

where $R_{(k,m)}(\rho,\rho')$ is an appropriate radial function of ρ, which in general must be some linear combination of $I_m(k\rho)$ and $K_m(k\rho)$. It will take two different forms in the two regions $\rho<\rho'$ and $\rho>\rho'$. From our construction of the general solution of Laplace's equation in cylindrical polar coordinates, it is manifest that $G(\vec{r},\vec{r}\,')$ in (5.74) satisfies Laplace's equation everywhere except at the transition where $\rho=\rho'$.

From the asymptotic forms of $I_\nu(x)$ and $K_\nu(x)$ described in section 5.2, it is evident that the solution well behaved at small ρ must be proportional to I_ν, whilst the solution well behaved at large ρ must be proportional to K_ν. Because of the symmetry of the Green function under exchanging ρ and ρ', we can expect that $R_{(k,m)}(\rho,\rho')$ will be proportional to $I_m(k\rho)K_m(k\rho')$ when $\rho<\rho'$, and proportional to $K_m(k\rho)I_m(k\rho')$ when $\rho>\rho'$. (We saw precisely the analogous feature previously, in the expansion in spherical polar coordinates.)

In fact, as we shall now show, the required function $R_{(k,m)}(\rho,\rho')$ is

$$R_{(k,m)}(\rho,\rho') = I_m(k\rho)K_m(k\rho')\vartheta(\rho'-\rho) + K_m(k\rho)I_m(k\rho')\vartheta(\rho-\rho')\,. \tag{5.75}$$

Differentiating $R_{(k,m)}(\rho,\rho')$, and making use of the standard properties of the derivative of the Heaviside ϑ function (i.e. $d\vartheta(\rho-\rho')/d\rho = \delta(\rho-\rho')$, etc., and the fact that $\delta(\rho-\rho')f(\rho') = \delta(\rho-\rho')f(\rho)$, we see that

$$\begin{aligned}
\frac{dR_{(k,m)}(\rho,\rho')}{d\rho} &= \alpha\Big[K_m(k\rho')\frac{dI_m(k\rho)}{d\rho}\vartheta(\rho'-\rho) + I_m(k\rho')\frac{dK_m(k\rho)}{d\rho}\vartheta(\rho-\rho')\Big]\,,\\
\frac{d^2R_{(k,m)}(\rho,\rho')}{d\rho^2} &= \alpha\Big[K_m(k\rho')\frac{d^2I_m(k\rho)}{d\rho^2}\vartheta(\rho'-\rho) + I_m(k\rho')\frac{d^2K_m(k\rho)}{d\rho^2}\vartheta(\rho-\rho')\Big]\\
&\quad +\Big[I_m(k\rho)\frac{dK_m(k\rho)}{d\rho} - K_m(k\rho)\frac{dI_m(k\rho)}{d\rho}\Big]\,\delta(\rho-\rho')\,.
\end{aligned} \tag{5.76}$$

When we apply the Laplace operator to the expression (5.74), all the terms in (5.76) where the ϑ functions remain undifferentiated will combine with the z and φ derivative terms to give the "bulk" result that $G(\vec{r},\vec{r}\,')$ satisfies Laplace's equation, in the way we discussed above. The only terms left over will be those in (5.76) involving the delta function. For these, we need to derive a simple result about properties of the solutions of the modified Bessel equation (5.62). Suppose $y_1(x)$ and $y_2(x)$ are two such solutions. Plugging y_1 into (5.62) and multiplying by y_2, and then subtracting the expression where the roles of y_1 and y_2 are exchanged, we obtain

$$xy_2(xy_1')' - xy_1(xy_2')' = 0\,. \tag{5.77}$$

(Note that $x^2y'' + xy'$ can be written as $x(xy')'$.) Dividing out by x, we can then rewrite this as

$$(xy_2y_1')' - xy_2'y_1' - (xy_1y_2')' + xy_1'y_2' = 0\,, \tag{5.78}$$

and hence $[x(y_1'y_2 - y_2'y_1)]' = 0$. From this, we deduce that for any two solutions of the modified Bessel equation, their Wronskian $W(y_1,y_2) \equiv y_1y_2' - y_2y_1'$ satisfies

$$W(y_1,y_2) = y_1y_2' - y_2y_1' = \frac{c}{x}\,, \tag{5.79}$$

where c is some constant that depends on which particular solutions y_1 and y_2 are chosen.

We can calculate the Wronskian for $I_\nu(x)$ and $K_\nu(x)$ easily, by using the small-x expansions (5.67) and (5.70). Since we have only to determine the value of the constant c in (5.79) for this particular pair of solutions, it suffices to use their small-x expansions (5.67) and (5.70), and calculate the leading-order term when substituting them into $I_m(x)K_m'(x) - K_m(x)I_m'(x)$. This gives

$$I_\nu(x)K_\nu'(x) - K_\nu(x)I_\nu'(x) = -\frac{1}{x}\,, \tag{5.80}$$

and hence we read off that $c = -1$.[38] Thus the prefactor of the $\delta(\rho - \rho')$ delta function on the bottom line of eqn (5.76) gives

$$I_m(k\rho)\frac{dK_m(k\rho)}{d\rho} - K_m(k\rho)\frac{dI_m(k\rho)}{d\rho} = -\frac{1}{\rho}\,. \tag{5.81}$$

This then implies that $G(\vec{r}, \vec{r}\,')$ in (5.74) satisfies

$$\nabla^2 G(\vec{r}, \vec{r}\,') = -4\pi\delta^3(\vec{r} - \vec{r}\,') = -\frac{4\pi}{\rho}\,\delta(\rho - \rho')\delta(\varphi - \varphi')\delta(z - z')\,. \tag{5.82}$$

To summarise, we have shown that in cylindrical polar coordinates we have the expansion

$$\boxed{\frac{1}{|\vec{r} - \vec{r}\,'|} = \frac{2}{\pi}\sum_{m=-\infty}^{\infty}\int_0^\infty dk\, I_m(k\rho_<)K_m(k\rho_>)e^{\mathrm{i}m(\varphi-\varphi')}\,\cos k(z - z')\,,} \tag{5.83}$$

where the notation here is that $\rho_<$ means whichever of ρ and ρ' is the smaller, and $\rho_>$ means whichever of them is the larger.

Recall that for the integer degrees, as we have here, I_n and K_n are each the same for positive and negative n (see eqn (5.65)). Thus we may also write (5.83) as

$$\boxed{\begin{aligned}\frac{1}{|\vec{r} - \vec{r}\,'|} \;&=\; \frac{4}{\pi}\int_0^\infty dk\,\left(\tfrac{1}{2}I_0(k\rho_<)K_0(k\rho_>) + \sum_{m\ge 1} I_m(k\rho_<)K_m(k\rho_>)\cos m(\varphi - \varphi')\right)\\ &\qquad\qquad \times \cos k(z - z')\,,\end{aligned}} \tag{5.84}$$

(5.84)

6 Multipole Expansion

The multipole expansion provides a way of organising the expression for the electrostatic potential due to a localised distribution of charges, as a sum over terms proportional to the total charge, the dipole moment, the quadrupole moment, and so on. In order to discuss this, it will be convenient first to introduce an index notation for vectors and tensors in three-dimensional Cartesian space.

[38] Note that by using the small-x expansion (5.70) for $K_\nu(x)$ we are assuming that ν is some generic non-integer value. Having thus obtained the result $c = -1$ for all such non-integer ν, we can trivially, by continuity, deduce that it continues to hold when ν *is* an integer, since there exist a continuum of non-integer ν values that are arbitrarily close to the integer.

6.1 Index notation for Cartesian vectors and tensors

Until now, we have typically been writing a vector V as an ordered triplet of components, in the form

$$\vec{V} = (V_x, V_y, V_z)\,. \tag{6.1}$$

In the index notation, we instead label the three components by the numbers 1, 2 and 3, rather than by x, y and z. Thus we write

$$\vec{V} = (V_1, V_2, V_3)\,. \tag{6.2}$$

In the same way, instead of writing $\vec{r} = (x, y, z)$, we can write it as

$$\vec{r} = (x_1, x_2, x_3)\,. \tag{6.3}$$

In other words, we call the three Cartesian coordinates x_1, x_2 and x_3, rather than x, y and z.

The scalar product $\vec{A} \cdot \vec{B}$ between any two vectors $\vec{A}$ and $\vec{B}$ can now be written as

$$\vec{A} \cdot \vec{B} = \sum_{i=1}^{3} A_i B_i\,. \tag{6.4}$$

At this point, the index notation is looking somewhat clumsy and long-winded. It becomes much simpler to write if we now adopt the *Einstein summation convention*. The idea here is that in any valid vector or tensor expression written using index notation, if a particular index occurs *exactly twice* in a term, then it will *always* be summed over, as in the right-hand side in (6.4). It is therefore redundant to write the summation explicitly, and so we can simply rewrite (6.4) as

$$\vec{A} \cdot \vec{B} = A_i B_i\,. \tag{6.5}$$

By the Einstein summation convention, it is understood that the index i is to be summed over.

The i index in (6.5) is called a *dummy index*, meaning that it is just an index "internal" to the term $A_i B_i$ that is summed over. The term could just as well be written as $A_j B_j$ or $A_k B_k$, etc. The dummy index is like a summation index in a computer programme; for example, in *Mathematica* one would write $A_i B_i$ as

$$\mathrm{Sum}[A[i]B[i], \{i, 1, 3\}]\,, \tag{6.6}$$

and obviously any other name could equally well have been chosen for the summation variable.

The index notation is extremely useful, but there are pitfalls for the unwary. Suppose, for example, we wish to write $(\vec{A}\cdot\vec{B})(\vec{C}\cdot\vec{D})$ in index notation. It will be

$$(\vec{A}\cdot\vec{B})(\vec{C}\cdot\vec{D}) = A_i B_i C_j D_j \,. \tag{6.7}$$

It is *absolutely essential* to make sure that a different index is chosen for the second factor, $\vec{C}\cdot\vec{D}$, that has not been used in writing the first factor, $\vec{A}\cdot\vec{B}$. Thus, for example,

$$A_i B_i C_i D_i \tag{6.8}$$

with i summed over 1, 2 and 3, would be very different from $(\vec{A}\cdot\vec{B})(\vec{C}\cdot\vec{D})$, and there will *never* be any occasion, when manipulating vector or tensor expressions, when such a thing would be needed.[39] More generally, there will *never* be any occasion when a valid vector or tensor expression has a term where the same index occurs more than twice.

In fact in a valid vector or tensor expression, a particular index can appear either once, or twice (or not at all) in each term. For example, suppose we want to write $\vec{V} = (\vec{A}\cdot\vec{B})\vec{C} + (\vec{E}\cdot\vec{F})\vec{G}$ in index notation. This is a vector-valued expression, and so we need a *free index* that labels the three components. We can write it as

$$V_i = A_j B_j C_i + E_j F_j G_i \,. \tag{6.9}$$

This could equally well be written, for example, as

$$V_i = A_j B_j C_i + E_k F_k G_i \,, \tag{6.10}$$

since the summation over the dummy index j in the first term on the right-hand side of (6.9) is completely independent of the summation over the dummy index j in the second term. But we must *not*, under any circumstances, choose the label i for either of the dummy indices in (6.10), since i is already being used as the free index.

The reason for emphasising so strongly the need to be careful about not violating the rules when using indices is that it is by far the most common mistake that people make when they first meet the index notation. If you find you have written a term such as $A_i B_i C_i$, or $A_i B_i C_i D_i$, it is simply wrong; there is no point in continuing the calculation until the source of the trouble has been located and corrected.

[39] If you think again in terms of writing a computer program to calculate $(\vec{A}\cdot\vec{B})(\vec{C}\cdot\vec{D})$, it should be clear that what is needed is

$$\mathrm{Sum}[A[i]B[i]C[j]D[j],\{i,1,3\},\{j,1,3\}]\,, \quad \textbf{not} \quad \mathrm{Sum}[A[i]B[i]C[i]D[i],\{i,1,3\}]\,.$$

One final piece of notation before we move on concerns the gradient operator $\vec{\nabla} = (\nabla_1, \nabla_2, \nabla_3) = (\partial/\partial x, \partial/\partial y, \partial/\partial z)$. In index notation we therefore have

$$\nabla_i = \frac{\partial}{\partial x_i} \,. \tag{6.11}$$

It will be convenient to abbreviate the writing of $\partial/\partial x_i$, by defining

$$\partial_i = \frac{\partial}{\partial x_i} \,. \tag{6.12}$$

Note that by the rules of partial differentiation, we have

$$\partial_i x_j = \delta_{ij} \,. \tag{6.13}$$

Note also, since it is a point that is often overlooked by those meeting the index notation for the first time, that

$$\delta_{ii} = 3 \,, \tag{6.14}$$

(since a summation over $i = 1, 2, 3$ is understood), and so also we have $\partial_i \, x_i = 3$. Another important observation is that

$$\delta_{ij} A_j = A_i \tag{6.15}$$

for any vector $\vec{A}$.

6.2 Multipole expansion in Cartesian coordinates

Now back to the multipole expansion. Consider the electrostatic potential of N point charges q_a, located at fixed positions $\vec{r}_a$. It is given by

$$\phi(\vec{r}) = \sum_{a=1}^{N} \frac{q_a}{|\vec{r} - \vec{r}_a|} \,. \tag{6.16}$$

In the continuum limit, the potential due to a charge distribution characterised by the charge density $\rho(\vec{r})$ is given by

$$\phi(\vec{r}) = \int \frac{\rho(\vec{r}\,')d^3\vec{r}\,'}{|\vec{r} - \vec{r}\,'|} \,. \tag{6.17}$$

Since we shall assume that the charges are confined to a finite region, it is useful to perform a *multipole expansion* of the potential far from the region where the charges are located. For this discussion it is convenient to choose the origin of the Cartesian coordinate system so that the localised charge distribution is in the neighbourhood of the origin. The multipole expansion then amounts to an expansion in inverse powers of $r = |\vec{r}|$. This can be achieved by performing a Taylor expansion of $|\vec{r} - \vec{r}\,'|^{-1}$.

Recall that in one dimension, Taylor's theorem gives

$$f(x+a) = f(x) + af'(x) + \frac{a^2}{2!}f''(x) + \frac{a^3}{3!}f'''(x) + \cdots . \tag{6.18}$$

In three dimensions, the analogous expansion is

$$f(\vec{r}+\vec{a}\,) = f(\vec{r}) + a_i\partial_i f(\vec{r}) + \frac{1}{2!}a_i a_j \partial_i\partial_j f(\vec{r}) + \frac{1}{3!}a_i a_j a_k \partial_i\partial_j\partial_k f(\vec{r}) + \cdots . \tag{6.19}$$

(You can easily verify this explicitly, for the first few orders in the expansion, by making repeated use of the one-dimensional Taylor expansion (6.18) applied in the x, y and z directions.)

We now apply this 3-dimensional Taylor expansion to the function $f(\vec{r}) = 1/|\vec{r}| = 1/r$, taking $\vec{a} = -\vec{r}\,'$. This gives

$$\frac{1}{|\vec{r}-\vec{r}\,'|} = \frac{1}{r} - x'_i\partial_i\frac{1}{r} + \frac{1}{2!}x'_i x'_j\partial_i\partial_j\frac{1}{r} - \frac{1}{3!}x'_i x'_j x'_k\partial_i\partial_j\partial_k\frac{1}{r} + \cdots . \tag{6.20}$$

Now since $r^2 = x_j x_j$, it follows that $\partial_i r^2 = 2r\,\partial_i r = \partial_i(x_j x_j) = 2x_j\,\partial_i x_j = 2x_j\,\delta_{ij} = 2x_i$, and so

$$\partial_i r = \frac{x_i}{r}\,. \tag{6.21}$$

Note that we have (assuming $r > 0$) that

$$\partial_i\partial_i\frac{1}{r} = \partial_i\Big(-\frac{x_i}{r^3}\Big) = -\frac{3}{r^3} + \frac{3x_i}{r^4}\,\frac{x_i}{r} = 0\,, \tag{6.22}$$

or, in other words[40]

$$\nabla^2\frac{1}{r} = 0\,, \qquad \text{for } r > 0\,. \tag{6.23}$$

A consequence of this is that the multiple derivatives

$$\partial_i\partial_j\frac{1}{r}\,, \qquad \partial_i\partial_j\partial_k\frac{1}{r}\,, \qquad \partial_i\partial_j\partial_k\partial_\ell\frac{1}{r}\,, \qquad \cdots \tag{6.24}$$

are all *traceless* on any pair of indices, in the sense that

$$\delta_{ij}\partial_i\partial_j\frac{1}{r} = 0\,, \qquad \delta_{ij}\partial_i\partial_j\partial_k\frac{1}{r} = 0\,, \qquad \text{etc.} \tag{6.25}$$

Note also that the expressions $\partial_i\partial_j\frac{1}{r}$, $\partial_i\partial_j\partial_k\frac{1}{r}$, etc. are *totally symmetric* in their indices, so

$$\partial_i\partial_j\frac{1}{r} = \partial_j\partial_i\frac{1}{r}\,, \qquad \partial_i\partial_j\partial_k\frac{1}{r} = \partial_j\partial_i\partial_k\frac{1}{r} = \partial_i\partial_k\partial_j\frac{1}{r} = \cdots\,, \qquad \text{etc.}\,, \tag{6.26}$$

[40] We already know, from an earlier discussion of point charges, that if we include the point $r = 0$ we shall have $\nabla^2\,\frac{1}{r} = -4\pi\,\delta^3(\vec{r})$. In our present discussion, however, r will be large, and so the singular behaviour at $r = 0$ will not concern us.

as a consequence of the fact that partial derivatives commute with one another.

We can use the traceless property (6.25) in order to replace the quantities

$$x'_i x'_j\,, \qquad x'_i x'_j x'_k\,, \qquad \cdots \tag{6.27}$$

that multiply the derivative terms in (6.20) by the totally tracefree quantities

$$(x'_i x'_j - \tfrac{1}{3}\delta_{ij} r'^2)\,, \qquad (x'_i x'_j x'_k - \tfrac{1}{5}[x'_i\delta_{jk} + x'_j\delta_{ik} + x'_k\delta_{ij}]r'^2)\,, \qquad \cdots \tag{6.28}$$

where $r'^2 = x'_i x'_i$. (We can do this because the trace terms that we are subtracting out here give zero when they are contracted onto the multiple derivatives of $1/r$ in (6.20).) It therefore follows from (6.17) and (6.20) that we have

$$\begin{aligned}\phi(\vec{r}) \;&=\; \frac{1}{r}\int \rho(\vec{r}\,')d^3\vec{r}\,' - \Big(\partial_i \frac{1}{r}\Big)\int x'_i\rho(\vec{r}\,')d^3\vec{r}\,' + \tfrac{1}{2}\Big(\partial_i\partial_j\frac{1}{r}\Big)\int (x'_i x'_j - \tfrac{1}{3}\delta_{ij}r'^2)\rho(\vec{r}\,')d^3\vec{r}\,' \\ &\quad -\tfrac{1}{6}\Big(\partial_i\partial_j\partial_k\frac{1}{r}\Big)\int (x'_i x'_j x'_k - \tfrac{1}{5}[x'_i\delta_{jk} + x'_j\delta_{ik} + x'_k\delta_{ij}]r'^2)\rho(\vec{r}\,')d^3\vec{r}\,' + \cdots\,. \end{aligned} \tag{6.29}$$

The expansion here can be written as

$$\phi(\vec{r}) = \frac{Q}{r} - p_i\partial_i\frac{1}{r} + \frac{1}{3\times 2!}Q_{ij}\partial_i\partial_j\frac{1}{r} - \frac{1}{5\times 3!}Q_{ijk}\partial_i\partial_j\partial_k\frac{1}{r} + \cdots \tag{6.30}$$

where

$$\begin{aligned} Q \;&=\; \int \rho(\vec{r}\,')d^3\vec{r}\,'\,, \\ p_i \;&=\; \int x'_i\rho(\vec{r}\,')d^3\vec{r}\,'\,, \\ Q_{ij} \;&=\; \int (3x'_i x'_j - \delta_{ij}r'^2)\rho(\vec{r}\,')d^3\vec{r}\,'\,, \\ Q_{ijk} \;&=\; \int (5x'_i x'_j x'_k - [x'_i\delta_{jk} + x'_j\delta_{ik} + x'_k\delta_{ij}]r'^2)\rho(\vec{r}\,')d^3\vec{r}\,'\,, \end{aligned} \tag{6.31}$$

and so on. The quantity Q is the total charge of the system, p_i is the dipole moment, Q_{ij} is the quadrupole moment, and Q_{ijk}, $Q_{ijk\ell}$, etc., are the higher multipole moments. Note that by construction, all the multipole moments with two or more indices are symmetric and traceless on all indices.

The reason for making a issue about the trace terms in the above discussion is that we need to have a clear picture of how many independent components arise in each of the multipole moments. If we didn't subtract out the trace terms in the way described above, one would end up with the incorrect impression that there were more independent components in the higher multipole moments (quadrupole and above) than is actually the case. We shall discuss this in more detail below.

Note that the terms in the multipole expansion (6.30) do indeed fall off with increasing inverse powers of r. For example, the dipole term is given by

$$\phi_{\rm Dipole} = -p_i \partial_i \frac{1}{r} = \frac{p_i x_i}{r^3} = \frac{p_i n_i}{r^2}\,, \tag{6.32}$$

which falls off like $1/r^2$, since $n_i \equiv x_i/r$ is a unit-length vector. The quadrupole term is given by

$$\phi_{\rm Quadrupole} = \tfrac{1}{6} Q_{ij} \partial_i \partial_j \frac{1}{r} = \tfrac{1}{6} Q_{ij} \frac{(3x_i x_j - r^2 \delta_{ij})}{r^5} = \tfrac{1}{2} Q_{ij} \frac{x_i x_j}{r^5} = \tfrac{1}{2} Q_{ij} \frac{n_i n_j}{r^3}\,, \tag{6.33}$$

which falls off like $1/r^3$. (The penultimate equality above follows because Q_{ij} is traceless.)

In summary, we see that the multipole expansion of the potential due to a localised charge distribution takes the form

$$\boxed{\phi(\vec{r}) = \frac{Q}{r} + \frac{\vec{p}\cdot\vec{n}}{r^2} + \frac{1}{2!}\frac{Q_{ij} n_i n_j}{r^3} + \frac{1}{3!}\frac{Q_{ijk} n_i n_j n_k}{r^4} + \cdots\,.} \tag{6.34}$$

The electric field due to the monopole potential $\phi_{\rm Monopole} = Q/r$ is the familiar one

$$\vec{E}_{\rm Monopole} = \frac{Q\vec{r}}{r^3} = \frac{Q\vec{n}}{r^2}\,, \tag{6.35}$$

which falls off as the square of the distance. For the dipole potential (6.32), the electric field is easily calculated using index notation:

$$\partial_i \Big(\frac{p_j x_j}{r^3}\Big) = \frac{p_j \delta_{ij}}{r^3} - \frac{3 p_j x_j x_i}{r^5} = -\frac{3 n_i n_j p_j - p_i}{r^3}\,, \tag{6.36}$$

and hence

$$\boxed{\vec{E}_{\rm Dipole} = \frac{3\vec{n}\,(\vec{n}\cdot\vec{p}) - \vec{p}}{r^3}\,.} \tag{6.37}$$

This falls off as the cube of the distance. The electric fields for the higher multipole terms can be calculated in a similar way.

As mentioned already, it is important to have a clear understanding of how many *independent* components there are in each of the multipoles in the expansion. The total charge Q (the electric monopole moment) is of course a single quantity. The dipole moment p_i is a 3-vector, so it has three independent components in general.

The quadrupole moment Q_{ij} is a symmetric 2-index tensor in three dimensions, which would mean $3 \times 4/2 = 6$ independent components. But it is also traceless, $Q_{ii} = 0$, which is one condition. Thus there are $6 - 1 = 5$ independent components.

The octopole moment Q_{ijk} is a 3-index symmetric tensor, which would mean $3 \times 4 \times 5/3! = 10$ independent components. But it is also traceless, $Q_{iij} = 0$, which is 3 conditions.

Thus the octopole has in general $10 - 3 = 7$ independent components. It is straightforward to see in the same way that the 2^ℓ-pole moment

$$Q_{i_1 i_2 \cdots i_\ell} = (2\ell - 1) \int (x'_{i_1} x'_{i_2} \cdots x'_{i_\ell} - \text{traces}) \rho(\vec{r}\,') d^3\vec{r}\,' \,, \qquad \ell \geq 1 \,, \tag{6.38}$$

has $(2\ell + 1)$ independent components.

6.3 Multipole expansion using spherical harmonics

In fact, the multipole expansion (6.30) is equivalent to an expansion in spherical polar coordinates, using the spherical harmonics $Y_{\ell m}(\theta, \phi)$:

$$\phi(r, \theta, \phi) = \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} B_{\ell m}\, Y_{\ell m}(\theta, \phi)\, \frac{1}{r^{\ell+1}} \,. \tag{6.39}$$

At a given value of ℓ the terms fall off like $r^{-\ell-1}$, and there are $(2\ell + 1)$ of them, with coefficients $B_{\ell m}$, since m ranges over the integers $-\ell \leq m \leq \ell$. For each value of ℓ, there is a linear relationship between the $(2\ell + 1)$ components of $B_{\ell m}$ and the $(2\ell + 1)$ components of the multipole moments Q, p_i. Q_{ij}, Q_{ijk}, etc. Likewise, for each ℓ there is a linear relationship between $r^{-\ell-1}\, Y_{\ell m}(\theta, \varphi)$ and the set of functions $\partial_{i_1} \partial_{i_2} \cdots \partial_{i_\ell} r^{-1}$.

Consider, for example, $\ell = 1$. The three functions $Z_i \equiv \partial_i r^{-1} = -x_i / r^3$ that arise in the dipole term in the multipole expansion (6.30) are given by

$$Z_1 = -\frac{\sin\theta \cos\varphi}{r^2} \,, \qquad Z_2 = -\frac{\sin\theta \sin\varphi}{r^2} \,, \qquad Z_3 = -\frac{\cos\theta}{r^2} \,, \tag{6.40}$$

when expressed in terms of spherical polar coordinates, for which

$$x = r \sin\theta \cos\varphi \,, \qquad y = r \sin\theta \sin\varphi \,, \qquad z = r \cos\theta \,. \tag{6.41}$$

On the other hand, the $\ell = 1$ spherical harmonics are given by (see (4.122))

$$Y_{11} = -\sqrt{\frac{3}{8\pi}}\, \sin\theta\, e^{\mathrm{i}\varphi} \,, \qquad Y_{10} = \sqrt{\frac{3}{4\pi}}\, \cos\theta \,, \qquad Y_{1,-1} = \sqrt{\frac{3}{8\pi}}\, \sin\theta\, e^{-\mathrm{i}\varphi} \,. \tag{6.42}$$

Thus we see that

$$Z_1 = \sqrt{\frac{8\pi}{3}}\, \frac{(Y_{11} - Y_{1,-1})}{2r^2} \,, \qquad Z_2 = \sqrt{\frac{8\pi}{3}}\, \frac{(Y_{11} + Y_{1,-1})}{2\mathrm{i}\, r^2} \,, \qquad Z_3 = -\sqrt{\frac{4\pi}{3}}\, \frac{Y_{10}}{r^2} \,. \tag{6.43}$$

Analogous relations can be seen for all higher values of ℓ, corresponding to the higher terms in the multipole expansion (6.30).

Working directly with the spherical harmonics, we may substitute the expansion (4.152) for $|\vec{r} - \vec{r}\,'|^{-1}$, i.e.

$$\frac{1}{|\vec{r} - \vec{r}\,'|} = \sum_{\ell \geq 0} \sum_{m=-\ell}^{\ell} \frac{4\pi}{2\ell + 1}\, \frac{r'^{\ell}}{r^{\ell+1}}\, \overline{Y}_{\ell m}(\theta', \varphi')\, Y_{\ell m}(\theta, \varphi) \,, \qquad r > r' \,, \tag{6.44}$$

into

$$\phi(\vec{r}) = \int \frac{\rho(\vec{r}\,')}{|\vec{r} - \vec{r}\,'|}\,, \tag{6.45}$$

obtaining

$$\phi(\vec{r}) = 4\pi \sum_{\ell,m} \frac{q_{\ell m}}{(2\ell+1)}\, \frac{1}{r^{\ell+1}}\, Y_{\ell m}(\theta,\varphi)\,, \tag{6.46}$$

where the multipole moments $q_{\ell m}$ are given by

$$q_{\ell m} \equiv \int \rho(\vec{r})\, r^\ell\, \overline{Y}_{\ell m}(\theta,\varphi) d^3\vec{r}\,. \tag{6.47}$$

(We have dropped the prime labels on the integration variables here.) Note that because of (4.114), they satisfy

$$\bar{q}_{\ell m} = (-1)^m\, q_{\ell,-m}\,. \tag{6.48}$$

Clearly, in view of (6.48), the total number of real quantities encoded in $q_{\ell m}$ for a given value of ℓ is $2\ell+1$, which is exactly the same as the number of independent components $Q_{i_1\cdots i_\ell}$ in the ℓ'th multipole moment tensor. Using the expressions (4.125) for the first few spherical harmonics, we can see that the $q_{\ell m}$ for $\ell = 0$, 1 and 2 are related to Q, p_i and Q_{ij} by

$$\begin{aligned}
q_{00} &= \frac{1}{\sqrt{4\pi}}\, Q\,, \\
q_{11} &= -\sqrt{\frac{3}{8\pi}}\,(p_1 - \mathrm{i}p_2)\,, \qquad q_{10} = \sqrt{\frac{3}{4\pi}}\, p_3\,, \\
q_{22} &= \tfrac{1}{12}\sqrt{\frac{15}{2\pi}}\,(Q_{11} - 2\mathrm{i}Q_{12} - Q_{22})\,, \quad q_{21} = -\tfrac{1}{3}\sqrt{\frac{15}{8\pi}}\,(Q_{13} - \mathrm{i}Q_{23})\,, \quad q_{20} = \tfrac{1}{2}\sqrt{\frac{5}{4\pi}}\, Q_{33}\,.
\end{aligned} \tag{6.49}$$

(The expressions for $q_{\ell m}$ with negative m follow from (6.48).) Analogous relations hold for all the $q_{\ell m}$.

6.4 Another construction of the spherical harmonics

The multipole expansion using Cartesian coordinates, which we discussed in section 6.2, points the way to a very simple construction of the spherical harmonics that sidesteps the need for the elaborate investigation of the separation of variables and the properties of the associated Legendre functions that we examined in great depth previously. We saw in section 6.2 that the tensor

$$\partial_{i_1} \cdots \partial_{i_\ell} \frac{1}{r} \tag{6.50}$$

is symmetric under the exchange of any pair of indices, and it is traceless with respect to the contraction of any pair of indices. It also satisfies Laplace's equation (for $r > 0$), since

$$\nabla^2\Big(\partial_{i_1} \cdots \partial_{i_\ell} \frac{1}{r}\Big) = \partial_{i_1} \cdots \partial_{i_\ell}\Big(\nabla^2 \frac{1}{r}\Big) = 0\,. \tag{6.51}$$

(Remember $\nabla^2 = \partial_j \partial_j$, and partial derivatives commute.)

Let us define

$$\psi_\ell^{(C)} \equiv C_{i_1 \cdots i_\ell}\, \partial_{i_1} \cdots \partial_{i_\ell} \frac{1}{r}\,, \tag{6.52}$$

where $C_{i_1 \cdots i_\ell}$ is a constant, traceless, symmetric tensor. The subscript ℓ and superscript (C) on ψ indicate that the function $\psi_\ell^{(C)}$ is constructed by applying ℓ partial derivatives to $\frac{1}{r}$, and that it depends upon the choice of constant tensor $C_{i_1 \cdots i_\ell}$.

In view of the fact that $\partial_i r = x_i/r = n_i$, and if we write the Cartesian coordinates (x, y, z) in terms of spherical polar coordinates (r, θ, φ) we will have

$$n_1 = \sin\theta \cos\varphi\,, \qquad n_2 = \sin\theta \sin\varphi\,, \qquad n_3 = \cos\theta\,. \tag{6.53}$$

It is evident that when we calculate the result of acting with all the derivatives in (6.52), we shall find that $\psi_\ell^{(C)}$ is of the form

$$\psi_\ell^{(C)} = \frac{1}{r^{\ell+1}}\, Y_\ell^{(C)}(\theta, \varphi)\,, \tag{6.54}$$

where $Y_\ell^{(C)}(\theta, \varphi)$ is some function that depends on θ and φ *but not on* r.[41] The subscript ℓ and superscript (C) on Y indicate that it depends on the specific choice for the constant ℓ-index tensor $C_{i_1 \cdots i_\ell}$.

Since we have seen that $\nabla^2 \psi_\ell^{(C)} = 0$, and since the Laplacian in spherical polar coordinates is given by

$$\nabla^2 = \frac{1}{r^2}\, \frac{\partial}{\partial r}\Big(r^2\, \frac{\partial}{\partial r}\Big) + \frac{1}{r^2}\, \nabla^2_{(\theta,\varphi)} = 0\,, \tag{6.55}$$

it follows that

$$\Big[\frac{1}{r^2}\, \frac{\partial}{\partial r}\Big(r^2\, \frac{\partial}{\partial r}\Big) \frac{1}{r^{\ell+1}}\Big]\, Y_\ell^{(C)}(\theta, \varphi) + \frac{1}{r^{\ell+3}}\, \nabla^2_{(\theta,\varphi)}\, Y_\ell^{(C)}(\theta, \varphi) = 0\,, \tag{6.56}$$

and hence

$$\frac{\ell(\ell+1)}{r^{\ell+3}}\, Y_\ell^{(C)}(\theta, \varphi) + \frac{1}{r^{\ell+3}}\, \nabla^2_{(\theta,\varphi)}\, Y_\ell^{(C)}(\theta, \varphi) = 0\,. \tag{6.57}$$

(Recall that $\nabla^2_{(\theta,\varphi)} = \frac{1}{\sin\theta}\, \frac{\partial}{\partial\theta}(\sin\theta \frac{\partial}{\partial\theta}) + \frac{1}{\sin^2\theta}\, \frac{\partial^2}{\partial\varphi^2}$ is the Laplacian on the unit sphere.) In other words, we have shown that $Y_\ell^{(C)}(\theta, \varphi)$, which is defined by (6.52) and (6.54), satisfies

$$\nabla^2_{(\theta,\varphi)}\, Y_\ell^{(C)}(\theta, \varphi) = -\ell(\ell+1) Y_\ell^{(C)}(\theta, \varphi)\,. \tag{6.58}$$

[41] One way to see this is to note that r has the dimensions of length, i.e. $[r] = L$, whilst ∂_i has dimension L^{-1}. Therefore if we take the coefficients $C_{i_1 \cdots i_\ell}$ to be dimensionless we have that $[\psi_\ell^{(C)}] = L^{-\ell-1}$. Since $\psi_\ell^{(C)}$ can be written in terms of r, θ and φ, and since θ and φ are dimensionless, it must be that $\psi_\ell^{(C)}$ is of the form $r^{-\ell-1}$ multiplied by a (necessarily dimensionless) function of θ and φ.

This is precisely the equation satisfied by the spherical harmonics $Y_{\ell m}(\theta, \varphi)$. So we have shown that the construction (6.52), together with (6.54), gives spherical harmonics.

It remains to verify that we get the full set of spherical harmonics by this means. In fact, we essentially already did this in section (6.2). We saw there that there are $(2\ell + 1)$ independent components to an ℓ-index symmetric traceless tensor. Thus, the construction in (6.52), together with (6.54), yields all $(2\ell + 1)$ spherical harmonics at each level ℓ. We obtain them all by considering all possible ℓ-index symmetric traceless tensors $C_{i_1 \cdots i_\ell}$. The relation between this construction and the usual one can be seen for the first few values of ℓ by comparing (4.122) and (4.125).

For those who are familiar with the theory of Lie algebras and groups, it is worth remarking that what has been done in the construction described above is to obtain the spherical harmonics as irreducible representations of the $SO(3)$ rotational symmetry group of the two dimensional sphere, described as the surface $x^i x^i = 1$ embedded in three-dimensional Euclidean space. Specifically, at level ℓ the representation in question is the symmetric, traceless ℓ-index tensor representation, which is $(2\ell + 1)$ dimensional.

It is also worth remarking that the same technique can be applied in any dimension. By this means one can easily construct the hyperspherical harmonics on an n-dimensional sphere. They will be organised as ℓ-index symmetric tensor representations of the $SO(n+1)$ rotational symmetry group of the n-sphere, with eigenvalues

$$\lambda = \ell\,(\ell + n - 1)\,. \tag{6.59}$$

6.5 Multipole expansion of the energy in an external field

Recall that the energy U of a localised charge distribution was discussed in section 1.9, and shown to be given by (1.85)

$$U_{\rm int} = \tfrac{1}{2} \int \rho(\vec{r})\phi(\vec{r}) d^3\vec{r}. \tag{6.60}$$

We have added the subscript "int" here to emphasise that that result gave the "self energy" or "internal energy" of the charge distribution itself, in its own self-generated electrostatic field.

A different question, with a different answer, concerns the energy of a charge distribution in an *externally-applied* electrostatic field. If the external field is expressed in terms of the potential $\Phi(\vec{r})$ (as opposed to the potential $\phi(\vec{r})$ in (6.60), which is the potential due to the charge distribution itself), then the "external" energy of the system is simply calculated by integrating up the energy of assembling all the charges that form the distribution $\rho(\vec{r})$.

This gives

$$U_{\rm ext} = \int \rho(\vec{r})\Phi(\vec{r})d^3\vec{r}. \tag{6.61}$$

We shall assume that the external electric field $\vec{E} = -\vec{\nabla}\Phi$ is generated by distant sources (i.e. distant charges), so that we can take $\nabla^2\Phi = 0$ in the region where the localised charge distribution ρ is non-zero. Let us choose the origin to lie in the vicinity of the localised charge distribution, and furthermore we assume that the external field is a *slowly varying* function of $\vec{r}$ in this region. We may then Taylor expand $\Phi(\vec{r})$, to give

$$\Phi(\vec{r}) = \Phi(0) + x_i\partial_i\Phi(0) + \tfrac{1}{2}x_ix_j\partial_i\partial_j\Phi(0) + \cdots . \tag{6.62}$$

(Of course $\partial_i\Phi(0)$ means $\partial_i\Phi(\vec{r})$ evaluated at $\vec{r} = 0$, and so on. In other words, first act with the partial derivatives on $\Phi(\vec{r})$, and then afterwards set $\vec{r}$ to zero.) Equation (6.62) can be written in terms of the external electric field as

$$\Phi(\vec{r}) = \Phi(0) - x_iE_i(0) - \tfrac{1}{2}x_ix_j\partial_iE_j(0) + \cdots . \tag{6.63}$$

Since we are assuming there are no sources for the external electric field within the localised region of interest, it follows that $\partial_iE_i = 0$, and so (6.63) may be re-expressed as

$$\Phi(\vec{r}) = \Phi(0) - x_iE_i(0) - \tfrac{1}{6}(3x_ix_j - r^2\delta_{ij})\partial_iE_j(0) + \cdots . \tag{6.64}$$

(i.e. the extra term we have added in is actually zero. We are making this trace subtraction for the usual reason, of wanting to deal with the true independent degrees of freedom in the multipole moments.)

Using the definitions (6.31) for the multipole moments, we see that when (6.64) is substituted into the expression (6.61) for the "external" energy, it gives

$$\boxed{U_{\rm ext} = Q\,\Phi(0) - p_i\,E_i(0) - \tfrac{1}{6}Q_{ij}\,\partial_iE_j(0) + \cdots .} \tag{6.65}$$

The first term is the familiar result for the energy of a charge Q in an electrostatic field. The second term is the energy $-\vec{p}\cdot\vec{E}$ of a dipole in an electric field. The third term, which depends on the gradient of the electric field, is the energy of a quadrupole in the external field. The energies for the higher multipole moments will be associated with higher derivatives of the electric field.

As an application of the result in equation (6.65), we may calculate the interaction energy between a pair of electric dipoles $\vec{p}_1$ and $\vec{p}_2$. Suppose they are located, respectively,

at points $\vec{r} = \vec{r}_1$ and $\vec{r} = \vec{r}_2$. From the expression (6.37) for the electric field due to a dipole, we see that the electric field at $\vec{r}_1$ due to a dipole moment $\vec{p}_2$ located at $\vec{r} = \vec{r}_2$ is given by

$$\vec{E}(\vec{r}_1) = \frac{3\vec{n}\,(\vec{n}\cdot\vec{p}_2) - \vec{p}_2}{|\vec{r}_1 - \vec{r}_2|^3}\,, \tag{6.66}$$

where $\vec{n}$ is the unit vector in the direction from $\vec{r}_2$ to $\vec{r}_1$. From (6.65), we then see that the energy of the dipole $\vec{p}_1$ in this electric field is given by

$$U_{12} = \frac{\vec{p}_1\cdot\vec{p}_2 - 3(\vec{n}\cdot\vec{p}_1)(\vec{n}\cdot\vec{p}_2)}{|\vec{r}_1 - \vec{r}_2|^3}\,. \tag{6.67}$$

As one would expect, this expression is completely symmetrical between $\vec{p}_1$ and $\vec{p}_2$.

7 Dielectric Media

7.1 Microscopic description

So far, we have considered situations in which isolated charges or isolated boundary surfaces are present in an otherwise free space (vacuum). In principle, to the extent that classical electromagnetism can be applied at all on the scale of atomic or subatomic particles, one could describe the electrostatic fields in any configuration of matter by means of the Maxwell equations in the form we have been using so far. However, this description would become extremely unwieldy in cases where one wanted to calculate the electric field inside a lump of matter.

Suppose, for example, we wanted to study the electric field inside an insulator such as salt crystal. At some level one could say that there is a huge array of positive Na ions and negative Cl ions arranged in a lattice. Close to an Na ion the electrostatic potential would grow to huge postive values, whilst near Cl ions the potential would grow to huge negative values. The electric field would, correspondingly, be fluctuating wildly in magnitude and direction, as a function of position within the crystal.

These huge fluctuations would be occurring on the atomic length scale. However, as we know very well, on a *macroscopic* length scale one sees no direct evidence of the wild goings-on at the atomic scale, and it must therefore be possible to *average out* over some appropriate intermediate length scale that lies somewhere between the atomic and the macroscopic scales. Having done this, one should arrive at a microscopic understanding of why the salt crystal has the properties that it does, and in the process one should obtain a phenomenological description of its macroscopic properties.

As far as studying electrostatics in the presence of the salt crystal is concerned, one should arrive at a macroscopic system of modified Maxwell equations that incorporate the

phenomenology whose underlying understanding stems from the appropriate averaging-out procedure. It will be crucial, in what follows, that this averaging procedure can be performed a scale that is very large in comparison to the atomic or molecular scale (which is of order 10^{-8} centimetres), and yet still very small compared to the macroscopic scale of the dielectric material (which is perhaps of order 1 centimetre).

Let us denote the actual, atomic-scale, electric field and charge density by $\vec{\mathcal{E}}$ and $\tilde{\rho}$. These will satisfy the standard free-space Maxwell equations

$$\vec{\nabla} \cdot \vec{\mathcal{E}} = 4\pi\tilde{\rho}\,, \qquad \vec{\nabla} \times \vec{\mathcal{E}} = 0\,. \tag{7.1}$$

From our earlier results, we know that $\vec{\mathcal{E}}$ can be calculated from $\tilde{\rho}$ using

$$\vec{\mathcal{E}}(\vec{r}) = -\vec{\nabla} \int \frac{\tilde{\rho}(\vec{r}\,')}{|\vec{r} - \vec{r}\,'|}\, d^3\vec{r}\,' = \int \frac{\vec{r} - \vec{r}\,'}{|\vec{r} - \vec{r}\,'|^3}\, \tilde{\rho}(\vec{r}\,')\, d^3\vec{r}\,'\,. \tag{7.2}$$

This, however, is the field mentioned above that fluctuates wildly, at the atomic length scale, as a function of $\vec{r}$.

We want to consider averaging over scales that are small by macroscopic standards, but which are huge in relation to the atomic length scale. Thus we consider averaged quantities

$$\begin{aligned} \langle \vec{\mathcal{E}}(\vec{r}) \rangle &= \frac{1}{\Delta V} \int_{\Delta V} \vec{\mathcal{E}}(\vec{r} + \vec{\xi})\, d^3\vec{\xi}\,, \\ \langle \tilde{\rho}(\vec{r}) \rangle &= \frac{1}{\Delta V} \int_{\Delta V} \tilde{\rho}(\vec{r} + \vec{\xi})\, d^3\vec{\xi}\,, \end{aligned} \tag{7.3}$$

where ΔV is some macroscopically small volume (in comparison to the volume of the entire piece of dielectric material) that nonetheless contains a huge number of atoms.

Note that actually, the ions in the substance are in motion and so one might think that it would be inappropriate to be trying to describe the situation purely by means of electrostatics. However, this motion is random in nature, resulting from thermal excitations, and so once the averaging is performed there will be no significant effects resulting from it. (At least, assuming that the temperatures are not so high that the motion could actually lead to significant electromagnetic radiation.)

Consider a substance composed of atoms or molecules, and first focus attention on the electric field due to the distribution of charge $\tilde{\rho}_\alpha$ in the α'th molecule, assumed to be centred at $\vec{r}_\alpha$. The potential outside it will be given by

$$\phi_\alpha(\vec{r}) = \int_{\text{Mol.}} \frac{\tilde{\rho}_\alpha(\vec{\zeta})}{|\vec{r} - \vec{r}_\alpha - \vec{\zeta}\,|}\, d^3\vec{\zeta}\,, \tag{7.4}$$

where the $\vec{\zeta}$ integration is over positions within the molecule. (Note that the charge density $\tilde{\rho}_\alpha(\vec{\zeta})$ of the α'th molecule is taken to be concentrated around $\vec{\zeta} = 0$.) Thus, the electric

field due to the α'th molecule will be

$$\vec{\mathcal{E}}_\alpha(\vec{r}) = -\vec{\nabla} \int_{\text{Mol.}} \frac{\tilde{\rho}_\alpha(\vec{\zeta})}{|\vec{r} - \vec{r}_\alpha - \vec{\zeta}|}\, d^3\vec{\zeta}\,, \tag{7.5}$$

Applying Taylor's theorem as we did earlier when considering the multipole expansion, with

$$\frac{1}{|\vec{r} - \vec{r}_\alpha - \vec{\zeta}|} = \frac{1}{|\vec{r} - \vec{r}_\alpha|} - \vec{\zeta}\cdot\vec{\nabla}\frac{1}{|\vec{r} - \vec{r}_\alpha|} + \cdots\,, \tag{7.6}$$

we find

$$\vec{\mathcal{E}}_\alpha(\vec{r}) \;\; = \;\; -\vec{\nabla}\left[\frac{q_\alpha}{|\vec{r} - \vec{r}_\alpha|} - \vec{p}_\alpha\cdot\vec{\nabla}\Big(\frac{1}{|\vec{r} - \vec{r}_\alpha|}\Big) + \cdots\right], \tag{7.7}$$

where q_α is the total charge of the α'th molecule and $\vec{p}_\alpha$ is its electric dipole moment:

$$q_\alpha = \int_{\text{Mol.}} \tilde{\rho}_\alpha(\vec{\zeta})d^3\vec{\zeta}\,, \qquad \vec{p}_\alpha = \int_{\text{Mol.}} \vec{\zeta}\,\tilde{\rho}_\alpha(\vec{\zeta})d^3\vec{\zeta}\,. \tag{7.8}$$

Note that typically q_α, the total charge of the α'th molecule, will be zero (because the positive and negative charges will typically balance). It is useful in any case to retain this term for now, because it is slightly pedagogically easier to understand the arguments that will now follow for this term before considering the analogous arguments for the dipole term that we are principally interested in. The total microscopic electric field is obtained by summing over the molecules:

$$\vec{\mathcal{E}}(\vec{r}) = -\vec{\nabla}\sum_\alpha \left[\frac{q_\alpha}{|\vec{r} - \vec{r}_\alpha|} - \vec{p}_\alpha\cdot\vec{\nabla}\Big(\frac{1}{|\vec{r} - \vec{r}_\alpha|}\Big)\right], \tag{7.9}$$

It is helpful at this stage to replace the discrete sum over molecules by a continuous integral, which is achieved in the standard way by representing their charges by a charge density with delta functions, and similarly for their dipole moments:

$$\rho_{\text{mol}}(\vec{r}) = \sum_\alpha q_\alpha\,\delta^3(\vec{r} - \vec{r}_\alpha)\,, \qquad \vec{\pi}_{\text{mol}}(\vec{r}) = \sum_\alpha \vec{p}_\alpha\,\delta^3(\vec{r} - \vec{r}_\alpha)\,. \tag{7.10}$$

Thus (7.9) becomes

$$\begin{aligned}\vec{\mathcal{E}}(\vec{r}) \;\; &= \;\; -\vec{\nabla}\int d^3\vec{r}\,'\left[\frac{\rho_{\text{mol}}(\vec{r}\,')}{|\vec{r} - \vec{r}\,'|} - \vec{\pi}_{\text{mol}}(\vec{r}\,')\cdot\vec{\nabla}\Big(\frac{1}{|\vec{r} - \vec{r}\,'|}\Big)\right], \\ &= \;\; -\vec{\nabla}\int d^3\vec{r}\,'\left[\frac{\rho_{\text{mol}}(\vec{r}\,')}{|\vec{r} - \vec{r}\,'|} + \vec{\pi}_{\text{mol}}(\vec{r}\,')\cdot\vec{\nabla}'\Big(\frac{1}{|\vec{r} - \vec{r}\,'|}\Big)\right],\end{aligned} \tag{7.11}$$

where the integration of $\vec{r}\,'$ now ranges over the entire volume of the substance. Note that in the last line $\vec{\nabla}'$ means the gradient with respect to the primed coordinates $\vec{r}\,'$, and we have used the fact that $\vec{\nabla} f(\vec{r} - \vec{r}\,') = -\vec{\nabla}'\, f(\vec{r} - \vec{r}\,')$ for any function of $\vec{r} - \vec{r}\,'$.

We now wish to perform the averaging procedure (7.3). Consider first the contribution from the charge term in (7.11); we shall then make an analogous analysis for the dipole term. From (7.3), the charge contribution in (7.11) will give

$$\langle \vec{\mathcal{E}}_{\rm charge}(\vec{r})\rangle = -\vec{\nabla}\left[\frac{1}{\Delta V}\int_{\Delta V} d^3\vec{\xi}\int d^3\vec{r}\,'\,\frac{\rho_{\rm mol}(\vec{r}\,')}{|\vec{r}+\vec{\xi}-\vec{r}\,'|}\right]. \tag{7.12}$$

By shifting variables according to

$$\vec{r}\,' \longrightarrow \vec{r}\,' + \vec{\xi}, \tag{7.13}$$

this becomes

$$\langle \vec{\mathcal{E}}_{\rm charge}(\vec{r})\rangle = -\vec{\nabla}\left[\frac{1}{\Delta V}\int_{\Delta V} d^3\vec{\xi}\int d^3\vec{r}\,'\,\frac{\rho_{\rm mol}(\vec{r}\,'+\vec{\xi})}{|\vec{r}-\vec{r}\,'|}\right]. \tag{7.14}$$

Note that the integration range for $\vec{\xi}$, which is over the averaging volume ΔV, is very small compared to the integration range for $\vec{r}\,'$, which is over the entire volume of the piece of dielectric material. This crucial observation means that when the change of variables (7.13) is performed, the associated change in the region over which $\vec{r}\,'$ is to be integrated is negligible, and so we may continue, with only a tiny error, to take the $\vec{r}\,'$ integral to be over the same entire volume of the material as before. This step in the calculation is a key one in the entire argument; an averaging over the $1/|\vec{r}-\vec{r}\,'|$ denominators in the $\vec{r}\,'$ integral has been turned into an averaging over the molecular charge density. It depends crucially on the fact that we don't need to bother about changing the integration range for the $\vec{r}\,'$ integration, because the change is so negligible.

After the change of variable in eqn (7.13), we then consider first the ξ integration. If we denote by $\langle q_{\rm mol}(\vec{r}\,')\rangle$ the average charge per molecule within the volume ΔV at $\vec{r}\,'$, and if we denote by $N(\vec{r}\,')$ the macroscopic number density of molecules at $\vec{r}\,'$, then we shall have

$$\frac{1}{\Delta V}\int_{\Delta V} d^3\vec{\xi}\,\rho_{\rm mol}(\vec{r}\,'+\vec{\xi}) = N(\vec{r}\,')\,\langle q_{\rm mol}(\vec{r}\,')\rangle\,. \tag{7.15}$$

The contribution (7.14) may then be written as

$$\langle \vec{\mathcal{E}}_{\rm charge}(\vec{r})\rangle = -\vec{\nabla}\int \frac{N(\vec{r}\,')\,\langle q_{\rm mol}(\vec{r}\,')\rangle}{|\vec{r}-\vec{r}\,'|}\,d^3\vec{r}\,'\,. \tag{7.16}$$

In a similar manner, we introduce the average dipole moment per molecule $\langle \vec{p}_{\rm mol}(\vec{r}\,')\rangle$ at $\vec{r}\,'$, so that

$$\frac{1}{\Delta V}\int_{\Delta V} d^3\vec{\xi}\,\vec{\pi}_{\rm mol}(\vec{r}\,'+\vec{\xi}) = N(\vec{r}\,')\,\langle \vec{p}_{\rm mol}(\vec{r}\,')\rangle\,. \tag{7.17}$$

The total average of the electric field (7.11) is therefore given by

$$\langle \vec{\mathcal{E}}(\vec{r})\rangle = -\vec{\nabla}\int N(\vec{r}\,')\left[\frac{\langle q_{\rm mol}(\vec{r}\,')\rangle}{|\vec{r}-\vec{r}\,'|} + \langle \vec{p}_{\rm mol}(\vec{r}\,')\rangle\cdot\vec{\nabla}'\left(\frac{1}{|\vec{r}-\vec{r}\,'|}\right)\right] d^3\vec{r}\,'\,. \tag{7.18}$$

Taking the divergence of (7.18), and recalling that $\nabla^2|\vec{r}-\vec{r}\,'|^{-1} = -4\pi\delta^3(\vec{r}-\vec{r}\,')$, gives

$$\vec{\nabla}\cdot\langle\vec{\mathcal{E}}(\vec{r})\rangle = 4\pi\int N(\vec{r}\,')\left[\langle q_{\rm mol}(\vec{r}\,')\rangle\delta^3(\vec{r}-\vec{r}\,') + \langle\vec{p}_{\rm mol}(\vec{r}\,')\rangle\cdot\vec{\nabla}'\,\delta^3(\vec{r}-\vec{r}\,')\right]\,d^3\vec{r}\,'\,. \quad (7.19)$$

Using the delta-functions to perform the $\vec{r}\,'$ integrations[42] then gives

$$\vec{\nabla}\cdot\langle\vec{\mathcal{E}}(\vec{r})\rangle = 4\pi N(\vec{r})\langle q_{\rm mol}(\vec{r})\rangle - 4\pi\vec{\nabla}\cdot\Big(N(\vec{r})\langle\vec{p}_{\rm mol}(\vec{r})\rangle\Big)\,. \quad (7.20)$$

Equation (7.20) is the phenomenological replacement for the Maxwell equation $\vec{\nabla}\cdot\vec{E} = 4\pi\rho$, in the case where we average out over molecules in a dielectric medium. Effectively, the usual charge density on the right-hand side is replaced by the sum of two terms. The first is the average charge per unit volume of the molecules themselves, and the second is the average polarisation charge per unit volume. The equation can be rewritten by taking this second term over to the left-hand side, giving

$$\vec{\nabla}\cdot\Big(\langle\vec{\mathcal{E}}(\vec{r})\rangle + 4\pi N(\vec{r})\langle\vec{p}_{\rm mol}(\vec{r})\rangle\Big) = 4\pi N(\vec{r})\langle q_{\rm mol}(\vec{r})\rangle\,. \quad (7.21)$$

We may now define macroscopic quantities as follows:

$$\vec{E}\equiv\langle\vec{\mathcal{E}}\rangle\,,\qquad \vec{P}\equiv N\,\langle\vec{p}_{\rm mol}\rangle\,,\qquad \hat{\rho}\equiv N\langle q_{\rm mol}\rangle\,,\qquad \vec{D}\equiv\vec{E}+4\pi\vec{P}\,. \quad (7.22)$$

$\vec{E}$ is called the electric field; $\vec{P}$ is the *polarisation* (i.e. the electric dipole moment per unit volume); $\vec{D}$ is called the *displacement*; and $\hat{\rho}$ is the charge density in the dielectric material itself. Equation (7.21) is now written, in terms of these new quantities, as

$$\vec{\nabla}\cdot\vec{D} = 4\pi\hat{\rho}\,. \quad (7.23)$$

If there are various different types of molecules or atoms comprising the medium, and if, furthermore, there are additional external charges present, then $\vec{P}$ and $\hat{\rho}$ defined above admit a natural generalisation, obtained by summing over all the contributions:

$$\vec{P} = \sum_i N_i\langle\vec{p}_i\rangle\,,\qquad \hat{\rho} = \rho + \sum_i N_i\langle q_i\rangle\,, \quad (7.24)$$

where ρ is the *external* charge density. Commonly, as mentioned previously, the molecules themselves are electrically neutral, and so the total charge density $\hat{\rho}$ is just given by the external charge density ρ.

[42] Recall that, using an integration by parts, one has the result for a one-dimensional delta-function that $\int f(x')(\partial\delta(x-x')/\partial x')\,dx' = -\int(\partial f(x')/\partial x')\,\delta(x-x')dx' = -\partial f(x)/\partial x$. Analogously, in three dimensions, one has $\int\vec{f}(\vec{r}\,')\cdot\vec{\nabla}'\delta^3(\vec{r}-\vec{r}\,')d^3\vec{r}\,' = -\vec{\nabla}\cdot\vec{f}(\vec{r})$ for any vector field $\vec{f}(\vec{r})$.

In terms of these definition, we see that the effective phenomenological Maxwell equations describing electrostatics in the presence of dielectric media will then be

$$\vec{\nabla}\cdot\vec{D} = 4\pi\rho\,, \qquad \vec{\nabla}\times\vec{E} = 0\,, \tag{7.25}$$

where ρ is just the density of free charges "external" to the molecular structure of the material. (The second equation follows by taking the curl of (7.18).)

It is evident from the second equation in (7.25) that $\vec{E}$ can still be written in terms of a potential,

$$\vec{E} = -\vec{\nabla}\phi\,. \tag{7.26}$$

Conversely, by integrating $\vec{E}$ along a path, one still obtains the potential difference between the initial and final points:

$$\int_A^B \vec{E}\cdot d\vec{\ell} = -\int \vec{\nabla}\phi\cdot d\vec{\ell} = \phi(A) - \phi(B)\,. \tag{7.27}$$

Since $\vec{D} = \vec{E} + 4\pi\vec{P}$, and $\vec{\nabla}\cdot\vec{D} = 4\pi\rho$, we have from $\vec{E} = -\vec{\nabla}\phi$ that $\nabla^2\phi = -4\pi(\rho - \vec{\nabla}\cdot\vec{P}\,)$, and so the potential ϕ can be solved for, in terms of the macroscopic quantities, as

$$\phi(\vec{r}\,) = \int d^3\vec{r}\,' \,\frac{[\rho(\vec{r}\,') - \vec{\nabla}'\cdot\vec{P}(\vec{r}\,')]}{|\vec{r}-\vec{r}\,'|}\,. \tag{7.28}$$

Thus we have

$$\phi(\vec{r}\,) = \int d^3\vec{r}\,' \,\frac{\rho(\vec{r}\,') + \rho_{\rm pol.}(\vec{r}\,')}{|\vec{r}-\vec{r}\,'|}\,, \tag{7.29}$$

where $\rho_{\rm pol.}$, defined by

$$\rho_{\rm pol.}(\vec{r}\,) = -\vec{\nabla}\cdot\vec{P}(\vec{r}\,)\,, \tag{7.30}$$

is called the *polarisation charge density*.

7.2 Examples of dielectric media

The polarisation $\vec{P}$ of the molecules or atoms of which a medium is composed is related to the local electric field in their neighbourhood. Typically, if no external electric field is applied the polarisation will be zero. To a very good approximation, the polarisation will increase linearly with increasing applied electric field. Thus, using the index notation, and Einstein summation convention, for vector and tensor fields, we can say that

$$P_i = \chi_{ij}\, E_j\,, \tag{7.31}$$

where the tensor χ_{ij} is independent of $\vec{E}$. It is known as the *electric susceptibility tensor*. In general, in a medium that is not homogeneous, it can depend on position. If the medium is homogeneous (i.e. it is the same throughout), then χ_{ij} will be a constant tensor.

If the medium is in addition *isotropic*, then the polarisation will be parallel to the electric field.[43] In such cases, we shall have

$$\chi_{ij} = \chi\, \delta_{ij}\,, \tag{7.32}$$

where χ is called the *electric susceptibility* of the medium. Then we shall have

$$\vec{P} = \chi\, \vec{E}\,. \tag{7.33}$$

From the definition of $\vec{D}$ in (7.22), we then have

$$\vec{D} = \epsilon\, \vec{E}\,, \tag{7.34}$$

where

$$\epsilon \equiv 1 + 4\pi\, \chi \tag{7.35}$$

is called the *dielectric constant* of the medium.

If we assume, therefore, a homogeneous isotropic medium, we shall have (7.34) for some constant ϵ, and so the effective Maxwell equation for $\vec{D}$ given in (7.25) implies

$$\vec{\nabla}\cdot\vec{E} = \frac{4\pi}{\epsilon}\,\rho\,. \tag{7.36}$$

This means that, compared with the free-space solutions that we discussed in previous chapters, the solutions in the presence of the medium will be exactly the same, except that $\vec{E}$ is scaled down by a factor of $1/\epsilon$. (The dielectric constant is usually greater than 1.)

For example, suppose we place a point charge q in an infinite medium of dielectric constant ϵ. Choosing coordinates so that the charge is located at the origin, it follows that the electrostatic potential will be given by

$$\phi = \frac{q}{\epsilon\, r}\,, \tag{7.37}$$

and therefore that the electric field $\vec{E} = -\vec{\nabla}\phi$ is given by

$$\vec{E} = \frac{q\,\vec{r}}{\epsilon\, r^3}\,. \tag{7.38}$$

The physical interpretation of what is happening in the medium is that when an external electric field is applied, the charges in the atoms or molecules tend to be displaced slightly, in such a way that the positive charges are pulled in the direction of the lower electrostatic potential, whilst the negative charges are pulled in the direction of the larger potential. This induces a dipole moment that points in the opposite direction to the electric field. In other words, the the dipole moment that develops when external charges are introduced is such as to tend to *oppose* the external charges.

[43] There are substances, such as calcite or quartz, which ar non-isotropic, and for these the electric susceptibility is necessarily described by a tensor.

7.3 Boundary-value problems with dielectric interfaces

In section 1.4, we derived the boundary conditions that must hold at an interface between two media. In the case of electrostatics, with dielectric media, the relevant conditions are (see eqn 1.34))

$$\vec{n} \cdot (\vec{D}_2 - \vec{D}_1) = 4\pi\sigma\,, \qquad \vec{n} \times (\vec{E}_2 - \vec{E}_1) = 0\,, \tag{7.39}$$

where the subscripts 1 and 2 indicate the fields on either side of the interface, in medium 1 and medium 2 respectively; $\vec{n}$ is the unit normal vector at the interface, pointing from medium 1 to medium 2; and σ is the surface charge density at the interface. Note that σ describes just the actual surface density of *free* charges; it does not include polarisation charges.

The two conditions in (7.39) came from integrating the phenomenological Maxwell equation $\vec{\nabla} \cdot \vec{D} = 4\pi\rho$ in (7.25) over a Gaussian pillbox straddling the interface, and integrating $\vec{\nabla} \times \vec{E} = 0$ around a loop straddling the interface, respectively. The first condition says that there is a discontinuity in the normal component of $\vec{D}$, given by 4π times the surface charge density. The second condition says that the tangential components of the electric field $\vec{E}$ must be continuous across the interface.

We now consider several example electrostatic problems involving dielectric media.

7.3.1 Method of images for two semi-infinite dielectrics

Suppose that a medium with dielectric constant ϵ_1 fills the entire half-space $z > 0$, and that a medium with dielectric constant ϵ_2 fills the other half of space, for $z < 0$. Thus we have an interface in the plane $z = 0$. Suppose that a point charge q is placed at the point $\vec{r} = (0, 0, d)$ in dielectric 1, and suppose that there is no free surface charge density at the interface, so $\sigma = 0$. The problem is to solve for the electric field everywhere.

From (7.25), the equations to be solved are

$$\begin{aligned}
\epsilon_1\, \vec{\nabla} \cdot \vec{E} &= 4\pi\rho = 4\pi q \delta(x)\delta(y)\delta(z-d)\,, \qquad z > 0\,,\\
\epsilon_2\, \vec{\nabla} \cdot \vec{E} &= 0\,, \qquad z < 0\,,\\
\vec{\nabla} \times \vec{E} &= 0\,, \qquad \text{all } z\,.
\end{aligned} \tag{7.40}$$

Since $\vec{\nabla} \times \vec{E} = 0$, we can write the electric field in terms of a scalar potential, $\vec{E} = -\vec{\nabla}\phi$,

as always. The boundary conditions (7.39) imply that

$$\begin{aligned} \epsilon_1 E_z(x,y,0^+) &= \epsilon_2 E_z(x,y,0^-)\,, \\ E_x(x,y,0^+) &= E_x(x,y,0^-)\,, \\ E_y(x,y,0^+) &= E_y(x,y,0^-)\,, \end{aligned} \tag{7.41}$$

where 0^+ indicates 0 approached from above (and so $z = 0^+$ means that z is in medium 1), and 0^- indicates 0 approached from below (medium 2).

In section (2.4), we solved the problem of a point charge q above an infinite planar conductor by introducing an image charge $-q$ at the mirror-image reflection point below the conductor. Here, we can apply a similar image-charge technique. However, because of the fact that we now have dielectric media rather than a conducting plane, the necessary procedure is less intuitively clear. A crucial point, though, is that, as we saw earlier, there is a *uniqueness theorem* in electrostatics which means that if, by hook or by crook, we obtain a solution to the equations that *works*, then we are guaranteed that it is the unique and correct solution to the problem. (By "works," we mean that is satisfies the equations and the given boundary conditions.)

The approach we shall take here in solving the problem will be to try some plausible guesses, and discover that they work. This then means that the problem is solved.

To proceed, we must choose suitable "trial solutions" in each of the two regions $z > 0$ and $z < 0$.

Considering the region $z > 0$ first, it is natural, by analogy with the usual method of images procedure, to guess that here we should introduce an image charge at the mirror-reflection of the location of the actual charge q; i.e. introduce an image charge at $\vec{r} = (0, 0, -d)$. However, clearly the *value* of the image charge will no longer in general be simply $-q$.[44] Thus for $z > 0$ we shall look for a solution of the form

$$\phi = \frac{q}{\epsilon_1\, R_1} + \frac{\alpha}{R_2}\,, \tag{7.42}$$

where α is an unknown constant, and

$$R_1^2 = x^2 + y^2 + (z-d)^2\,, \qquad R_2^2 = x^2 + y^2 + (z+d)^2\,. \tag{7.43}$$

It is convenient to reparameterise the unknown constant α by writing $\alpha = q'/\epsilon_1$, so the trial

[44] This is obvious from the fact that if $\epsilon_1 = \epsilon_2 = 1$, so that the original charge q is in an infinite free space, no image charge at all is needed. On the other hand, if $\epsilon_1 = 1$ and $\epsilon_2 = \infty$, we are back to the situation of the infinite planar conductor, for which we know the image charge should be $-q$.

solution for $z > 0$ will be

$$\boxed{\phi_> = \frac{1}{\epsilon_1}\Big(\frac{q}{R_1} + \frac{q'}{R_2}\Big)\,, \qquad z > 0\,,} \tag{7.44}$$

where q' is an as-yet unknown constant.

The expression (7.44) is intended to apply *only* to observation points in the upper half-space. We also need an expression for the potential that is valid instead in the lower half-space, $z < 0$. Here, we make the plausible guess that it takes the form

$$\phi = \frac{\beta}{R_1}\,, \tag{7.45}$$

where β is an unknown constant. In other words, we are guessing that the potential in the region $z < 0$ has the form that would result from a charge placed at the location of the original physical charge q, but with some as-yet unknown coefficient. (Note that the trial expression for ϕ in the region $z < 0$) should certainly *not* contain any charges that are themselves located in the region $z < 0$; if there were any such charge, it would imply (incorrectly) that the potential in $z < 0$ would diverge at the location of that charge.) It is convenient to reparameterise the unknown constant β as $\beta = q''/\epsilon_2$, where q'' is an unknown constant, and so in the region $z < 0$ we choose the trial solution

$$\boxed{\phi_< = \frac{q''}{\epsilon_2 R_1}\,, \qquad z < 0\,,} \tag{7.46}$$

where q'' is an as-yet unknown constant.

In case the reader is feeling doubtful at this point, recall again that the *only* thing that matters in the end is that the proposed solution should (a) satisfy the equations (7.25), and (b) satisfy the boundary conditions (7.41) and at infinity. Thus, we take (7.44) and (7.46) as the guess[45], and now check to see if we can choose q' and q'' so that the boundary conditions (7.41) are satisfied.

To impose the boundary conditions, we need to calculate the derivatives of $1/R_1$ and $1/R_2$ with respect to x, y and z, evaluated at the interface $z = 0$. We see that

$$\begin{aligned}
\frac{\partial}{\partial x}\Big(\frac{1}{R_1}\Big)\Big|_{z=0} = \frac{\partial}{\partial x}\Big(\frac{1}{R_2}\Big)\Big|_{z=0} &= -\frac{x}{(x^2+y^2+d^2)^{3/2}}\,,\\
\frac{\partial}{\partial y}\Big(\frac{1}{R_1}\Big)\Big|_{z=0} = \frac{\partial}{\partial y}\Big(\frac{1}{R_2}\Big)\Big|_{z=0} &= -\frac{y}{(x^2+y^2+d^2)^{3/2}}\,,\\
\frac{\partial}{\partial z}\Big(\frac{1}{R_1}\Big)\Big|_{z=0} = -\frac{\partial}{\partial z}\Big(\frac{1}{R_2}\Big)\Big|_{z=0} &= \frac{d}{(x^2+y^2+d^2)^{3/2}}\,.
\end{aligned} \tag{7.47}$$

[45] Or *ansatz*, to use a slightly more scientific-sounding word that means a guess.

Thus we see that the boundary conditions (7.41) imply

$$\begin{aligned} q'' &= q - q' \,, \\ \frac{1}{\epsilon_1}\,(q + q') &= \frac{1}{\epsilon_2}\, q'' \,. \end{aligned} \tag{7.48}$$

The solution to these equations is

$$\boxed{q' = \frac{\epsilon_1 - \epsilon_2}{\epsilon_1 + \epsilon_2}\, q \,, \qquad q'' = \frac{2\epsilon_2}{\epsilon_1 + \epsilon_2}\, q \,.} \tag{7.49}$$

Note that it is non-trivial that we were able to find a solution by this means. the boundary conditions (7.41) must be satisfied at the junction $z = 0$ *for all values of* x *and* y, and it was not *a priori* guaranteed that this would have worked out OK, given the ansatz we were making.

Observe that (7.49) clearly makes sense in two limiting cases. Firstly, consider the case when $\epsilon_1 = \epsilon_2$, meaning that all of space is filled with a single dielectric medium. Obviously, there is no interface at all in this case, and so no image charge is needed. And indeed, (7.49) gives $q' = 0$. Furthermore, the formula for the potential that is valid when $z < 0$ should be the same as the one valid for $z > 0$. And indeed, $q'' = q$ in this case.

Secondly, consider the limit when $\epsilon_2 = \infty$. This means that effectively, the region $z < 0$ is just a semi-infinite slab of conductor. And indeed, we see that the expression for the potential in the region $z > 0$ just reduces to the original image-charge result for an infinite conductor, as discussed in section (2.4).

In the derivation above, we have been careful not to attach any direct physical significance to the "fictitious" charges q' and q'' that were introduced in order to obtain the correct solution for the electrostatic potential in the two half-spaces $z > 0$ and $z < 0$. Observe, indeed, that it was never necessary to give them any direct interpretation; they are just coefficients of terms we introduced in the candidate expressions for the potential in the two half-spaces, and by "lucky chance," it turned out that by tuning q' and q'' appropriately, we were able to solve the problem.

By definition, the *polarisation charge density* is given by (see (7.30))

$$\rho_{\rm pol} = -\vec{\nabla} \cdot \vec{P} \,. \tag{7.50}$$

Inside each dielectric, we have (see (7.33))

$$\vec{P} = \chi\, \vec{E} \,, \tag{7.51}$$

where $\chi = (\epsilon_1 - 1)/(4\pi)$ in medium 1, and $\chi = (\epsilon_2 - 1)/(4\pi)$ in medium 2. Since $\vec{\nabla} \cdot \vec{E} = 0$ everywhere in each region (except for the location of the physical charge q), it follows that the polarisation charge density is zero inside each medium.

However, there is a discontinuity in χ at the interface $z = 0$; it jumps by an amount

$$\Delta\chi = \frac{\epsilon_1 - \epsilon_2}{4\pi}\,. \tag{7.52}$$

This implies that there is a *surface* polarisation charge density $\sigma_{\rm pol}$ on the boundary surface, given by integrating (7.50) over a pill-box straddling the interface at $z = 0$:

$$\sigma_{\rm pol} = -(\vec{P}_2 - \vec{P}_1) \cdot \vec{n}\,, \tag{7.53}$$

where $\vec{n}$ is the unit normal vector pointing from medium 1 to medium 2. $\vec{P}_1$ and $\vec{P}_2$ are the polarisations in the two media, either side of the boundary surface at $z = 0$.

With

$$\vec{P}_i = \frac{\epsilon_i - 1}{4\pi}\,\vec{E} = -\frac{\epsilon_i - 1}{4\pi}\,\vec{\nabla}\phi\,, \tag{7.54}$$

for $i = 1$ or 2 in the two media, we see that

$$\sigma_{\rm pol} = -\frac{q\,(\epsilon_2 - \epsilon_1)}{2\pi\epsilon_1(\epsilon_2 + \epsilon_1)}\,\frac{d}{(x^2 + y^2 + d^2)^{3/2}} = \frac{q'}{2\pi\epsilon_1}\,\frac{d}{(x^2 + y^2 + d^2)^{3/2}}\,. \tag{7.55}$$

If this is integrated over the entire $z = 0$ plane, it gives a total polarisation charge

$$Q_{\rm pol} = \frac{q'}{\epsilon_1}\,. \tag{7.56}$$

7.3.2 Dielectric sphere

Consider a sphere of radius a and dielectric constant ϵ, placed in an originally-uniform electric field $\vec{E}$. Thus we imagine that prior to introducing the sphere, there is a uniform electric field $\vec{E}$ directed along the z direction.[46] Let its magnitude be denoted by E_0. Thus in the absence of the dielectric sphere, we have

$$\phi = -E_0\,z\,, \qquad \Rightarrow \qquad \vec{E} = (0, 0, E_0)\,. \tag{7.57}$$

Note that in spherical polar coordinates we therefore have

$$\phi = -E_0\,r\cos\theta\,. \tag{7.58}$$

[46]The slightly vague statement of an "originally-uniform electric field" is one that one quite often encounters in the wording of electrostatics problems. Stated more precisely, what it means is an electric field whose asymptotic form at large distance is $\vec{E} \longrightarrow \vec{E}_0$, where $\vec{E}_0$ is a constant vector.

After we have introduced the spherical dielectric, we shall take (7.58) as the asymptotic form of the potential at large r. (This implements the notion of the "originally uniform electric field.")

Now we introduce the dielectric sphere. For convenience, we choose our coordinates so that it is centred on the origin of spherical polar coordinates. There will be no free charges either inside or outside the sphere, and so in each region we must have $\nabla^2\phi = 0$. In the regions inside and outside the sphere we therefore must have

$$\begin{aligned} r < a: \qquad & \phi_<(r,\theta) = \sum_{\ell\geq 0} A_\ell\, r^\ell\, P_\ell(\cos\theta)\,, \\ r > a: \qquad & \phi_>(r,\theta) = \sum_{\ell\geq 0} \Big(\widetilde{A}_\ell\, r^\ell + B_\ell\, r^{-\ell-1}\Big)\, P_\ell(\cos\theta)\,. \end{aligned} \tag{7.59}$$

Note that normally, we would have assumed in the large-r region that only the terms with inverse powers of r should be included, since the potential is normally assumed to fall off as r goes to infinity. However, here we are going to impose the boundary condition (7.58) at large r, so we require

$$\phi_>(r,\theta) \longrightarrow -E_0\, r\, P_1(\cos\theta) \qquad \text{as} \quad r\longrightarrow\infty\,. \tag{7.60}$$

Notice that this asymptotic boundary condition drives the entire solution, in the sense that if E_0 were zero (no asymptotic electric field), the solution everywhere would simply be $\phi(r,\theta) = 0$. Because we are solving linear equations, this means that only the terms with $\ell = 1$ in the expansions (7.59) will be non-zero. So we can in fact just replace the general expansions in (7.59) by

$$\begin{aligned} r < a: \qquad & \phi_<(r,\theta) = A\, r\, \cos\theta\,, \\ r > a: \qquad & \phi_>(r,\theta) = \Big(-E_0\, r + B\, r^{-2}\Big)\, \cos\theta\,, \end{aligned} \tag{7.61}$$

where the two constants A and B are to be determined by the boundary conditions at $r = a$.[47] Note that we have already solved the asymptotic boundary condition (7.60) by taking $\widetilde{A}_1 = -E_0$.

[47] There is, of course, nothing wrong with retaining all the terms in the full expansions (7.59). Eventually, after imposing the boundary conditions at $r = a$ and $r = \infty$, one would discover that the solutions would give $A_\ell = B_\ell = \widetilde{A}_\ell = 0$ for all $\ell \neq 1$. But there is no point in making extra work by including all these terms, since one can see on general grounds that they must be zero. It is worth remembering also that the uniqueness theorem assures us that once the boundary conditions are specified the solution must be unique, and so if we can show that we have satisfied all the boundary conditions just with the subset of $\ell = 1$ terms, then the solution we thereby obtain must be the one and only correct solution.

The boundary conditions are given by eqns (7.39):

$$\begin{aligned} -\frac{1}{a}\,\frac{\partial\phi_<(a,\theta)}{\partial\theta} &= -\frac{1}{a}\,\frac{\partial\phi_>(a,\theta)}{\partial\theta}\,, \\ -\epsilon\,\frac{\partial\phi_<(r,\theta)}{\partial r}\Big|_{r=a} &= -\frac{\partial\phi_>(r,\theta)}{\partial r}\Big|_{r=a}\,. \end{aligned} \tag{7.62}$$

(The first equation comes from matching the tangential components of $\vec{E}$ across the boundary, and the second from matching the normal component of $\vec{D}$.) It is worth remarking at this point that the matching of the tangential derivatives of ϕ in the first equation in (7.62) can instead be expressed, essentially equivalently but more simply, as a matching of the potential itself across the boundary:

$$\phi_<(a,\theta) = \phi_>(a,\theta)\,. \tag{7.63}$$

This is because integrating the first equation in (7.62) with respect to θ gives $\phi_<(a,\theta) = \phi_>(a,\theta) + c$ where c is a constant. If the constant c were non-zero then this would mean that there would be discontinuity in $\phi(r,\theta)$ as one passed from $r < a$ to $r > a$, and this would imply a delta-function "shell" of infinite radial electric field at $r = a$. Since this obviously does not occur, it must mean that $c = 0$, and hence the first equation in (7.62) is exactly equivalent to (7.63). This observation applies rather generally to all electrostatics problems of this sort; one can always re-express the junction condition stating the continuity of the tangential components of the electric field as the simpler condition stating that the electrostatic potential must be continuous across the junction.[48]

In the simplified form, the boundary conditions at $r = a$ are

$$\phi_<(a,\theta) = \phi_>(a,\theta)\,, \qquad \epsilon\,\frac{\partial\phi_<(r,\theta)}{\partial r}\Big|_{r=a} = \frac{\partial\phi_>(r,\theta)}{\partial r}\Big|_{r=a}\,. \tag{7.64}$$

With $\phi_<(r,\theta)$ and $\phi_>(r,\theta)$ given by eqns (7.61), we therefore obtain $A = -E_0 + Ba^{-3}$ and $\epsilon\,A = -E_0 - 2Ba^{-3}$, with the solution

$$A = -\left(\frac{3}{\epsilon+2}\right)\,E_0\,, \qquad B = \left(\frac{\epsilon-1}{\epsilon+2}\right)\,a^3\,E_0\,. \tag{7.65}$$

[48] Another example where the junction condition $\vec{n}\times(\vec{E}_2 - \vec{E}_1) = 0$ can be expressed more simply as the condition of continuity of the potential, $(\phi_2 - \phi_1) = 0$, is in the previous example discussed in section 7.3.1. As can easily be checked, the second equation in (7.48), which came from the boundary conditions on the x and y components (i.e. the tangential components) of the electric field (second and third equations in eqns (7.41)), is precisely the same as the equation one gets by equating the two expressions $\phi_>$ (in (7.44)) and $\phi_<$ (in (7.46)) for the potentials at the $z = 0$ boundary.

Thus the potentials inside and outside the sphere are given by

$$\begin{aligned}\phi_<(r,\theta) &= -\left(\frac{3}{\epsilon+2}\right) E_0\, r\, \cos\theta = -\left(\frac{3}{\epsilon+2}\right) E_0\, z\,,\\ \phi_>(r,\theta) &= -E_0\, r\, \cos\theta + \left(\frac{\epsilon-1}{\epsilon+2}\right) E_0\, \frac{a^3}{r^2}\, \cos\theta\,. \end{aligned} \tag{7.66}$$

The potential inside the sphere implies that the electric field is uniform for $r < a$, and parallel to the external applied field (so it lies along the z direction). Its magnitude is given by

$$E_< = \frac{3}{\epsilon+2}\, E_0 \qquad \text{and so} \quad \vec{E}_< = \frac{3}{\epsilon+2}\, \vec{E}_0\,. \tag{7.67}$$

The potential outside the sphere is a sum of two terms. The first just gives the original uniform electric field E_0 (see (7.58)). The second term in the expression for $\phi_>$ in (7.66) can be written as

$$\left(\frac{\epsilon-1}{\epsilon+2}\right) a^3\, E_0\, \frac{\vec{z}\cdot\vec{r}}{r^3}\,, \tag{7.68}$$

where $\vec{z} = (0,0,1)$ is the unit vector along the z axis. This term can be recognised (see (6.32) as the potential due to a dipole of moment

$$\vec{p} = \left(\frac{\epsilon-1}{\epsilon+2}\right) a^3\, \vec{E}_0\,, \tag{7.69}$$

pointing along the direction of the applied electric field (i.e. along the z axis).

Inside the sphere, the polarisation $\vec{P}$ is given by

$$\vec{P} = \chi\, \vec{E} = \left(\frac{\epsilon-1}{4\pi}\right) \vec{E} = \frac{3}{4\pi}\, \left(\frac{\epsilon-1}{\epsilon+2}\right) \vec{E}_0\,. \tag{7.70}$$

Since this is constant inside the sphere, we see that its volume integral is given by

$$\int_{\text{sphere}} \vec{P} = \tfrac{4}{3}\,\pi a^3\, \vec{P} = \left(\frac{\epsilon-1}{\epsilon+2}\right) a^3\, \vec{E}_0\,. \tag{7.71}$$

Comparing with (7.69), we see that the dipole moment $\vec{p}$ is nothing but the volume integral of the polarisation $\vec{P}$ inside the sphere.

From (7.53), it can be seen that the polarisation surface-charge density on the surface of the sphere is given by $\sigma_{\text{pol}} = (\vec{n}\cdot\vec{P})|_{r=a}$ with $\vec{n} = \vec{r}/r$, and so so

$$\sigma_{\text{pol}} = \frac{3}{4\pi}\, \left(\frac{\epsilon-1}{\epsilon+2}\right) E_0\, \cos\theta\,. \tag{7.72}$$

This can be understood as generating an internal electric field that tends to oppose the applied field, thus reducing the electric field inside the sphere to the value given by (7.67).

7.4 Electrostatic energy in dielectric media

In free space, we saw that the electrostatic self energy of a charge distribution ρ was given by

$$U = \tfrac{1}{2} \int \rho(\vec{r})\phi(\vec{r}) d^3\vec{r}. \tag{7.73}$$

Using $\vec{\nabla} \cdot \vec{E} = 4\pi\rho$ and $\vec{E} = -\vec{\nabla}\phi$, this could be rewritten as

$$U = \frac{1}{8\pi} \int |\vec{E}|^2 \, d^3\vec{r}. \tag{7.74}$$

We may derive the analogous expression in a dielectric medium as follows. Suppose that a localised charge distribution ρ gives rise to the potential ϕ, and that we then make an infinitesimal change $\delta\rho$ to the charge distribution. The work done in making this change will be given by

$$\delta U = \int \delta\rho \, \phi \, d^3\vec{r}. \tag{7.75}$$

Using $\vec{\nabla} \cdot \vec{D} = 4\pi\rho$, we have $\delta\rho = (\vec{\nabla} \cdot \delta\vec{D})/(4\pi)$, and so

$$\delta U = \frac{1}{4\pi} \int \phi(\vec{\nabla} \cdot \delta\vec{D}) \, d^3\vec{r} = -\frac{1}{4\pi} \int (\vec{\nabla}\phi) \cdot \delta\vec{D} \, d^3\vec{r}, \tag{7.76}$$

where we have dropped the boundary term in the integration by parts, on account of the fact that the charge distribution is taken to be localised in some finite region. Using $\vec{E} = -\vec{\nabla}\phi$, we therefore have

$$\delta U = \frac{1}{4\pi} \int \vec{E} \cdot \delta\vec{D} \, d^3\vec{r}. \tag{7.77}$$

The energy of the charge distribution can thus be calculated by integrating up from $\vec{D} = 0$ to its final value. In general, this integration could be non-trivial, since $\vec{D}$ and $\vec{E}$ could be related in a complicated way.

Suppose, however, that the medium is *linear*, in the sense that $\vec{D}$ is *proportional* to $\vec{E}$. This is true, for example, if $\vec{D} = \epsilon\vec{E}$. For such a linear medium, we must have that

$$\vec{E} \cdot \delta\vec{D} = \vec{E} \cdot (\epsilon \, \delta\vec{E}) = \epsilon \, \vec{E} \cdot \delta\vec{E} = \vec{D} \cdot \delta\vec{E} \,, \tag{7.78}$$

and hence we can write

$$\vec{E} \cdot \delta\vec{D} = \tfrac{1}{2}\vec{E} \cdot \delta\vec{D} + \tfrac{1}{2}\vec{E} \cdot \delta\vec{D} = \tfrac{1}{2}\vec{E} \cdot \delta\vec{D} + \tfrac{1}{2}\vec{D} \cdot \delta\vec{E} = \tfrac{1}{2}\delta(\vec{E} \cdot \vec{D}) \,, \tag{7.79}$$

and so from eqn (7.77) we have

$$\delta U = \frac{1}{8\pi} \int \delta(\vec{E} \cdot \vec{D}) \, d^3\vec{r} = \delta\Big(\frac{1}{8\pi} \int \vec{E} \cdot \vec{D} \, d^3\vec{r}\Big) \,. \tag{7.80}$$

It is then straightforward to integrate up from $\vec{D} = 0$ to its final value, giving

$$\boxed{U = \frac{1}{8\pi} \int \vec{E} \cdot \vec{D}\, d^3\vec{r}.} \tag{7.81}$$

Interestingly, if we now use $\vec{E} = -\vec{\nabla}\phi$ in (7.81), integrate by parts, and then use $\vec{\nabla} \cdot \vec{D} = 4\pi\rho$, we obtain

$$U = \tfrac{1}{2} \int \rho(\vec{r})\, \phi(\vec{r}) d^3\vec{r}, \tag{7.82}$$

which is the same as the free-space result. However, it should be emphasised that in deriving (7.81) it was necessary to use the assumption of *linearity* of the relation between $\vec{D}$ and $\vec{E}$, and so although (7.82) holds in dielectric media with a linear response, it does not hold if the relation between $\vec{D}$ and $\vec{E}$ is non-linear.

Suppose that a distribution of charge $\rho_0(x)$ inside a medium of dielectric constant ϵ_0 gives rise to an electric field $\vec{E}_0$. We may allow ϵ_0 to be position dependent in the medium.[49] We know from the discussion above that this configuration will have electrostatic energy U_0 given by

$$U_0 = \frac{1}{8\pi} \int \vec{E}_0 \cdot \vec{D}_0\, d^3\vec{r}, \tag{7.83}$$

with $\vec{D}_0 = \epsilon_0 \vec{E}_0$. The medium in this discussion might be just a vacuum, in which case $\epsilon_0 = 1$, or it might be a fluid (liquid or gas) with $\epsilon_0 > 1$. In any case, for the next step in the discussion, we need to suppose that it is possible to introduce a piece of solid material with a different dielectric constant into the same space.

While keeping the charge distribution fixed, suppose now that a piece of dielectric material of volume V and dielectric constant ϵ_1 is introduced into the field. This has the effect of making the dielectric constant ϵ equal to ϵ_1 inside V, while it remains at its original value ϵ_0 outside V. We can think of $\epsilon = \epsilon(\vec{r})$, with $\epsilon(\vec{r})$ changing smoothly but rapidly from ϵ_1 inside V to ϵ_0 outside V. The electric field is now $\vec{E}$. The energy is given by

$$U = \frac{1}{8\pi} \int \vec{E} \cdot \vec{D}\, d^3\vec{r}, \tag{7.84}$$

where $\vec{D}(\vec{r}) = \epsilon(\vec{r})\, \vec{E}(\vec{r})$.

The change in energy, $\Delta U = U - U_0$, is given by

$$\Delta U = \frac{1}{8\pi} \int (\vec{E} \cdot \vec{D} - \vec{E}_0 \cdot \vec{D}_0) d^3\vec{r}. \tag{7.85}$$

[49] Do not confuse ϵ_0 here with the symbol ϵ_0 that is used in SI units and is known as the "permittivity of free space." Here, ϵ_0 is just being used to denote whatever permittivity the medium happens to have.

Clearly this can be written as

$$\Delta U = \frac{1}{8\pi}\int (\vec{E}\cdot\vec{D}_0 - \vec{E}_0\cdot\vec{D})d^3\vec{r} + \frac{1}{8\pi}\int (\vec{E}+\vec{E}_0)\cdot(\vec{D}-\vec{D}_0)d^3\vec{r}. \qquad (7.86)$$

Since $\vec{\nabla}\times\vec{E}=0$ and $\vec{\nabla}\times\vec{E}_0=0$, it follows that $\vec{\nabla}\times(\vec{E}+\vec{E}_0)=0$ and so we can write $\vec{E}+\vec{E}_0=\vec{\nabla}\psi$ for some scalar function ψ. The second integral in (7.86) therefore gives

$$\frac{1}{8\pi}\int \vec{\nabla}\psi\cdot(\vec{D}-\vec{D}_0)d^3\vec{r} = -\frac{1}{8\pi}\int \psi\vec{\nabla}\cdot(\vec{D}-\vec{D}_0)d^3\vec{r}. \qquad (7.87)$$

Since we assumed that the charge distribution ρ_0 was unaltered by the introduction of the dielectric medium ϵ_1, it follows that $\vec{\nabla}\cdot\vec{D}=\vec{\nabla}\cdot\vec{D}_0=4\pi\rho_0$, and hence the integral (7.87) gives zero. Thus we have

$$\Delta U = \frac{1}{8\pi}\int (\vec{E}\cdot\vec{D}_0 - \vec{E}_0\cdot\vec{D})d^3\vec{r}. \qquad (7.88)$$

Since $\vec{D}=\epsilon_0\vec{E}$ at points outside V, it follows that the integrand in (7.88) vanishes at points outside the volume V, and so the integration in (7.88) need be performed only within V, giving

$$\Delta U = -\frac{1}{8\pi}\int_V (\epsilon_1-\epsilon_0)\vec{E}\cdot\vec{E}_0\, d^3\vec{r}. \qquad (7.89)$$

If the original medium was in fact just free space, so that $\epsilon_0=1$, then it follows from $\vec{D}=\vec{E}+4\pi\vec{P}$, and $\vec{D}=\epsilon\vec{E}$, that (7.89) becomes

$$\Delta U = -\tfrac{1}{2}\int_V \vec{P}\cdot\vec{E}_0\, d^3\vec{r}, \qquad (7.90)$$

where $\vec{P}$ is the polarisation of the dielectric material. From this we see that if a dielectric object has polarisation $\vec{P}$ when placed in an electric field $\vec{E}_0$ that is generated by fixed sources, then it will have an energy density

$$w = -\tfrac{1}{2}\vec{P}\cdot\vec{E}_0\,. \qquad (7.91)$$

This expression is analogous to the contribution in (6.65) giving the energy of a dipole $\vec{p}$ in an electric field. The reason for the factor of $\frac{1}{2}$ in (7.91) is that, rather than being like a *fixed* dipole moment, the polarisation $\vec{P}$ in (7.91) is itself *caused* by the presence of the electric field $\vec{E}_0$.

The implication of the above calculation is that a dielectric object (with $\epsilon>1$) will tend to be drawn into a region of increasing electric field $\vec{E}_0$. The force can be calculated by making a virtual infinitesimal displacement $\delta\vec{r}$ of the object, leading to a change δU in the energy. Since the charges are held fixed, there is no change in the external energy, and so

the change in the field energy can be interpreted as a change in the potential energy of the object. There will therefore be a force acting on the object, given by

$$\vec{F} = -\vec{\nabla} U \,. \tag{7.92}$$

It is important to emphasise that this is *when we hold the charges that generate the electric field fixed.*

An alternative situation, and in fact one that in practice is more likely to arise, is that the electric field is generated by means of electrodes or plates that are held at a *fixed potential*, rather than with *fixed charge.* This would arise if the electrodes are attached to a battery or some other constant-voltage source. The situation is then like the classic problem in elementary electrostatics, in which one calculates the atttractive force between the plates of a parallel plate capacitor. The easier calculation is when one assumes that the charge on the plates is fixed (achieved by connecting the plates to the battery to charge the capacitor and then disconnecting the battery). The force is calculated by working out the energy change when a small virtual displacement is made. A more complicated way of reaching the same conclusion is to keep the battery connected during the virtual displacement, so that now the potential, instead of the charge, is held fixed. Naively, one now arrives at the (incorrect) conclusion that the potential energy is *increased* if the plates are moved together, implying a repulsive force that is equal in magnitude to the previously-calculated attractive force. Of course, what the constant-potential argument has neglected is that the battery has to do work *in order to keep the potential fixed* during the virtual displacement. In fact it does twice as much work as the field energy-change, and so when the changes in field energy *plus* work done by the battery are added, the net energy change is the same in magnitdude *and in sign* as the net energy change in the constant-charge virtual displacement. Thus the force between the capacitor plates comes out to be the same in magnitude and sign, regardless of whether the potential difference or the charge is held fixed.

Of course it is clear from common sense that this must be the case. A fixed capacitor (assumed to be ideal, with no leakage) holds the same potential difference whether the battery is connected or not, and so the force cannot possibly suddenly switch in sign depending on whether the battery is connected or not.

In the same way, one can do the analogous constant-potential calculation for a dielectric object being drawn into an electric field. The final conclusion is that it is drawn in, with the same force, whether the battery is connected or not.

8 Magnetostatics

We now turn to the magnetic analogue of electrostatics, namely magnetostatics. This describes the situation when everything is independent of time, and only magnetic fields and electric currents are present.

8.1 Ampère's law and the Biot-Savat law

To begin, we shall consider the situation where no magnetically permeable media are present. From the general form of Maxwell's equations (1.3), we see that the only non-trivial equations to be considered are

$$\vec{\nabla}\cdot\vec{B}=0\,,\qquad \vec{\nabla}\times\vec{B}=\frac{4\pi}{c}\,\vec{J}\,. \tag{8.1}$$

(Recall that we are using Gaussian units, and that c is the speed of light.)

In view of $\vec{\nabla}\cdot\vec{B}=0$, we can write $\vec{B}$ in terms of a vector potential $\vec{A}$, as

$$\vec{B}=\vec{\nabla}\times\vec{A}\,. \tag{8.2}$$

The choice of vector potential is not unique, and we can perform a *gauge transformation* of the form

$$\vec{A}\longrightarrow\vec{A}+\vec{\nabla}\lambda\,, \tag{8.3}$$

where λ is any function of $\vec{r}$, since obviously when (8.3) is substituted into (8.2), the term involving λ gives zero.

We may employ the *gauge invariance* of the system under the transformations (8.3) in order to impose a convenient *gauge condition* called the *Coulomb gauge*

$$\vec{\nabla}\cdot\vec{A}=0\,. \tag{8.4}$$

This can be seen by supposing that we started with a gauge potential $\widetilde{\vec{A}}$ that was not in the Coulomb gauge, and transforming it according to (8.3) to give

$$\vec{A}=\widetilde{\vec{A}}+\vec{\nabla}\lambda\,, \tag{8.5}$$

where we require that $\vec{A}$ *is* in Coulomb gauge; i.e. $\vec{\nabla}\cdot\vec{A}=0$. Taking the divergence of (8.5), we therefore see that to achieve this, λ must satisfy

$$\nabla^2\lambda=-\vec{\nabla}\cdot\widetilde{\vec{A}}\,. \tag{8.6}$$

This Poisson equation can always be solved; in fact, we are familiar with its solution from solving such an equation in electrostatics (i.e. $\nabla^2\phi = -4\pi\rho$). Thus we can immediately write down the solution for the required gauge transformation function λ:

$$\lambda(\vec{r}) = \frac{1}{4\pi}\int \frac{\vec{\nabla}'\cdot\widetilde{\vec{A}}(\vec{r}\,')}{|\vec{r}-\vec{r}\,'|}\, d^3\vec{r}\,' \,. \tag{8.7}$$

Having shown that it must always be possible to perform a gauge transformation to put the vector potential into the Coulomb gauge, we can now proceed under the assumption that this has been done, and that $\vec{A}$ is indeed in Coulomb gauge. Substituting into the second equation in (8.1), we find

$$\begin{aligned} \frac{4\pi}{c}\,\vec{J} &= \vec{\nabla}\times\vec{B} = \vec{\nabla}\times(\vec{\nabla}\times\vec{A}) = \vec{\nabla}(\vec{\nabla}\cdot\vec{A}) - \nabla^2\vec{A}\,, \\ &= -\nabla^2\vec{A}\,, \end{aligned} \tag{8.8}$$

where we are assuming here that Cartesian vectors are used, with ∇^2 meaning the standard scalar Laplacian, $\nabla^2 = \partial_i\partial_i$, and so $\nabla^2\vec{A}$ means the Cartesian vector with components $(\nabla^2 A_x, \nabla^2 A_y, \nabla^2 A_z)$. The equation can be solved immediately; it is just a vector-valued version of the equation $\nabla^2\phi = -4\pi\rho$ from electrostatics. Thus we have

$$\vec{A}(\vec{r}) = \frac{1}{c}\int \frac{\vec{J}(\vec{r}\,')}{|\vec{r}-\vec{r}\,'|}\, d^3\vec{r}\,' \,. \tag{8.9}$$

Again, we emphasise that this result is valid *only* if we use Cartesian vectors. For example if we use spherical polar coordinates, with respect to which the vectors $\vec{A}$ and $\vec{J}$ have components A_r, A_θ, A_φ and J_r, J_θ and J_φ, then it is *not* true that the spherical-polar components of $\vec{A}$ are given by inserting the spherical polar components of $\vec{J}$ into the right-hand side of eqn (8.9). We shall see this explicitly a bit later on.

Rerturning now to Cartesian coordinates, we may easily calculate the magnetic field $\vec{B}$ from (8.9), by taking the curl. It is useful to note that for any vector $\vec{V}$ and any scalar f, we have the identity

$$\begin{aligned} \vec{\nabla}\times(f\,\vec{V}) &= (\vec{\nabla}f)\times\vec{V} + f\,\vec{\nabla}\times\vec{V}\,, \\ &= -\vec{V}\times\vec{\nabla}f + f\,\vec{\nabla}\times\vec{V}\,. \end{aligned} \tag{8.10}$$

Applying this with $f = |\vec{r}-\vec{r}\,'|^{-1}$ and $V = \vec{J}(\vec{r}\,')$ (and bearing in mind that $\vec{\nabla}$ acts only on functions of $\vec{r}$, and not on functions of $\vec{r}\,'$), we see from (8.9) that

$$\vec{B}(\vec{r}) = -\frac{1}{c}\int \vec{J}(\vec{r}\,')\times\vec{\nabla}\Big(\frac{1}{|\vec{r}-\vec{r}\,'|}\Big)\, d^3\vec{r}\,' \,, \tag{8.11}$$

and hence

$$\vec{B}(\vec{r}) = \frac{1}{c}\int \vec{J}(\vec{r}\,') \times \frac{(\vec{r}-\vec{r}\,')}{|\vec{r}-\vec{r}\,'|^3}\, d^3\vec{r}\,' \,, \tag{8.12}$$

Suppose we consider a charge q moving along the path $\vec{r} = \vec{r}_0(t)$. It will give rise to the current density

$$\vec{J}(\vec{r}) = q\,\delta^3\big(\vec{r}-\vec{r}_0(t)\big)\,\frac{d\vec{r}_0(t)}{dt}\,. \tag{8.13}$$

Substituting this into (8.12), we therefore have

$$\begin{aligned}\vec{B}(\vec{r}) &= \frac{q}{c}\int \delta^3\big(\vec{r}\,'-\vec{r}_0(t)\big)\,\frac{d\vec{r}_0(t)}{dt} \times \frac{(\vec{r}-\vec{r}\,')}{|\vec{r}-\vec{r}\,'|^3}\, d^3\vec{r}\,' \,,\\ &= \frac{q}{c}\,\frac{d\vec{r}_0(t)}{dt} \times \frac{\big(\vec{r}-\vec{r}_0(t)\big)}{|\vec{r}-\vec{r}_0(t)|^3}\,.\end{aligned} \tag{8.14}$$

Suppose at some instant $t = t_0$, the particle is at the origin, so $\vec{r}_0(t_0) = 0$. Writing its velocity as $\vec{v} = d\vec{r}_0(t)/dt$, we therefore have

$$\vec{B} = \frac{q\vec{v}\times\vec{r}}{cr^3}\,. \tag{8.15}$$

This result assumes that the velocity $\vec{v}$ is small compared with the speed of light, so that relativistic effects can be neglected. In fact, we were really making a kind of "quasi-static" approximation when we derived the result above, in which we assume that the magnetic fields are sufficiently slowly varying that to a good approximation we can neglect the time derivative terms in the full Maxwell equations (1.3), and thus we can assume

$$\vec{\nabla}\times\vec{B}(\vec{r},t) = \frac{4\pi}{c}\,\vec{J}(\vec{r},t)\,, \qquad \vec{\nabla}\cdot\vec{B}(\vec{r},t) = 0\,. \tag{8.16}$$

One can think of the charge moving along an infinitesimal segment of its path as being equivalent to a current I passing through the corresponding infinitesimal line element $d\vec{\ell}$, in the sense that $q\vec{v} \longrightarrow I d\vec{\ell}$. Thus, if the current element is located at the origin, then it gives a contribution

$$d\vec{B}(\vec{r}) = \frac{I\,(d\vec{\ell}\times\vec{r}\,)}{c\,r^3} \tag{8.17}$$

to the magnetic field. This result is known as the Biot-Savat law. Note that it only really makes sense in the context of an integral around a current loop or circuit, since current cannot simply materialise from nowhere and then disappear again. (It would violate charge conservation.) Historically, of course, the Biot-Savat law came first, and (8.12) was deduced from it.

Suppose that a current element $I_1\,d\vec{r}_1$ is placed in a magnetic field $\vec{B}$. Experiments by Ampère in the 19th century established that it would experience a force $d\vec{F}$ given by

$$d\vec{F} = \frac{I_1}{c}\,d\vec{r}_1\times\vec{B}\,. \tag{8.18}$$

If the magnetic field is itself due to an infinitesimal current element $I_2\, d\vec{r}_2$, then from (8.17), the (doubly infinitesimal) force $d\vec{F}_{12}$ experienced by the first current element will be given by

$$d\vec{F}_{12} = \frac{I_1\, I_2}{c^2}\, \frac{d\vec{r}_1 \times (d\vec{r}_2 \times \vec{r}_{12})}{r_{12}^3}\,, \tag{8.19}$$

where $\vec{r}_{12}$ is the vector from $d\vec{r}_2$ to $d\vec{r}_1$. In other words, $\vec{r}_{12} = \vec{r}_2 - \vec{r}_1$.

The expression (8.19) can be integrated up around the two current loops, to give

$$\vec{F}_{12} = \frac{I_1\, I_2}{c^2} \oint\oint \frac{d\vec{r}_1 \times (d\vec{r}_2 \times \vec{r}_{12})}{r_{12}^3}\,, \tag{8.20}$$

As it stands, this expression is not manifestly (anti)symmetric under the exchange of the rôles of the two current loops, but of course it should be, since by Newton's third law the force on loop 1 due to the current in loop 2 should be equal and opposite to the force on loop 2 due to the the current in loop 1. However, this can be made manifest as follows:

Using the standard identity for the vector triple product, we can write

$$d\vec{r}_1 \times (d\vec{r}_2 \times \vec{r}_{12}) = d\vec{r}_2\, (d\vec{r}_1 \cdot \vec{r}_{12}) - (d\vec{r}_1 \cdot d\vec{r}_2)\, \vec{r}_{12}\,, \tag{8.21}$$

and so (8.19) can be written as

$$d\vec{F}_{12} = \frac{I_1\, I_2}{c^2}\, d\vec{r}_2 \Big(\frac{d\vec{r}_1 \cdot \vec{r}_{12}}{r_{12}^3}\Big) - \frac{I_1\, I_2}{c^2}\, (d\vec{r}_1 \cdot d\vec{r}_2)\, \frac{\vec{r}_{12}}{r_{12}^3}\,. \tag{8.22}$$

The second term here *is* antisymmetric under exchanging the labels 1 and 2. We just need to work on the first term, therefore.

If we consider just the Loop 1 integration for now, keeping $\vec{r}_2$ fixed, the first term in (8.22) can be written as the exact differential

$$\frac{I_1\, I_2}{c^2}\, d\vec{r}_2\, d\Big(\frac{1}{r_{12}}\Big)\,. \tag{8.23}$$

This follows from the fact that $r_{12}^2 = (\vec{r}_2 - \vec{r}_1) \cdot (\vec{r}_2 - \vec{r}_1)$, and so (with $\vec{r}_2$ held fixed),

$$2r_{12}\, dr_{12} = -2d\vec{r}_1 \cdot (\vec{r}_2 - \vec{r}_1) = -2d\vec{r}_1 \cdot \vec{r}_{12}\,. \tag{8.24}$$

Thus if we first integrate (8.22) around Loop 1, the first term gives zero (since $\oint df = 0$ whenever any exact differential df is integrated around a *closed* loop). Integrating up the remaining second term in eqn (8.22), we find that the total force $\vec{F}_{12}$ is given by

$$\vec{F}_{12} = -\frac{I_1\, I_2}{c^2} \oint\oint \frac{(d\vec{r}_1 \cdot d\vec{r}_2)\, \vec{r}_{12}}{r_{12}^3} = -\vec{F}_{21}\,. \tag{8.25}$$

This makes manifest the total antisymmetry between the rôles of the two loops.

The expression (8.18) for the force on a current element $I\,d\vec{\ell}$ in a magnetic field $\vec{B}$ can be generalised immediately to the situation where there is a current density $\vec{J}$ in an external $\vec{B}$ field. The infinitesimal force on the current density in the volume element $d^3\vec{r}$ will be given by

$$d\vec{F}(\vec{r}) = \frac{1}{c}\,\vec{J}(\vec{r}) \times \vec{B}(\vec{r})\,d^3\vec{r}, \tag{8.26}$$

and so the total force on the current distribution will be given by

$$\vec{F} = \frac{1}{c}\int \vec{J}(\vec{r}) \wedge \vec{B}(\vec{r})\,d^3\vec{r}. \tag{8.27}$$

It also follows from (8.26) that the infinitesimal torque on the element $d^3\vec{r}$ will be

$$d\vec{N}(\vec{r}) = \vec{r} \times d\vec{F}(\vec{r}) = \frac{1}{c}\,\vec{r} \times \Big(\vec{J}(\vec{r}) \times \vec{B}(\vec{r})\Big)\,d^3\vec{r}, \tag{8.28}$$

and so the total torque (measured relative to the origin) is

$$\vec{N} = \frac{1}{c}\int \vec{r} \times \Big(\vec{J}(\vec{r}) \times \vec{B}(\vec{r})\Big)\,d^3\vec{r}. \tag{8.29}$$

A further application of the expression (8.18) for the force on a current element in an external magnetic field is to the situation where a particle of charge q is moving with velocity $\vec{v}$ in the field $\vec{B}$. It then follows that it will experience a force given by

$$\vec{F} = \frac{q}{c}\,\vec{v} \times \vec{B}\,. \tag{8.30}$$

This is known as the *Lorentz force*. (We are assuming here that the velocity $\vec{v}$ is small in comparison to the speed of light.)

Finally, in this section, we note that the Maxwell equation $\vec{\nabla} \times \vec{B} = (4\pi/c)\,\vec{J}$ can be integrated to give the result known as *Ampère's Law*. Thus, integrating over an open surface Σ, with closed 1-dimensional boundary C, we have, by using Stokes' theorem

$$\int_\Sigma (\vec{\nabla} \times \vec{B}) \cdot d\vec{S} = \oint_C \vec{B} \cdot d\vec{\ell}, \tag{8.31}$$

the result that

$$\oint_C \vec{B} \cdot d\vec{\ell} = \frac{4\pi}{c}\int_\Sigma \vec{J} \cdot d\vec{S}\,. \tag{8.32}$$

The integral $\vec{J} \cdot d\vec{S}$ on the right-hand side is equal to the total current I passing through the area bounded by the loop C, and so we obtain Ampère's law

$$\oint_C \vec{B} \cdot d\vec{\ell} = \frac{4\pi}{c}\,I\,. \tag{8.33}$$

(Of course historically, Ampère's law was discovered (empirically) first, and only later was it re-expressed in the differential form $\vec{\nabla} \times \vec{B} = (4\pi/c)\,\vec{J}$.)

8.2 Magnetic field of a circular current loop

A relatively simple, and highly symmetrical, illustration of some of the methods of the previous section is provided by considering a circular conducting loop, of radius a, around which a current I is passing. We may take the loop to be centred on the origin, lying in the (x, y) plane. Using spherical polar coordinates, it is described by a current density $\vec{J}$ whose spherical-polar components J_r and J_θ vanish, whilst

$$J_\varphi(r', \theta', \varphi') = \frac{I}{a}\,\delta(\cos\theta')\,\delta(r'-a) = \frac{I}{a}\,\delta(\theta' - \tfrac{1}{2}\pi)\,\delta(r'-a)\,. \tag{8.34}$$

(Note that using the standard properties of the delta function, $\delta(\cos\theta')$ is the same thing as $\delta(\theta' - \frac{1}{2}\pi)$ when θ' lies in the interval $0 \le \theta' \le \pi$.)

It is actually helpful to re-express the current density in terms of its Cartesian coordinate components, the reason being that only in this case can we employ equation (8.9) giving the vector potential $\vec{A}$ as an integral involving $\vec{J}$. Thus we shall have

$$J_x = -J_\varphi\,\sin\varphi'\,, \qquad J_y = J_\varphi\,\cos\varphi'\,, \qquad J_z = 0\,. \tag{8.35}$$

Clearly, the system is azimuthally symmetric, and so we can, without loss of generality, take the observation point to be at $\varphi = 0$ in order to simplify the calculation. In other words, we take the observation point to lie in the (x, z) plane. This means that only the y component of $\vec{A}$ will be non-vanishing. Note that just like the expressions in (8.35) for the current density, we shall also have the relations

$$A_x = -A_\varphi\,\sin\varphi\,, \qquad A_y = A_\varphi\,\cos\varphi\,, \qquad A_z = 0\,, \tag{8.36}$$

and so at $\varphi = 0$ we shall have $A_y = A_\varphi$. Thus, from (8.9) and (8.34) we shall have[50]

$$A_\varphi(r,\theta) = A_y\Big|_{\varphi=0} = \frac{I}{ac}\int dr' d\Omega' \frac{r'^2\,\cos\varphi'\,\delta(\theta' - \frac{1}{2}\pi)\,\delta(r'-a)}{|\vec{r} - \vec{r}\,'|}\Big|_{\varphi=0}\,, \tag{8.37}$$

since $d^3\vec{r}\,' = r'^2\,d\Omega'$, where $d\Omega' = \sin\theta'\,d\theta' d\varphi'$. Since we are taking $\varphi = 0$ we have

$$\vec{r} = (r\sin\theta, 0, r\cos\theta)\,, \qquad \vec{r}\,' = (r'\,\sin\theta'\cos\varphi', r'\,\sin\theta'\sin\varphi', r'\,\cos\theta')\,, \tag{8.38}$$

implying $|\vec{r} - \vec{r}\,'|^2 = r^2 + r'^2 - 2rr'\,(\cos\theta\cos\theta' + \sin\theta\sin\theta'\cos\varphi')$, and so, after performing the integrations over the delta functions, we have

$$A_\varphi(r,\theta) = \frac{Ia}{c}\int_0^{2\pi} \frac{\cos\varphi'\,d\varphi'}{(a^2 + r^2 - 2ar\sin\theta\cos\varphi')^{1/2}}\,. \tag{8.39}$$

[50] Note that if one were to make the mistake of thinking that one could use eqn (8.9) directly for the spherical polar components of $\vec{A}$ and $\vec{J}$, and thus writing the (incorrect) equation $A_\varphi(r,\theta,\varphi) = c^{-1}\int J_\varphi(r',\theta',\varphi')\,|\vec{r} - \vec{r}\,'|^{-1}\,d^3\vec{r}\,'$, one would get the wrong answer; the factor of $\cos\varphi'$ in the integrand of eqn (8.37) would be missing.

The integral in (8.39) can actually be performed explicitly, as we shall see later, although the result involves the complete elliptic integrals $K(k)$ and $E(k)$. First, let's just look at the leading-order approximation to (8.39), if we assume that the dimensionless quantity

$$\frac{2ar\sin\theta}{r^2+a^2} \tag{8.40}$$

is much less than 1. Notice that this condition will be satisfied, in particular, if either $r >> a$ or if $r << a$. We may then expand the integrand in (8.39) as

$$\begin{aligned}
[r^2+a^2-2ar\sin\theta\,\cos\varphi']^{-1/2}\,\cos\varphi' &= \frac{\cos\varphi'}{\sqrt{r^2+a^2}}\Big[1-\frac{2ar\sin\theta\,\cos\varphi'}{r^2+a^2}\Big]^{-1/2} \\
&= \frac{\cos\varphi'}{\sqrt{r^2+a^2}}\Big[1+\frac{ar\sin\theta}{r^2+a^2}\,\cos\varphi'+\cdots\Big] \\
&= \frac{\cos\varphi'}{\sqrt{r^2+a^2}}+\frac{ar\sin\theta\cos^2\varphi'}{(r^2+a^2)^{3/2}}+\cdots\,. \qquad (8.41)
\end{aligned}$$

Integrating this term by term, and noting that

$$\int_0^{2\pi}\cos\varphi'\,d\varphi'=0\,,\qquad \int_0^{2\pi}\cos^2\varphi'\,d\varphi'=\pi\,, \tag{8.42}$$

we see from (8.39) that the leading-order contribution in this expansion for A_φ is given by the second term in (8.41), and thus

$$A_\varphi=\frac{I\pi a^2\,r\sin\theta}{c\,(r^2+a^2)^{3/2}}+\cdots\,. \tag{8.43}$$

Let us consider, in particular, the regime where $r >> a$. From (8.43) we therefore have

$$A_\varphi=\frac{I\,\pi a^2\,\sin\theta}{r^2}+\mathcal{O}\Big(\frac{1}{r^3}\Big)\,. \tag{8.44}$$

In spherical polar coordinates, the components of the magnetic field $\vec{B}=\vec{\nabla}\times\vec{A}$ are in general given by

$$\begin{aligned}
B_r &= \frac{1}{r\sin\theta}\Big[\frac{\partial(\sin\theta\,A_\varphi)}{\partial\theta}-\frac{\partial A_\theta}{\partial\varphi}\Big]\,, \\
B_\theta &= \frac{1}{r\sin\theta}\,\frac{\partial A_r}{\partial\varphi}-\frac{1}{r}\,\frac{\partial(r\,A_\varphi)}{\partial r}\,, \\
B_\varphi &= \frac{1}{r}\Big[\frac{\partial(r\,A_\theta)}{\partial r}-\frac{\partial A_r}{\partial\theta}\Big]\,. \qquad (8.45)
\end{aligned}$$

In our case, $A_r=0$ and $A_\theta=0$, and so the components of the $\vec{B}$ field are given by

$$B_r=\frac{1}{r\sin\theta}\frac{\partial(\sin\theta\,A_\varphi)}{\partial\theta}\,,\qquad B_\theta=-\frac{1}{r}\frac{\partial(rA_\varphi)}{\partial r}\,,\qquad B_\varphi=0\,. \tag{8.46}$$

Plugging in the leading-order large-r expansion (8.44), we therefore have

$$A_\varphi\approx\frac{I\pi a^2}{c}\,\frac{\sin\theta}{r^2}\,, \tag{8.47}$$

and so

$$B_r \approx \frac{I\pi a^2}{c}\,\frac{2\cos\theta}{r^3}\,, \qquad B_\theta \approx \frac{I\pi a^2}{c}\,\frac{\sin\theta}{r^3}\,, \qquad B_\varphi = 0\,. \tag{8.48}$$

It is interesting to compare the $\vec{B}$ field of a current loop with the $\vec{E}$ field due to an electric dipole. Recall from (6.32) that the potential due to an electric dipole $\vec{p}$ at the origin is given by

$$\phi = \frac{\vec{p}\cdot\vec{r}}{r^3}\,, \tag{8.49}$$

and so if it is parallel to the z axis we shall have

$$\phi(r,\theta) = \frac{p\cos\theta}{r^2}\,. \tag{8.50}$$

This implies that the spherical-polar components of $\vec{E}$ will be given by

$$\begin{aligned} E_r &= -\frac{\partial\phi}{\partial r} = \frac{2p\cos\theta}{r^3}\,, \\ E_\theta &= -\frac{1}{r}\frac{\partial\phi}{\partial\theta} = \frac{p\sin\theta}{r^3}\,, \\ E_\varphi &= -\frac{1}{r\sin\theta}\frac{\partial\phi}{\partial\varphi} = 0\,. \end{aligned} \tag{8.51}$$

Comparison with (8.48) shows that the magnetic field of a current loop, seen from afar, is dipole in character. We also see that it is natural to define the *magnetic dipole moment* m for the current loop by

$$m = \frac{I\pi a^2}{c}\,, \tag{8.52}$$

so that at the leading order in $1/r$ the vector potential is given by

$$A_\varphi = \frac{m\,\sin\theta}{r^2}\,, \qquad A_r = 0\,, \qquad A_\theta = 0\,, \tag{8.53}$$

and then the components of the magnetic field at leading order ar given by

$$\begin{aligned} B_r &= \frac{2m\cos\theta}{r^3}\,, \\ B_\theta &= \frac{m\sin\theta}{r^3}\,, \\ B_\varphi &= 0\,. \end{aligned} \tag{8.54}$$

Note that if we merely assume that the quantity given in (8.40) is small compared to 1, which may be achieved if $r >> a$ or $r << a$ or if θ is close to either 0 or π (the north or south poles of the sphere), then going back to the leading-order expression (8.43) for A_φ we then have $B_\varphi = 0$ and the approximate expressions

$$\begin{aligned} B_r &\approx \frac{I\pi a^2}{c}\,\frac{(2a^2+2r^2+ar\sin\theta)}{(a^2+r^2+2ar\sin\theta)^{5/2}}\,\cos\theta\,, \\ B_\theta &\approx -\frac{I\pi a^2}{c}\,\frac{(2a^2-r^2+ar\sin\theta)}{(a^2+r^2+2ar\sin\theta)^{5/2}}\,\sin\theta\,. \end{aligned} \tag{8.55}$$

As mentioned previously, an exact closed-form expression for A_φ may be obtained, at the price of employing elliptic functions. Thus (8.39) turns out to be given by

$$A_\varphi(r,\theta) = \frac{4Ia}{c\sqrt{a^2+r^2+2ar\sin\theta}} \left(\frac{(2-k^2)K(k^2) - 2E(k^2)}{k^2} \right) , \tag{8.56}$$

where

$$k^2 \equiv \frac{4ar\sin\theta}{a^2+r^2+2ar\sin\theta} , \tag{8.57}$$

and the complete elliptic integrals are defined by

$$K(z) = \int_0^{\pi/2} \frac{d\psi}{\sqrt{1-z\sin^2\psi}} , \qquad E(z) = \int_0^{\pi/2} \sqrt{1-z\sin^2\psi}\, d\psi . \tag{8.58}$$

Note that k^2 defined in (8.57) is just twice the quantity defined in eqn (8.40), which we assumed to be small in our previous expansion. If z is small, the elliptic integrals (8.58) admit power-series expansions

$$\begin{aligned} K(z) &= \frac{\pi}{2}\left(1 + \frac{1}{4}z + \frac{9}{64}z^2 + \frac{25}{256}z^3 + \frac{1225}{16384}z^4 + \cdots\right) , \\ E(z) &= \frac{\pi}{2}\left(1 - \frac{1}{4}z - \frac{3}{64}z^2 - \frac{5}{256}z^3 - \frac{175}{16384}z^4 + \cdots\right) . \end{aligned} \tag{8.59}$$

Thus if k^2 is small, we can obtain a series expansion for the vector potential A_φ by using

$$\frac{(2-k^2)K(k^2) - 2E(k^2)}{k^2} = \frac{\pi k^2}{16}\left(1 + \frac{3}{4}k^2 + \frac{75}{128}k^4 + \cdots\right) \tag{8.60}$$

in the expression for A_φ given in (8.56). The leading-order term coming from this expansion reproduces the expression (8.43) that we obtained previously.

Although we were able to obtain the exact expression (8.56) for the vector potential due to a circular current loop, we ended up looking at a power-series expansion in order to understand the nature of the solution. We could also have performed a series expansion prior to evaluating the integral (8.39), as we discussed previously, thereby avoiding the need to introduce the elliptic integrals. A slightly different way of doing this, in a systematic fashion, is to go back to the expression (8.37), and then to substitute in the expansion (4.154), which we reproduce here:

$$\frac{1}{|\vec{r}-\vec{r}\,'|} = \sum_{\ell\geq 0}\sum_{m=-\ell}^{\ell} \frac{4\pi}{2\ell+1}\frac{r_<^\ell}{r_>^{\ell+1}} \bar{Y}_{\ell m}(\theta',\varphi')\, Y_{\ell m}(\theta,\varphi) . \tag{8.61}$$

Writing $\cos\varphi' = \Re(e^{i\varphi'})$, where $\Re$ denotes the real part, and recalling that $Y_{\ell m}(\theta',\varphi')$ has φ' dependence of the form $e^{im\varphi'}$ (i.e. $Y_{\ell m}(\theta',\varphi') = e^{im\varphi'}\, Y_{\ell m}(\theta',0)$, see (4.112)), we can

perform the φ' integration, as well as the r' and θ' integrations, to give

$$\begin{aligned} A_\varphi &= \frac{Ia}{c}\sum_{\ell\geq 0}\sum_{m=-\ell}^{\ell}\frac{4\pi}{2\ell+1}\,\frac{r_<^\ell}{r_>^{\ell+1}}\,\Re\int_0^{2\pi} d\varphi'\,e^{\mathrm{i}(1-m)\varphi'}\,Y_{\ell,m}(\tfrac{1}{2}\pi,0)\,Y_{\ell,m}(\theta,\varphi)\Big|_{\varphi=0}\,, \\ &= \frac{8I\pi^2 a}{c}\left(\sum_{\ell\geq 1}\frac{Y_{\ell,1}(\frac{1}{2}\pi,0)}{2\ell+1}\,\frac{r_<^\ell}{r_>^{\ell+1}}\,Y_{\ell,1}(\theta,0)\right)\,, \end{aligned} \tag{8.62}$$

where $r_<$ and $r_>$ denotes the lesser, and greater, of a and r. (Note that having integrated over φ', with $\int_0^{2\pi} d\varphi'\,e^{\mathrm{i}(1-m)\varphi'} = 2\pi\,\delta_{m,1}$, and having set $\varphi = 0$ (since, as before, the prescription for calculating A_φ in (8.37) involves evaluating A_y at $\varphi = 0$), the expression in the second line of (8.62) is already real, and so we no longer need to take the real part.)

The quantities $Y_{\ell,1}(\frac{1}{2}\pi,0)$ are just ℓ-dependent constants that can be read off from the definition (4.112):

$$Y_{\ell,1}(\tfrac{1}{2}\pi,0) = \sqrt{\frac{2\ell+1}{4\pi\ell(\ell+1)}}\,P_\ell^1(0)\,. \tag{8.63}$$

Thus $Y_{\ell,1}(\frac{1}{2}\pi,0) = 0$ when ℓ is even (recall that $P_\ell^m(x)$ is given by the generalised Rodrigues formula (4.101), which shows that it is an *odd* function of x when $\ell + m$ is odd). When $\ell = 2n+1$, we have

$$Y_{2n+1,1}(\tfrac{1}{2}\pi,0) = \sqrt{\frac{2\ell+1}{4\pi\ell(\ell+1)}}\,\left[\frac{(-1)^{n+1}\Gamma(n+\frac{3}{2})}{\Gamma(n+1)\Gamma(\frac{3}{2})}\right]\,. \tag{8.64}$$

The gauge potential for the circular current loop can therefore be written as

$$A_\varphi = -\frac{\pi Ia}{c}\sum_{n\geq 0}\frac{(-1)^n(2n-1)!!}{2^n\,(n+1)!}\,\frac{r_<^{2n+1}}{r_>^{2n+2}}\,P_{2n+1}^1(\cos\theta)\,, \tag{8.65}$$

where

$$(2n-1)!! \equiv (2n-1)(2n-3)(2n-5)\times\cdots\times 5\times 3\times 1\,, \tag{8.66}$$

and by definition $(2n-1)!!$ is equal to 1 when $n = 0$.

The fact that the associated Legendre functions $P_\ell^m(\cos\theta)$ with $m = 1$ enter here in the expansion for A_φ is a reflection of the vector-like nature of the potential $\vec{A}$. (In contrast to the expansion for an azimuthally-symmetric scalar potential, where, as we have seen in electrostatics, the $m = 0$ associated Legendre functions $P_\ell^0(\cos\theta)$, otherwise known as the Legendre polynomials $P_\ell(\cos\theta)$, occur.) If you look back to the derivation of A_φ in eqn (8.62), you will see that the reason why it is the $m = 1$ harmonics that arise here is because of that $\cos\varphi'$ factor in the integrand in eqn (8.37). In turn, as emphasised in footnote 50, that $\cos\varphi'$ factor arose from the proper handling of the transformation from spherical polar vectors into Cartesian vectors in order to make use of eqn (8.9).

In order to calculate the component B_r of the $\vec{B}$ field, we see from (8.46) that it is necessary to evaluate the derivative of the associated Legendre functions $P_\ell^1(x)$ with respect to x. Recalling from (4.99) that

$$P_\ell^m(x) \equiv (-1)^m\,(1-x^2)^{m/2}\,\frac{d^m}{dx^m}\,P_\ell(x)\,, \tag{8.67}$$

we see that

$$\frac{d}{dx}\Big((1-x^2)^{1/2}P_\ell^1(x)\Big) = -\frac{d}{dx}\Big((1-x^2)\frac{dP_\ell(x)}{dx}\Big) = \ell(\ell+1)P_\ell(x)\,, \tag{8.68}$$

since we know that the Legendre polynomials $P_\ell(x)$ satisfy the standard Legendre equation $\big((1-x^2)P_\ell'\big)' + \ell(\ell+1)P_\ell = 0$.

The upshot from this is that the radial component of the $\vec{B}$ field for the circular current loop is given by

$$B_r = \frac{2\pi I a}{cr}\sum_{n\ge 0}\frac{(-1)^n(2n+1)!!}{2^n\,n!}\,\frac{r_<^{2n+1}}{r_>^{2n+2}}\,P_{2n+1}(\cos\theta)\,. \tag{8.69}$$

The θ component of $\vec{B}$ is given by

$$B_\theta = \frac{2\pi I}{ca}\sum_{n\ge 0}\frac{(-1)^n(2n-1)!!}{2^n\,n!}\,\Big(\frac{r}{a}\Big)^{2n}\,P_{2n+1}^1(\cos\theta)\,,\quad \text{when}\quad r<a\,, \tag{8.70}$$

or by

$$B_\theta = -\frac{\pi I a^2}{cr^3}\sum_{n\ge 0}\frac{(-1)^n(2n+1)!!}{2^n\,(n+1)!}\,\Big(\frac{a}{r}\Big)^{2n}\,P_{2n+1}^1(\cos\theta)\,,\quad \text{when}\quad r>a\,, \tag{8.71}$$

and the φ component is zero.

8.3 Localised Current Distribution

Suppose now we consider a localised region of space within which currents are flowing, described by the current density $\vec{J}(\vec{r})$. Outside this region, it is assumed that $\vec{J}(\vec{r}) = 0$. We can proceed in a manner that is precisely analogous to the earlier discussion we gave of localised charge distributions, now using the expression

$$\vec{A}(\vec{r}) = \frac{1}{c}\int\frac{\vec{J}(\vec{r}\,')}{|\vec{r}-\vec{r}\,'|}\,d^3\vec{r}\,' \tag{8.72}$$

for the magnetic vector potential, and then Taylor expanding $|\vec{r}-\vec{r}\,'|^{-1}$ in inverse powers of r, using (6.20). Thus, keeping just the first couple of orders in the expansion, we shall

have

$$\begin{aligned}\frac{1}{|\vec{r}-\vec{r}\,'|} &= \frac{1}{r} - x'_i\,\partial_i\frac{1}{r} + \cdots\,,\\ &= 1 + \frac{x_i\,x'_i}{r^3} + \cdots\,,\\ &= \frac{1}{r} + \frac{\vec{r}\,'\cdot\vec{r}}{r^3} + \cdots\,. \end{aligned} \tag{8.73}$$

It follows from (8.72) that we shall have

$$A_i(\vec{r}) = \frac{1}{c\,r}\int \vec{J}_i(\vec{r}\,')d^3\vec{r}\,' + \frac{x_j}{c\,r^3}\int x'_j\,J_i(\vec{r}\,')\,d^3\vec{r}\,' + \cdots\,. \tag{8.74}$$

The first term in (8.74) vanishes, as can be seen from the following argument. We know that the current density is conserved, $\vec{\nabla}\cdot\vec{J}=0$, i.e. $\partial_i J_i = 0$. Now consider the quantity $\partial_i(x_j\,J_i)$, which is therefore given by

$$\begin{aligned}\partial_i(x_j\,J_i) &= (\partial_i x_j)\,J_i + x_j\,\partial_i J_i\,,\\ &= \delta_{ij}\,J_i\,,\\ &= J_j\,. \end{aligned} \tag{8.75}$$

If we integrate $\partial_i(x_j\,J_i)$ over all space it will give zero by the divergence theorem, since $\vec{J}$ vanishes outside some bounded domain:

$$\int \partial_i(x_j\,J_i)\,d^3\vec{r} = \int_{(\text{sphere at }\infty)}(x_j\,J_i)\,dS_i = 0\,, \tag{8.76}$$

and hence we conclude that $\int J_j\,d^3\vec{r}=0$.

To discuss the second term in (8.74), it is useful first to review the way in which one describes the vector product using index notation. This is done by introducing the totally antisymmetric tensor ϵ_{ijk}, which is then defined by

$$\epsilon_{123} = 1\,. \tag{8.77}$$

The total antisymmetry under the exchange of any pair of indices implies that we must therefore have

$$\epsilon_{123} = \epsilon_{231} = \epsilon_{312} = +1\,, \qquad \epsilon_{132} = \epsilon_{321} = \epsilon_{213} = -1\,, \tag{8.78}$$

with all other components vanishing. Using this, we can clearly write the vector product $\vec{V} = \vec{A}\times\vec{B}$ as

$$V_i = \epsilon_{ijk}\,A_j\,B_k\,. \tag{8.79}$$

It is straightforward to show, by enumerating all the possible assignments for the free indices i, j, k and ℓ, that[51]

$$\epsilon_{ijm}\,\epsilon_{k\ell m} = \delta_{ik}\,\delta_{j\ell} - \delta_{i\ell}\,\delta_{jk}\,. \tag{8.80}$$

Using this, many vector identities involving a pair of vector products can easily be proven. For example, if $\vec{V} \equiv \vec{A} \times (\vec{B} \times \vec{C})$, then we shall have

$$\begin{aligned} V_i &= \epsilon_{ijm}\,A_j\,(\vec{B} \times \vec{C})_m\,, \\ &= \epsilon_{ijm}\,A_j\,\epsilon_{mk\ell}B_k\,C_\ell\,, \\ &= \epsilon_{ijm}\,\epsilon_{k\ell m}\,A_j\,B_k\,C_\ell\,, \\ &= (\delta_{ik}\,\delta_{j\ell} - \delta_{i\ell}\,\delta_{jk})\,A_j\,B_k\,C_\ell\,, \\ &= B_i\,A_j\,C_j - C_i\,A_j\,B_j\,, \end{aligned} \tag{8.81}$$

or, in 3-vector notation, the well-known identity[52]

$$\vec{V} \equiv \vec{A} \times (\vec{B} \times \vec{C}) = \vec{B}\,(\vec{A}\cdot\vec{C}) - \vec{C}\,(\vec{A}\cdot\vec{B})\,. \tag{8.82}$$

(On the first line of eqn (8.81) we are using the notation $(\vec{B} \times \vec{C})_m$ to mean the m'th component of $\vec{B} \times \vec{C}$.)

Returning to the problem in hand we note, since $\partial_i J_i = 0$, that

$$\begin{aligned} \partial_i(x_j\,x_k\,J_i) &= (\partial_i x_j)\,x_k\,J_i + x_j\,(\partial_i x_k)\,J_i + x_j\,x_k\,\partial_i J_i \\ &= \delta_{ij}\,x_k\,J_i + \delta_{ik}\,x_j\,J_i \\ &= x_k\,J_j + x_j\,J_k\,. \end{aligned} \tag{8.83}$$

Using the divergence theorem, the integral of the left-hand side over all space is zero, and so we conclude that

$$\int x_j\,J_k\,d^3\vec{r} = -\int x_k\,J_j\,d^3\vec{r}\,. \tag{8.84}$$

This means that the integral in the second term in (8.74) can be written as

$$\int x'_j\,J_i(\vec{r}\,')\,d^3\vec{r}\,' = -\tfrac{1}{2}\int (x'_i\,J_j - x'_j\,J_i)\,d^3\vec{r}\,'\,. \tag{8.85}$$

Now, as can easily be seen using (8.80),

$$\tfrac{1}{2}(x'_i\,J_j - x'_j\,J_i) = \tfrac{1}{2}\epsilon_{ijk}\,(\vec{r}\,' \times \vec{J})_k\,, \tag{8.86}$$

[51] The labour involved in this proof is hugely reduced, becoming almost a triviality, if one takes note of the symmetries of the problem!

[52] Remembering just the simple expression (8.80) enables one to prove easily, "on demand," almost all the identities of three-dimensional vector calculus.

where $(\vec{r}\,' \times \vec{J})_k$ denotes the k'th component of the vector $(\vec{r}\,' \times \vec{J})$, and so we conclude that the second term in (8.74) can be written as

$$A_i(\vec{r}) = -\frac{x_j}{2c\,r^3}\,\epsilon_{ijk}\int \big(\vec{r}\,' \times \vec{J}(\vec{r}\,')\big)_k\,d^3\vec{r}\,', \tag{8.87}$$

and hence

$$\vec{A}(\vec{r}) = \frac{\vec{m}\times\vec{r}}{r^3}, \tag{8.88}$$

where we have defined the magnetic moment $\vec{m}$

$$\vec{m} = \frac{1}{2c}\int \vec{r}\times\vec{J}(\vec{r})\,d^3\vec{r} \tag{8.89}$$

of the current distribution $\vec{J}$. Prior to performing the volume integration, we may define the magnetisation $\vec{\mathcal{M}}$ by

$$\vec{\mathcal{M}}(\vec{r}) = \frac{1}{2c}\,\vec{r}\times\vec{J}, \qquad \text{and so} \quad \vec{m} = \int \vec{\mathcal{M}}(\vec{r})\,d^3\vec{r}. \tag{8.90}$$

We have shown that (8.88) is the leading-order term in the multipole expansion describing the magnetic vector potential $\vec{A}$ of a localised current distribution. Let us now calculate the magnetic field $\vec{B} = \vec{\nabla}\times\vec{A}$. In index notation we have

$$\begin{aligned} B_i &= \epsilon_{ijk}\,\partial_j A_k = \epsilon_{ijk}\,\epsilon_{k\ell m}\,\partial_j\Big(\frac{m_\ell\,x_m}{r^3}\Big),\\ &= (\delta_{i\ell}\,\delta_{jm} - \delta_{im}\,\delta_{j\ell})\,\partial_j\Big(\frac{m_\ell\,x_m}{r^3}\Big),\\ &= m_i\,\partial_j\Big(\frac{x_j}{r^3}\Big) - m_j\,\partial_j\Big(\frac{x_i}{r^3}\Big). \end{aligned} \tag{8.91}$$

(The Cartesian components m_i of the magnetic moment vector are, of course, constants.) Now, away from the origin (i.e. for $r > 0$) we have

$$\partial_j\Big(\frac{x_j}{r^3}\Big) = \frac{\partial_j x_j}{r^3} - \frac{3x_j}{r^4}\,\frac{x_j}{r} = \frac{3}{r^3} - \frac{3}{r^3} = 0, \tag{8.92}$$

and

$$\partial_j\Big(\frac{x_i}{r^3}\Big) = \frac{\delta_{ij}}{r^3} - \frac{3x_i\,x_j}{r^5}, \tag{8.93}$$

and hence we find

$$\vec{B} = \frac{3(\vec{m}\cdot\vec{n})\,\vec{n} - \vec{m}}{r^3}, \qquad \text{for} \qquad r > 0, \tag{8.94}$$

where as usual we define the unit vector $\vec{n} = \vec{r}/r$. Thus $\vec{B}$ has exactly the same form as the electric field of an electric dipole $\vec{p}$ (see (6.37)).

It is of interest also to consider the expression for $\vec{B}$ including the origin $r = 0$.[53] Recalling that $\nabla^2(\frac{1}{r}) = -4\pi\,\delta^3(\vec{r})$, and that $\frac{x_j}{r^3} = -\partial_j(\frac{1}{r})$, we see that

$$\partial_j\Big(\frac{x_j}{r^3}\Big) = 4\pi\,\delta^3(\vec{r}). \tag{8.95}$$

[53]Note that it only really makes sense to do this in the case of an idealised "point dipole" of zero size.

What about the other term in the last line of (8.91)? It is clear that $\partial_j(\frac{x_i}{r^3})$ must also have a delta-function term at the origin, since we already know that taking its trace gives $\partial_j(\frac{x_j}{r^3}) = 4\pi\,\delta^3(\vec{r})$. Obviously no direction in 3-dimensional space can be preferred over any other, and so the delta-function term in $\partial_j(\frac{x_i}{r^3})$ must be *isotropic*, i.e. proportional to δ_{ij}. Thus it must be that

$$\partial_i\Big(\frac{x_j}{r^3}\Big) = \frac{\delta_{ij}}{r^3} - \frac{3x_i\,x_j}{r^5} + a\,\delta_{ij}\,\delta^3(\vec{r})\,, \tag{8.96}$$

where a is some constant to be determined. Taking the trace of this formula, the right-hand side must give the right-hand side of eqn (8.95), and so we see that $a = \frac{4\pi}{3}$. Thus we obtain the result that

$$\partial_j\Big(\frac{x_i}{r^3}\Big) = \frac{\delta_{ij}}{r^3} - \frac{3x_i\,x_j}{r^5} + \frac{4\pi}{3}\delta_{ij}\,\delta^3(\vec{r})\,, \tag{8.97}$$

Putting all the above results together, it follows from (8.91) that

$$\vec{B} = \frac{3(\vec{m}\cdot\vec{n})\,\vec{n} - \vec{m}}{r^3} + \frac{8\pi}{3}\,\vec{m}\,\delta(\vec{r})\,. \tag{8.98}$$

(We emphasise again that the additional $\delta(\vec{r})$ term is not relevant for the discussion of the field due to a locaised current source as viewed from some distant poistion at $\vec{r}$, but it would be relevant in a discussion of an idealised "point magnetic dipole" located at the origin, where it could make sense to consider the magnetic field in the limit where $\vec{r}$ goes to zero.)

If the current distribution takes the form of a planar closed loop of current (for example, a current flowing round a planar wire loop), then the general expression (8.89) for the magnetic moment reduces to

$$\vec{m} = \frac{I}{2c}\oint \vec{r}\times d\vec{r}, \tag{8.99}$$

where I is the current. Note that the magnetic moment is perpendicular to the plane of the loop. Since $\frac{1}{2}\vec{r}\times d\vec{r}$ is the area element of the triangular wedge whose vertices lie at the origin, and the points $\vec{r}$ and $\vec{r}+d\vec{r}$ on the loop, it follows that $\frac{1}{2}\oint \vec{r}\times d\vec{r}$ gives the area of the loop, and so the magnitude m of the magnetic moment for a planar current loop of area A is just given by

$$m = \frac{IA}{c}\,. \tag{8.100}$$

This result generalises, to a planar loop of arbitray shape, the result obtained in (8.52) in the case of a circular current loop.

Another special case we may consider is when the current distribution is generated by N point charges q_a, for $1 \le a \le N$, located at points $\vec{r}_a(t)$ and moving with velocities

$\vec{v}_a = \dfrac{d\vec{r}_a(t)}{dt}$. The current density is therefore given by

$$\vec{J}(\vec{r}) = \sum_{a=1}^{N} q_a\, \vec{v}_a\, \delta^3\big(\vec{r} - \vec{r}_a(t)\big)\,. \tag{8.101}$$

Inserting this into (8.89), the integration can be performed, giving

$$\vec{m} = \frac{1}{2c} \sum_{a=1}^{N} q_a\, (\vec{r}_a \times \vec{v}_a(t))\,. \tag{8.102}$$

Now, if the a'th particle has mas M_a, then its orbital angular momentum is given by

$$\vec{L}_a = M_a\, (\vec{r}_a \times \vec{v}_a(t))\,, \tag{8.103}$$

and so we have

$$\vec{m} = \sum_{a=1}^{N} \frac{q_a}{2cM_a}\, \vec{L}_a\,. \tag{8.104}$$

If all the particles have the same charge to mass ratio,

$$\frac{q_a}{M_a} = \frac{q}{M}\,, \qquad \text{for each } a\,, \tag{8.105}$$

then we have the simple relation

$$\vec{m} = \frac{q}{2Mc}\, \vec{L}\,, \tag{8.106}$$

where

$$\vec{L} = \sum_{a=1}^{N} \vec{L}_a \tag{8.107}$$

is the total orbital angular momentum of the system of particles.

8.4 Force on a current distribution in an external $\vec{B}$ field

Suppose a localised current distribution is located in a region where there is an externally-generated magnetic field $\vec{B}(\vec{r})$, which may be position dependent. If we assume that the magnetic field varies slowly with position, then we can make a Taylor expansion of $\vec{B}(\vec{r})$ around some point (which can conveniently be taken to be the origin), and keep just the leading-order terms. Thus we shall have

$$B_i(\vec{r}) = B_i(0) + \vec{r}\cdot\vec{\nabla}B_i(0) + \cdots\,, \tag{8.108}$$

where, of course, in the second term the argument is set to zero *after* taking the gradient (i.e. $\vec{\nabla}B_i(0)$ means $\vec{\nabla}B_i(\vec{r})\big|_{\vec{r}=0}$). Substituting into the expression

$$\vec{F} = \frac{1}{c} \int \vec{J}\,(\vec{r}\,) \wedge \vec{B}(\vec{r}\,)\, d^3\vec{r} \tag{8.109}$$

for the force on a current distribution, we therefore find

$$\vec{F} = -\frac{1}{c}\,\vec{B}(0) \times \int \vec{J}\,(\vec{r})\,d^3\vec{r} + \frac{1}{c}\int \vec{J}\,(\vec{r}) \times [(\vec{r}\cdot\vec{\nabla})\vec{B}(0)]\,d^3\vec{r} + \cdots\,. \tag{8.110}$$

As we already saw earlier, the integral in the first term vanishes, and so the leading-order contribution to the force comes from the second term. In index notation, the second term is

$$F_i = \frac{1}{c}\,\epsilon_{ijk}\,(\partial_\ell B_k)(0)\int x_\ell\,J_j(\vec{r})\,d^3\vec{r}\,. \tag{8.111}$$

Using (8.85), (8.86) and (8.89), we therefore have

$$\begin{aligned} F_i &= -\epsilon_{ijk}\,\epsilon_{j\ell m}\,m_m\,\partial_\ell B_k(0)\,, \\ &= -m_i\,\partial_k B_k(0) + m_k\,\partial_i B_k(0)\,, \\ &= m_k\,\partial_i B_k(0)\,, \end{aligned} \tag{8.112}$$

where, in getting to the final line, we have used $\vec{\nabla}\cdot\vec{B} = 0$. Thus, since $\vec{m}$ is a constant, we can write $F_i = \partial_i\,(m_k\,B_k)$ and so

$$\vec{F} = \vec{\nabla}(\vec{m}\cdot\vec{B})\,. \tag{8.113}$$

The expression (8.113) for the force on a magnetic dipole $\vec{m}$ in a magnetic field $\vec{B}$ shows that we can define a potential energy

$$U = -\vec{m}\cdot\vec{B}\,, \tag{8.114}$$

in terms of which the force is given by $\vec{F} = -\vec{\nabla}U$.

Note that since $\vec{B}$ is assumed to be generated by distant current sources, and so $\vec{\nabla}\times\vec{B} = 0$ in the region under consideration, we have $\epsilon_{k\ell m}\partial_\ell B_m = 0$ and hence $\epsilon_{ijk}\,\epsilon_{k\ell m}\partial_\ell B_m = 0$. Using (8.80), this implies that

$$\partial_i B_j - \partial_j B_i = 0\,. \tag{8.115}$$

Using also the constancy of $\vec{m}$, we can then manipulate (8.113) to give

$$F_i = \partial_i(m_j B_j) = m_j\,\partial_i B_j = m_j\,\partial_j B_i\,, \tag{8.116}$$

and so we can give the alternative expression

$$\vec{F} = (\vec{m}\cdot\vec{\nabla})\vec{B} \tag{8.117}$$

for the force on the magnetic moment.

To calculate the torque on the current distribution, we substitute (8.108) into the general expression for the torque that we derived previously,

$$\vec{N} = \frac{1}{c} \int \vec{r} \times \Big(\vec{J}(\vec{r}) \times \vec{B}(\vec{r}) \Big) \, d^3\vec{r}. \tag{8.118}$$

Unlike in the calculation of the force, here the first term in (8.108) gives a non-zero contribution, and so to leading order we have

$$\vec{N} = \frac{1}{c} \int \vec{r} \times \Big(\vec{J}(\vec{r}) \times \vec{B}(0) \Big) \, d^3\vec{r}. \tag{8.119}$$

Expanding out the vector triple product, and writing in index notation, we have

$$N_i = \frac{1}{c} \, B_j(0) \int x_j \, J_i(\vec{r}) \, d^3\vec{r} - \frac{1}{c} \, B_i(0) \int x_j \, J_j(\vec{r}) \, d^3\vec{r}. \tag{8.120}$$

The second term in (8.120) integrates to zero. This can be seen by integrating the identity

$$\vec{\nabla} \cdot (r^2 \, \vec{J}\,) = 2\vec{r} \cdot \vec{J} + r^2 \, \vec{\nabla} \cdot \vec{J} = 2\vec{r} \cdot \vec{J} \tag{8.121}$$

over all space, and using the divergence theorem to turn the left-hand side into a surface integral at infinity. (Or, equivalently, just use the result in eqn (8.158) showing that $\int x_i \, J_j \, d^3\vec{r}$ is antisymmetric in i and j, from which it immediately follows that if we contract with δ_{ij} (which is symmetric), the result will be zero.) Using (8.85), (8.86) and (8.89), the first term in (8.120) can be seen to give

$$\vec{N} = \vec{m} \times \vec{B}(0) \, . \tag{8.122}$$

8.5 Magnetically permeable media

In section 7, we discussed the phenomenological description of dielectric media, in which one introduces a macroscopic $\vec{D}$ field in addition to the fundamental electric field $\vec{E}$. The essential idea is that for many purposes, one can give a macroscopic description of the effect of a piece of dielectric medium, such as a salt crystal or a block of glass, in which the microscopic contributions of each atom or molecule within the medium are averaged over, so that on the large scale a relatively simple description of the electrical properties of the material as a whole can be given.

In a very similar vein, in magnetostatics one may give a macroscopic description of magnetically permeable materials. As in the case of electrostatics, the idea here is that whilst one could, in principle, simply apply the fundamental Maxwell equations of magnetostatics in a vacuum, i.e.

$$\vec{\nabla} \cdot \vec{B} = 0 \, , \qquad \vec{\nabla} \times \vec{B} = \frac{4\pi}{c} \, \vec{J}, \tag{8.123}$$

at the microscopic level, to the entire system of atoms or molecules, it instead suffices, for many purposes, to work with a macroscopically averaged description. In this description, one introduces a phenomenological field $\vec{H}$ in addition to the fundamental field $\vec{B}$ of magnetostatics. Rather confusingly, $\vec{H}$ is called the *magnetic field*, while the fundamental field $\vec{B}$ is called the *magnetic induction*.

Since the derivation of the phenomenological equations for magnetostatics is closely parallel to the derivation for electrostatics in section 7, we shall just present the final conclusions here; a derivation can be found in any of the standard textbooks.

The effect of the ensemble of atoms or molecules in the medium is to give rise to a magnetic moment density $\vec{M}(\vec{r})$, which comes from an averaging over all the atoms or molecules. The magnetic induction $\vec{B}$ still satisfies the original Maxwell equation $\vec{\nabla}\cdot\vec{B}=0$, and so we may still write

$$\vec{B}=\vec{\nabla}\times\vec{A}\,, \tag{8.124}$$

but the equation (8.72) giving the vector potential for a localised current distribution is now modified to

$$\vec{A}(\vec{r})=\frac{1}{c}\int\frac{\vec{J}(\vec{r}\,')+c\,\vec{\nabla}'\times\vec{M}(\vec{r}\,')}{|\vec{r}-\vec{r}\,'|}\,d^3\vec{r}\,'\,. \tag{8.125}$$

The magnetisation can thus be thought of as giving rise to an effective magnetisation current density

$$\vec{J}_M=c\,\vec{\nabla}\times\vec{M}\,. \tag{8.126}$$

The second of the Maxwell equations in (8.123) now becomes

$$\vec{\nabla}\times\vec{B}=\frac{4\pi}{c}\,\vec{J}+4\pi\,\vec{\nabla}\times\vec{M}\,, \tag{8.127}$$

or, in other words,

$$\vec{\nabla}\times\Big(\vec{B}-4\pi\,\vec{M}\Big)=\frac{4\pi}{c}\,\vec{J}\,. \tag{8.128}$$

We are therefore led to define a new field $\vec{H}$ by

$$\vec{H}=\vec{B}-4\pi\,\vec{M}\,. \tag{8.129}$$

The free-space Maxwell equations (8.123) therefore become

$$\vec{\nabla}\cdot\vec{B}=0\,,\qquad\vec{\nabla}\times\vec{H}=\frac{4\pi}{c}\,\vec{J} \tag{8.130}$$

in the presence of a magnetically-permeable medium. Recall that in electrostatics, we had the Maxwell equations

$$\vec{\nabla}\cdot\vec{D}=4\pi\,\rho\,,\qquad\vec{\nabla}\times\vec{E}=0\,, \tag{8.131}$$

in the presence of a dielectric medium. Note that in each case it is the field equation with the source term on the right-hand side that is modified in the presence of the medium, whilst the equation that has no source term is unmodified. In other words, the phenomenological description of dielectric or permeable media modifies the *field equations* but leaves the *Bianchi identities* unchanged.

In order to apply the phenomenological description of magnetostatics, it is necessary to know the so-called *constitutive relation* between $\vec{B}$ and $\vec{H}$. This can be quite a complicated business in general, especially in the case of ferromagnetic materials where there may not even be a single-valued functional relation between $\vec{B}$ and $\vec{H}$.

The simplest materials to consider are paramagnetic and diamagnetic media. For these, there is a linear relation between $\vec{B}$ and $\vec{H}$, with

$$\vec{B} = \mu \vec{H} \,, \tag{8.132}$$

where μ is a constant called the magnetic permeability. For paramagnetic materials, μ is slightly greater than 1, whilst for diamagnetic materials, μ is slightly less than 1. (The deviations from 1 are typically only a few parts in 10^5 for diamagnetic or paramagnetic materials.)

For many ferromagnetic materials, a relation of the form (8.132) is approximately valid, provided the fields are sufficiently weak. In these materials, μ is typically in the range between 10 and 10^4.

Another common situation is that of a permanent magnetic material, where there is magnetisation even in the absence of an applied field.

In much of the remainder of our consideration of magnetically-permeable media, we shall assume that the simple linear constitutive relation (8.132) holds.

8.6 Boundary conditions at medium interfaces

In section (1.4) we gave a discussion of the boundary condition that must be satisfied at an interface between two magnetically-permeable media. These were obtained by considering a "pill-box" volume integration of the equation $\vec{\nabla}\cdot\vec{B} = 0$, and integrations of $\vec{\nabla}\times\vec{H} = (4\pi/c)\vec{J}$ over slender rectangular loops straddling the boundary. These lead, respectively, to the conditions

$$\begin{aligned} \vec{n}\cdot(\vec{B}_2 - \vec{B}_1) &= 0 \,, \\ \vec{n}\times(\vec{H}_2 - \vec{H}_1) &= \frac{4\pi}{c}\,\vec{K} \,, \end{aligned} \tag{8.133}$$

where $\vec{n}$ is the unit vector normal to the interface, pointing from medium 1 into medium 2, and $\vec{K}$ is the surface current density (i.e. current per unit length, flowing in the surface).

The first equation in (8.133) says that the normal component of $\vec{B}$ must be continuous across the boundary. In many common situations there will be no surface currents at the interface, and then the second equation in (8.133) just says that the tangential components of $\vec{H}$ must be continuous across the boundary.

8.7 Techniques for solving boundary-value problems in magnetostatics

There are various techniques that can be applied in order to solve for the magnetic fields in a boundary-value problem in magnetostatics. Depending upon the the circumstances, one or another may be more convenient. For concreteness and simplicity, we shall typically focus on situations where the simple relation $\vec{B} = \mu \vec{H}$ holds, with the assumption that in any given region μ is a constant. (We shall allow μ to take different constant values in different regions, as would be the case in a typical boundary-value problem with an interface between media.)

8.7.1 Using the vector potential

The most direct way to formulate the problem is just to take the Maxell equations of magnetostatics,

$$\vec{\nabla} \cdot \vec{B} = 0\,, \qquad \vec{\nabla} \times \vec{H} = \frac{4\pi}{c}\,\vec{J}\,, \tag{8.134}$$

solve the first equation by writing $\vec{B}$ in terms of the vector potential,

$$\vec{B} = \vec{\nabla} \times \vec{A}\,, \tag{8.135}$$

and then plug this into the second of the Maxwell equations. Of course in general, this will give a very complicated result if there is a complicated constitutive relation between $\vec{B}$ and $\vec{H}$, but if we make the assumption that $\vec{B} = \mu \vec{H}$, where μ is piecewise constant, then in the various regions we shall have $\vec{\nabla} \times (\vec{\nabla} \times \vec{A}) = (4\pi\mu/c)\,\vec{J}$. Using Cartesian coordinates and expanding out the vector triple product, we then obtain, by working in the Coulomb gauge $\vec{\nabla} \cdot \vec{A} = 0$, the equation

$$\nabla^2 \vec{A} = -\frac{4\pi\mu}{c}\,\vec{J}\,, \tag{8.136}$$

where ∇^2 is the usual scalar Laplacian. This is very similar to the equation in free space, except that now we have the permeability μ appearing on the right-hand side.

8.7.2 Magnetic scalar potential satisfying Laplace's equation

If there are no free currents in the problem, as is often the case, the Maxwell field equation becomes

$$\vec{\nabla} \times \vec{H} = 0\,, \tag{8.137}$$

and this can be solved by writing $\vec{H}$ as the gradient of a scalar, just as one does for the electric field in electrostatics. Thus we may write

$$\vec{H} = -\vec{\nabla}\Phi_M\,, \tag{8.138}$$

where Φ_M is the *magnetic scalar potential*. If we assume the simple constitutive relation $\vec{B} = \mu\,\vec{H}$ between $\vec{B}$ and $\vec{H}$, where μ is piecewise constant, then in each region the remaining Maxwell equation $\vec{\nabla} \cdot \vec{B} = 0$ becomes

$$\nabla^2\Phi_M = 0\,. \tag{8.139}$$

Solving boundary-value magnetostatics problems where $\vec{J} = 0$ is thus closely analogous to solving boundary-value problems in electrostatics; one just has to solve Laplace's equation in the various regions, and then impose the boundary conditions at interfaces as discussed in section 8.6.

As an illustration of the method, consider a spherical shell of material with magnetic permeability μ, with inner radius a and outer radius b. The regions inside and outside the shell are assumed to be empty space. Outside the shell, we take the magnetic field to be asymptotically uniform, $\vec{B} \longrightarrow \vec{B}_0$, with $\vec{B}_0$ directed along the z axis. The problem is to solve for the magnetic fields everywhere. A particular point of interest will be to calculate the magnetic field inside the shell, in order to study the phenomenon of *magnetic shielding*.

In the regions $0 \le r < a$ and $r > b$ we have $\vec{B} = \vec{H}$, whilst in the region $a < r < b$ we have $\vec{B} = \mu\,\vec{H}$. There are no external currents, and so the boundary conditions (8.133) will apply at the interfaces at $r = a$ and $r = b$. There is azimuthal symmetry around the z axis, and so in spherical polar coordinates the magnetic scalar potential will depend only on r and θ. In order to avoid cluttering the subsequent equations with the subscript M on all the expressions for the magnetic scalar potentials, we shall omit it from now on in this discussion, and just write Φ. Since it must satisfy Laplace's equation in the three regions,

it must take the form

$$\begin{aligned}
0 < r < a: \qquad & \Phi_{(1)}(r,\theta) = \sum_{\ell\geq 0} \alpha_\ell\, r^\ell\, P_\ell(\cos\theta)\,, \\
a < r < b: \qquad & \Phi_{(2)}(r,\theta) = \sum_{\ell\geq 0} \Big(\beta_\ell\, r^\ell + \frac{\gamma_\ell}{r^{\ell+1}}\Big)\, P_\ell(\cos\theta)\,, \\
r > b: \qquad & \Phi_{(3)}(r,\theta) = -B_0\, r\cos\theta + \sum_{\ell\geq 0} \frac{\delta_\ell}{r^{\ell+1}}\, P_\ell(\cos\theta)\,.
\end{aligned} \tag{8.140}$$

In just the same way as we saw previously when solving boundary-value problems in electrostatics, the only modes that will actually occur in the expansions are the ones that arise in the specification of the boundary conditions (and, in more general problems, modes that arise in any source terms). In the present case, this means that only the $\ell = 1$ modes can arise, since the asymptotic boundary condition $\Phi \longrightarrow -B_0\, r\, \cos\theta = -B_0\, r\, P_1(\cos\theta)$ is the only one that is causing the magnetic field to be non-zero. Thus instead of the full expansions (8.140), we can just write

$$\begin{aligned}
0 < r < a: \qquad & \Phi_{(1)}(r,\theta) = \alpha\, r\, \cos\theta\,, \\
a < r < b: \qquad & \Phi_{(2)}(r,\theta) = \Big(\beta\, r + \frac{\gamma}{r^2}\Big)\, \cos\theta\,, \\
r > b: \qquad & \Phi_{(3)}(r,\theta) = \Big(-B_0\, r + \frac{\delta}{r^2}\Big)\, \cos\theta\,,
\end{aligned} \tag{8.141}$$

with the four constants α, β, γ and δ to be determined by the boundary conditions at $r = a$ and $r = b$.

These boundary conditions require that the radial components of $\vec{B}$ be continuous, and the tangential components of $\vec{H}$ be continuous, as in (8.133), at each of $r = a$ and $r = b$ (with the surface current $\vec{K} = 0$). We therefore have the conditions

$$\begin{aligned}
\mu\, \frac{\partial\Phi_{(2)}}{\partial r}\Big|_{r=a} &= \frac{\partial\Phi_{(1)}}{\partial r}\Big|_{r=a}\,, \qquad & \frac{\partial\Phi_{(2)}}{\partial\theta}\Big|_{r=a} &= \frac{\partial\Phi_{(1)}}{\partial\theta}\Big|_{r=a}\,, \\
\mu\, \frac{\partial\Phi_{(2)}}{\partial r}\Big|_{r=b} &= \frac{\partial\Phi_{(3)}}{\partial r}\Big|_{r=b}\,, \qquad & \frac{\partial\Phi_{(2)}}{\partial\theta}\Big|_{r=b} &= \frac{\partial\Phi_{(3)}}{\partial\theta}\Big|_{r=b}\,.
\end{aligned} \tag{8.142}$$

Note that, as in the analogous discussion given previously in the electrostatic case, the tangential matching conditions, involving $\dfrac{\partial\Phi}{\partial\theta}$, are actually equivalent to the conditions that the magnetic scalar potential itself should be continuous at the boundaries. In practice it is always simpler just to match Φ rather than $\dfrac{\partial\Phi}{\partial\theta}$ at the boundaries. Thus the boundary conditions can be written more simply as

$$\begin{aligned}
\mu\, \frac{\partial\Phi_{(2)}}{\partial r}\Big|_{r=a} &= \frac{\partial\Phi_{(1)}}{\partial r}\Big|_{r=a}\,, \qquad & \Phi_{(2)}(a,\theta) &= \Phi_{(1)}(a,\theta)\,, \\
\mu\, \frac{\partial\Phi_{(2)}}{\partial r}\Big|_{r=b} &= \frac{\partial\Phi_{(3)}}{\partial r}\Big|_{r=b}\,, \qquad & \Phi_{(2)}(b,\theta) &= \Phi_{(3)}(b,\theta)\,,
\end{aligned} \tag{8.143}$$

Plugging the expansions (8.141) into (8.143), we obtain four equations for the four unknowns α, β, γ and δ, namely

$$\begin{aligned}
\mu\,\beta - 2\mu\,\gamma\,a^{-3} - \alpha &= 0\,,\\
\beta\,a + \gamma\,a^{-2} - \alpha\,a &= \ ,\\
\mu\,\beta\ - 2\mu\,b^{-3} + 2\delta\,b^{-3} &= -B_0\,,\\
\beta\,b + \gamma\,b^{-2} - \delta\,b^{-2} &= -b\,B_0\,.
\end{aligned} \tag{8.144}$$

Solving these equations, we find

$$\begin{aligned}
\alpha &= -\frac{9\mu}{X}\,B_0\,,\\
\beta &= -\frac{3(2\mu+1)\,a^3}{X}\,B_0\,,\\
\gamma &= -\frac{3(\mu-1)\,a^3}{X}\,B_0\,,\\
\delta &= \frac{(2\mu+1)(\mu-1)(b^3-a^3)}{X}\,B_0\,,
\end{aligned} \tag{8.145}$$

where

$$X \equiv (\mu+2)(2\mu+1) - 2a^3\,b^{-3}\,(\mu-1)^2\,. \tag{8.146}$$

The magnetic field inside the shell is given by the potential

$$\Phi_{(1)} = \alpha\,r\,\cos\theta = -\frac{9\mu}{X}\,B_0\,r\,\cos\theta = -\frac{9\mu}{X}\,B_0\,z\,, \tag{8.147}$$

and so inside the shell the magnetic field is uniform, and given by

$$\vec{B} = \frac{9\mu}{X}\,\vec{B}_0\,. \tag{8.148}$$

If $\mu >> 1$, as is easily possible for a high permeability material, which might have $\mu \sim 10^6$, we see that

$$\alpha \approx -\frac{9}{2\mu(1-a^3\,b^{-3})}\,B_0\,, \tag{8.149}$$

and so

$$\vec{B} \approx \frac{9}{2\mu(1-a^3\,b^{-3})}\,\vec{B}_0\,, \qquad r < a\,. \tag{8.150}$$

This shows that even with a fairly thin shell of the permeable medium, a high degree of magnetic shielding can be achieved.

Outside the shell, the magnetic field is the sum of the uniform field $\vec{B}_0$ plus a magnetic dipole term of magnitude δ.

8.7.3 Magnetic scalar potential for hard ferromagnets

Another circumstance where we can make use of the magnetic scalar potential is in the case of a so-called *hard ferromagnet*. Such a ferromagnet is commonly also called a *permanent magnet*, namely a ferromagnetic material that has its own intrinsic magnetisation, even in the absence of externally-applied magnetic fields. In fact, to a good approximation, the magnetisation inside the hard ferromagnet is independent of any external magnetic field. Thus a common situation is to have a hard ferromagnet, with known magnetisation $\vec{M}$, immersed in free space. One might also include externally-applied magnetic fields.

Let us consider such a situations where there are free currents, and so we have the Maxwell equations

$$\vec{\nabla}\cdot\vec{B} = 0\,, \qquad \vec{\nabla}\times\vec{H} = 0\,, \tag{8.151}$$

where, as always,

$$\vec{H} = \vec{B} - 4\pi\,\vec{M}\,. \tag{8.152}$$

We solve the second equation in (8.151) by writing

$$\vec{H} = -\vec{\nabla}\Phi_M\,. \tag{8.153}$$

The first equation becomes, upon using (8.152),

$$\vec{\nabla}\cdot\vec{H} = -4\pi\,\vec{\nabla}\cdot\vec{M}\,, \tag{8.154}$$

and hence

$$\nabla^2\Phi_M = 4\pi\,\vec{\nabla}\cdot\vec{M}\,. \tag{8.155}$$

This may be solved by writing

$$\Phi_M(\vec{r}) = -\int \frac{\vec{\nabla}'\cdot\vec{M}(\vec{r}\,')}{|\vec{r}-\vec{r}\,'|}\,d^3\vec{r}\,'\,. \tag{8.156}$$

The magnetisation $\vec{M}$ will be non-zero only inside the ferromagnet.

Let us consider, as an example, a uniformly magnetised sphere of ferromagnetic material in free space, with no externally-appplied magnetic fields. We shall assume that inside the sphere, of radius a, there is a uniform magnetisation $\vec{M}_0$; i.e. the Cartesian components of $\vec{M}_0$ are constant. It will be convenient presently to assume that axes are chosen so that $\vec{M}_0$ is aligned along the z direction, and that the sphere is centred on the origin.

Naively, one might think that since the magnetisation is piecewise constant in this problem (equal to $\vec{M}_0$ inside the sphere, and equal to zero outside the sphere), then $\vec{\nabla}\cdot\vec{M}$ would be zero and hence (8.156) would be zero. (A manifestly absurd conclusion, since

obviously the magnet will produce a non-vanishing magnetic field outside the sphere.) The flaw in the naive argument is that the magnetisation undergoes a discontinuous jump at the surface of the sphere, and so $\vec{\nabla}\cdot\vec{M}$ will be equal to a delta-function at $r=a$.

One way to calculate (8.156) is to note, using the vector identity $\vec{\nabla}\cdot(f\,\vec{M}) = \vec{M}\cdot\vec{\nabla}f + f\,\vec{\nabla}\cdot\vec{M}$, that Φ_M can be written as

$$\Phi_M(\vec{r}) = -\int \vec{\nabla}'\cdot\Big[\frac{\vec{M}(\vec{r}\,')}{|\vec{r}-\vec{r}\,'|}\Big]\,d^3\vec{r}\,' + \int \vec{M}(\vec{r}\,')\cdot\vec{\nabla}'\Big[\frac{1}{|\vec{r}-\vec{r}\,'|}\Big]\,d^3\vec{r}\,'\,. \tag{8.157}$$

Using the divergence theorem, the first term can be turned into a surface integral over the sphere at infinity, and therefore it gives zero since $\vec{M}=0$ outside the ferromagnet. For the second term, we use the fact that

$$\vec{\nabla}'\Big[\frac{1}{|\vec{r}-\vec{r}\,'|}\Big] = -\vec{\nabla}\Big[\frac{1}{|\vec{r}-\vec{r}\,'|}\Big]\,, \tag{8.158}$$

and so we have

$$\begin{aligned}\Phi_M(\vec{r}) &= -\int \vec{M}(\vec{r}\,')\,\vec{\nabla}\Big[\frac{1}{|\vec{r}-\vec{r}\,'|}\Big]\,d^3\vec{r}\,' = -\int \vec{\nabla}\cdot\Big[\frac{\vec{M}(\vec{r}\,')}{|\vec{r}-\vec{r}\,'|}\Big]\,d^3\vec{r}\,'\,,\\ &= -\vec{\nabla}\cdot\int\frac{\vec{M}(\vec{r}\,')}{|\vec{r}-\vec{r}\,'|}\,d^3\vec{r}\,'\,.\end{aligned} \tag{8.159}$$

(Note that one must, as always, pay careful attention to the distinction between the position vector $\vec{r}$ and the primed position vector (i.e. the integration variable) $\vec{r}\,'$.)

In the problem at hand, the magnetisation $\vec{M}$ is equal to the constant vector $\vec{M}_0$, parallel to the z axis, inside the sphere of radius a. Outside the sphere, we have $\vec{M}=0$. Thus (8.159) becomes

$$\begin{aligned}\Phi_M(\vec{r}) &= \vec{\nabla}\cdot\int_0^a r'^2\,dr'\int d\Omega'\frac{\vec{M}_0}{|\vec{r}-\vec{r}\,'|}\,,\\ &= -\vec{M}_0\cdot\vec{\nabla}\int_0^a r'^2\,dr'\int d\Omega'\frac{1}{|\vec{r}-\vec{r}\,'|}\,,\end{aligned} \tag{8.160}$$

where $d\Omega' = \sin\theta'\,d\theta'\,d\varphi'$. Now recall from equation (4.154) that we can write

$$\frac{1}{|\vec{r}-\vec{r}\,'|} = \sum_{\ell\ge0}\sum_{m=-\ell}^{\ell}\frac{4\pi}{2\ell+1}\,\frac{r_<^\ell}{r_>^{\ell+1}}\,\bar{Y}_{\ell m}(\theta',\varphi')\,Y_{\ell m}(\theta,\varphi)\,. \tag{8.161}$$

Thus we have

$$\int\frac{d\Omega'}{|\vec{r}-\vec{r}\,'|} = \sum_{\ell\ge0}\sum_{m=-\ell}^{\ell}\frac{4\pi}{2\ell+1}\,\frac{r_<^\ell}{r_>^{\ell+1}}\,Y_{\ell m}(\theta,\varphi)\int d\Omega'\,\bar{Y}_{\ell m}(\theta',\varphi')\,,\,,$$

with $r_>$ being the greater of r and r'.

Since

$$\int d\Omega\, \bar{Y}_{\ell' m'}(\theta\,\varphi)\, Y_{\ell m}(\theta,\varphi) = \delta_{\ell\ell'}\,\delta_{mm'}\,, \tag{8.162}$$

and $Y_{0,0}(\theta,\varphi) = 1/\sqrt{4\pi}$, it follows that

$$\int d\Omega'\, \bar{Y}_{\ell m}(\theta',\varphi') = \sqrt{4\pi}\int d\Omega'\, \bar{Y}_{\ell' m'}(\theta\,\varphi)\, Y_{0,0}(\theta',\varphi') = \sqrt{4\pi}\,\delta_{\ell,0}\,\delta_{m,0}\,, \tag{8.163}$$

and so[54]

$$\int \frac{d\Omega'}{|\vec{r}-\vec{r}\,'|} = \frac{(4\pi)^{3/2}}{r_>}\, Y_{0,0}(\theta,\varphi) = \frac{4\pi}{r_>}\,. \tag{8.164}$$

Equation (8.160) therefore becomes

$$\Phi_M(\vec{r}) = -4\pi\,\vec{M}_0\cdot\vec{\nabla}\int_0^a \frac{r'^2\,dr'}{r_>}\,. \tag{8.165}$$

At points outside the sphere, we have $r > a$. On the other hand, the integration over r' runs only up to $r' = a$. Therefore in the exterior region we have $r_> = r$, and so (8.165) gives

$$\Phi_M(\vec{r}) = -\frac{4\pi a^3}{3}\,\vec{M}_0\cdot\vec{\nabla}\frac{1}{r}\,, \qquad (r > a)\,. \tag{8.166}$$

When r is instead inside the sphere, i.e. for points with $0 < r < a$, then $r_>$ is equal to r for $0 \le r' < r$, while $r_>$ is equal to r' for $r < r' < a$. In this case we have

$$\int_0^a \frac{r'^2\,dr'}{r_>} = \frac{1}{r}\int_0^r r'^2\,dr' + \int_r^a r'\,dr' = -\tfrac{1}{6}r^2 + \tfrac{1}{2}a^2\,, \tag{8.167}$$

and so the magnetic scalar potential inside the sphere is given by

$$\Phi_M(\vec{r}) = \frac{2\pi}{3}\,\vec{M}_0\cdot\vec{\nabla}r^2\,, \qquad (r < a)\,. \tag{8.168}$$

Since we are taking $\vec{M}_0$ to be parallel to the z axis, and since $\vec{\nabla}r = \vec{r}/r$, we therefore have

$$\begin{aligned}
\Phi_M &= \frac{4\pi a^3\, M_0}{3}\,\frac{\cos\theta}{r^2}\,, \qquad (r > a)\,,\\
\Phi_M &= \frac{4\pi\, M_0}{3}\, r\,\cos\theta = \frac{4\pi\, M_0}{3}\, z\,, \qquad (r < a)\,.
\end{aligned} \tag{8.169}$$

We see that outside the sphere, where $\mu = 1$, the magnetic field is given by $\vec{H} = \vec{B} = -\vec{\nabla}\Phi_M$, and it is a magnetic dipole field with magnetic moment

$$\vec{m} = \frac{4\pi a^3}{3}\,\vec{M}_0\,. \tag{8.170}$$

[54]Using the rather involved formula (8.161) in order to arrive eventually at the very simple expression (8.162) may seem like taking a sledge hammer to crack a nut, but since the sledge hammer is available, it is as good a method as any other for the purpose!

The magnetic dipole moment m should be given by the volume integral of the magnetisation, and since this is just the constant $\vec{M}_0$ inside the sphere we just have to multiply by the volume of the sphere, $\dfrac{4\pi a^3}{3}$, and this indeed gives the expression in eqn (8.170). Inside sphere, we have $\vec{H} = -\vec{\nabla}\Phi_M = -\vec{\nabla}(4\pi\, M_0\, z/3)$ and so

$$\vec{H} = -\frac{4\pi}{3}\,\vec{M}_0\,. \tag{8.171}$$

Using (8.152), we then have

$$\vec{B} = \frac{8\pi}{3}\,\vec{M}_0\,. \tag{8.172}$$

8.7.4 Magnetised sphere in an external field

As an elaboration of the discussion above of a magnetised sphere, we can superpose on it a uniform magnetic induction $\vec{B}_0$ throughout all space. This means that inside the sphere, the expressions (8.171) and (8.172) will be modified to become

$$\vec{B} = \vec{B}_0 + \frac{8\pi}{3}\,\vec{M}_0\,, \qquad \vec{H} = \vec{B}_0 - \frac{4\pi}{3}\,\vec{M}_0\,. \tag{8.173}$$

One can now replace the original problem, in which the sphere was composed of a permanently magnetised ferromagnetic material, by a problem in which the sphere is instead a permeable medium of magnetic permeability μ. This means that $\vec{B}$ and $\vec{H}$ in (8.173) will now be related by $\vec{B} = \mu\,\vec{H}$. This allows us to solve for the magnetisation $\vec{M}_0$, finding

$$\vec{M}_0 = \frac{3(\mu-1)}{4\pi(\mu+2)}\,\vec{B}_0\,. \tag{8.174}$$

This result is a direct magnetic analogue of the expression (7.70) that we found when calculating the polarisation of a dielectric sphere in an asymptotically uniform electric field. It is easy to check that the expression (8.174) is in agreement with the results we obtained earlier for a spherical shell of permeable medium in an asymptotically uniform magnetic field, if we send the inner radius a to zero.

9 Electromagnetism and Quantum Mechanics

9.1 The Schrödinger equation and gauge transformations

The Schrödinger equation for a particle of mass m and charge e in an electromagnetic field is

$$-\frac{\hbar^2}{2m}\Big(\vec{\nabla} - \frac{\mathrm{i}\,e}{\hbar c}\,\vec{A}\Big)^2\,\psi + e\phi\,\psi = \mathrm{i}\,\hbar\,\frac{\partial\psi}{\partial t}\,. \tag{9.1}$$

The Schrödinger equation (9.1) is written in terms of the scalar and vector potentials ϕ and $\vec{A}$ that describe the electromagnetic field. Thus, if we perform a gauge transformation

$$\vec{A} \longrightarrow \vec{A}' = \vec{A} + \vec{\nabla}\lambda\,, \qquad \phi \longrightarrow \phi' = \phi - \frac{1}{c}\,\frac{\partial\lambda}{\partial t}\,, \tag{9.2}$$

the Schrödinger equation will change its form. On the other hand, we expect that the physics should be unaltered by a mere gauge transformation, since this leaves the physically-observable electric and magnetic fields unchanged. It turns out that we should simultaneously perform a very specific phase transformation on the wavefunction ψ,

$$\psi \longrightarrow \psi' = e^{\mathrm{i}\,e\lambda/(\hbar c)}\,\psi \tag{9.3}$$

then the Schrödinger equation expressed entirely in terms of the primed quantities (i.e. wavefunction ψ' and electromagnetic potentials ϕ' and $\vec{A}'$) will take the identical form to the original unprimed equation (9.1). Thus, we may say that the Schrödinger equation transforms *covariantly* under gauge transformations.

To see the details of how this works, it is useful first to define what are called *covariant derivatives*. We do this both for the three spatial derivatives, and also for the time derivative. Thus we define

$$D_i \equiv \partial_i - \frac{\mathrm{i}\,e}{\hbar c}\,A_i\,, \qquad D_0 \equiv \frac{\partial}{\partial t} + \frac{\mathrm{i}\,e}{\hbar}\,\phi\,. \tag{9.4}$$

Note that the original Schrödinger equation (9.1) is now written simply as

$$-\frac{\hbar^2}{2m}\,D_iD_i\psi - \mathrm{i}\,\hbar\,D_0\psi = 0\,. \tag{9.5}$$

Next, perform the transformations

$$\begin{aligned} \vec{A} \;&\longrightarrow\; \vec{A}' = \vec{A} + \vec{\nabla}\lambda\,, \qquad \phi \longrightarrow \phi' = \phi - \frac{1}{c}\,\frac{\partial\lambda}{\partial t}\,, \\ \psi \;&\longrightarrow\; \psi' = e^{\mathrm{i}\,e\lambda/(\hbar c)}\,\psi \end{aligned} \tag{9.6}$$

The crucial point about this is that we have the following:

$$\begin{aligned} D'_i\,\psi' \;&\equiv\; \Big(\partial_i - \frac{\mathrm{i}\,e}{\hbar c}\,A'_i\Big)\,\psi' = \Big(\partial_i - \frac{\mathrm{i}\,e}{\hbar c}\,A_i - \frac{\mathrm{i}\,e}{\hbar c}\,(\partial_i\lambda)\Big)\,\Big(e^{\mathrm{i}\,e\lambda/(\hbar c)}\,\psi\Big)\,, \\ &=\; e^{\mathrm{i}\,e\lambda/(\hbar c)}\,\Big(\partial_i - \frac{\mathrm{i}\,e}{\hbar c}\,A_i - \frac{\mathrm{i}\,e}{\hbar c}\,(\partial_i\lambda) + \frac{\mathrm{i}\,e}{\hbar c}\,(\partial_i\lambda)\Big)\,\psi\,, \\ &=\; e^{\mathrm{i}\,e\lambda/(\hbar c)}\,\Big(\partial_i - \frac{\mathrm{i}\,e}{\hbar c}\,A_i\Big)\,\psi\,, \end{aligned} \tag{9.7}$$

and

$$\begin{aligned} D'_0\,\psi' \;&\equiv\; \Big(\frac{\partial}{\partial t} + \frac{\mathrm{i}\,e}{\hbar}\,\phi'\Big)\,\psi' = \Big(\frac{\partial}{\partial t} + \frac{\mathrm{i}\,e}{\hbar}\,\phi - \frac{\mathrm{i}\,e}{\hbar c}\,\frac{\partial\lambda}{\partial t}\Big)\,\Big(e^{\mathrm{i}\,e\lambda/(\hbar c)}\,\psi\Big)\,, \\ &=\; e^{\mathrm{i}\,e\lambda/(\hbar c)}\,\Big(\frac{\partial}{\partial t} + \frac{\mathrm{i}\,e}{\hbar}\,\phi - \frac{\mathrm{i}\,e}{\hbar c}\,\frac{\partial\lambda}{\partial t} + \frac{\mathrm{i}\,e}{\hbar c}\,\frac{\partial\lambda}{\partial t}\Big)\,\psi\,, \\ &=\; e^{\mathrm{i}\,e\lambda/(\hbar c)}\,\Big(\frac{\partial}{\partial t} + \frac{\mathrm{i}\,e}{\hbar}\,\phi\Big)\,\psi\,. \end{aligned} \tag{9.8}$$

In other words, we have

$$D'_i\psi' = e^{\mathrm{i}\,e\lambda/(\hbar c)}\,D_i\psi\,,\qquad D'_0\psi' = e^{\mathrm{i}\,e\lambda/(\hbar c)}\,D_0\psi\,. \tag{9.9}$$

This means that $D_i\psi$ and $D_0\psi$ transform the same way as ψ itself under the gauge transformations (9.6), namely just by acquiring the phase factor $e^{\mathrm{i}\,e\lambda/\hbar}$. This is a non-trivial statement, since the gauge parameter λ is an arbitrary function of space and time. Had we been considering standard partial derivatives ∂_i and $\partial/\partial t$ rather than the covariant deriavtives defined in (9.4), it would most certainly not have been true. For example,

$$\partial_i\psi' = \partial_i\Big(e^{\mathrm{i}\,e\lambda/(\hbar c)}\,\psi\Big) = e^{\mathrm{i}\,e\lambda/(\hbar c)}\,\partial_i\psi + e^{\mathrm{i}\,e\lambda/(\hbar c)}\,\frac{\mathrm{i}\,e}{\hbar c}\,(\partial_i\lambda)\,\psi \neq e^{\mathrm{i}\,e\lambda/(\hbar c)}\,\partial_i\psi\,, \tag{9.10}$$

precisely because the derivative can land on the space-time dependent gauge-transformation parameter λ and thus give the second term, which spoils the covariance of the transformation. The point about the covariant derivatives is that the contributions from the gauge transformation of the gauge potentials precisely cancels the "unwanted" second term in (9.10).

By iterating the calculation, it also follows that $D'_iD'_i\psi' = e^{\mathrm{i}\,e\lambda/\hbar}\,D_iD_i\psi$, and so we see that the Schrödinger equation (9.5) written in terms of the primed fields, i.e.

$$-\frac{\hbar^2}{2m}\,D'_iD'_i\psi' - \mathrm{i}\,\hbar\,D'_0\psi' = 0\,, \tag{9.11}$$

just implies the Schrödinger equation in terms of unprimed fields, since

$$\begin{aligned} 0 &= -\frac{\hbar^2}{2m}\,D'_iD'_i\psi' - \mathrm{i}\,\hbar\,D'_0\psi'\,,\\ &= e^{\mathrm{i}\,e\lambda/(\hbar c)}\,\Big(-\frac{\hbar^2}{2m}\,D_iD_i\psi - \mathrm{i}\,\hbar\,D_0\psi\Big)\,. \end{aligned} \tag{9.12}$$

What we have proved above is that the Schrödinger equation transforms covariantly under electromagnetic gauge transformations, provided that at the same time the wave function is scaled by a space-time dependent phase factor, as in (9.6).

9.2 Magnetic monopoles

Particles with magnetic charge, known as magnetic monopoles, have never been seen in nature. However, there seems to be no reason in principle why they should not exist, and it is of interest to explore their properties in a little more detail. A point electric charge e has an electric field given by

$$\vec{E} = \frac{e\,\vec{r}}{r^3}\,. \tag{9.13}$$

Thus by analogy, a point magnetic monopole, with magnetic charge g, will have a magnetic field given by

$$\vec{B} = \frac{g\,\vec{r}}{r^3}\,. \tag{9.14}$$

This satisfies

$$\vec{\nabla}\cdot\vec{B} = 4\pi\,\rho_M\,, \qquad \rho_M = g\,\delta^3(\vec{r})\,, \tag{9.15}$$

where $\rho_M = J^0_M$ is the magnetic charge density. (Note that if magnetic monompoles did exist, the Maxwell equation $\vec{\nabla}\cdot\vec{B} = 0$ would have to change, to allow for the magnetic charge density ρ_M on the right-hand side.)

We shall be interested in studying the quantum mechanics of electrically-charged particles in the background of a magnetic monopole. Since the Schrödinger equation is written in terms of the potentials ϕ and $\vec{A}$, we shall therefore need to write down the 3-vector potential $\vec{A}$ for the magnetic monopole. To do this, we introduce Cartesian coordinates (x, y, z), related to spherical polar coordinates (r, θ, φ) in the standard way,

$$x = r\sin\theta\,\cos\varphi\,, \quad y = r\sin\theta\,\sin\varphi\,, \quad x = r\cos\theta\,, \tag{9.16}$$

and we also define

$$\rho^2 = x^2 + y^2\,. \tag{9.17}$$

Consider the 3-vector potential

$$\vec{A} = (A_x, A_y, A_z) = \Big(\frac{g\,zy}{r\rho^2}, -\frac{g\,zx}{r\rho^2}, 0\Big)\,. \tag{9.18}$$

Using

$$\begin{aligned}
\frac{\partial r}{\partial x} &= \frac{x}{r}\,, \qquad \frac{\partial r}{\partial y} = \frac{y}{r}\,, \qquad \frac{\partial r}{\partial z} = \frac{z}{r}\,,\\
\frac{\partial \rho}{\partial x} &= \frac{x}{\rho}\,, \qquad \frac{\partial \rho}{\partial y} = \frac{y}{\rho}\,, \qquad \frac{\partial \rho}{\partial z} = 0\,,
\end{aligned} \tag{9.19}$$

it is easily seen that

$$B_x = \partial_y A_z - \partial_z A_y = g\partial_z\Big(\frac{zx}{r\rho^2}\Big) = \frac{gx}{r\rho^2} - \frac{gxz^2}{r^3\rho^2} = \frac{gx}{r^3}\,, \tag{9.20}$$

and similarly

$$B_y = \frac{gy}{r^3}\,, \qquad B_z = \frac{gz}{r^3}\,. \tag{9.21}$$

Thus indeed we find that

$$\vec{\nabla}\times\vec{A} = \frac{g\,\vec{r}}{r^3}\,, \tag{9.22}$$

and so the 3-vector potential (9.18) describes the magnetic monopole field (9.14).

In terms of spherical polar coordinates we have $\rho^2 = x^2 + y^2 = r^2 \sin^2\theta$, and so (9.18) can be written as

$$\vec{A} = \frac{g\cot\theta}{r}\,(\sin\varphi, -\cos\varphi, 0)\,. \tag{9.23}$$

(Note that we are still writing $\vec{A}$ as a Cartesian vector here, with components A_x, A_y and A_z, but we are writing the coordinates x, y and z in terms of the spherical polar coordinates r, θ and φ.) Not surprisingly, this potential is singular at $r = 0$, since we are describing an idealised point magnetic charge. In exactly the same way, the potential $\phi = e/r$ describing a point electric charge diverges at $r = 0$ also. However, the potential (9.23) also diverges everywhere along the z axis, i.e. at $\theta = 0$ and $\theta = \pi$. It turns out that these latter singularities are "unphysical," in the sense that they can be removed by making gauge transformations. This is not too surprising, when we note that the magnetic field itself, given by (9.14) has no singularity along the z axis. It is, of course, genuinely divergent at $r = 0$, so that is a real physical singularity.

To see the unphysical nature of the singularities in (9.23) along $\theta = 0$ and $\theta = \pi$, we need to make gauge transformations, under which

$$\vec{A} \longrightarrow \vec{A} + \vec{\nabla}\lambda\,. \tag{9.24}$$

Consider first taking

$$\lambda = g\,\varphi = g\,\arctan\frac{y}{x}\,. \tag{9.25}$$

From this, we find

$$\vec{\nabla}\lambda = -\frac{g}{r}\,\mathrm{cosec}\theta\,(\sin\varphi, -\cos\varphi, 0)\,. \tag{9.26}$$

Letting the gauge-transformed potential be $\vec{A}'$, we therefore find

$$\vec{A}' = \vec{A} + \vec{\nabla}\lambda = \frac{g}{r}\,\frac{\cos\theta - 1}{\sin\theta}\,(\sin\varphi, -\cos\varphi, 0) = -\frac{g}{r}\,\tan\tfrac{1}{2}\theta\,(\sin\varphi, -\cos\varphi, 0)\,. \tag{9.27}$$

It can be seen that $\vec{A}$ is completely non-singular along $\theta = 0$ (i.e. along the positive z axis). It is, however, singular along $\theta = \pi$ (i.e. along the negative z axis).

We could, on the other hand, perform a gauge transformation with λ given by

$$\lambda = -g\,\varphi = -g\,\arctan\frac{y}{x} \tag{9.28}$$

instead of (9.25). Defining the gauge-transformed potential as $\vec{A}''$ in this case, we find

$$\vec{A}'' = \frac{g}{r}\,\frac{\cos\theta + 1}{\sin\theta}\,(\sin\varphi, -\cos\varphi, 0) = \frac{g}{r}\,\cot\tfrac{1}{2}\theta\,(\sin\varphi, -\cos\varphi, 0)\,. \tag{9.29}$$

This time, we have obtained a gauge potential that is non-singular along $\theta = \pi$ (i.e. the negative z axis), but it is singular along $\theta = 0$ (the positive z axis).

There is no single choice of gauge in which the 3-vector potential for the magnetic monopole is completely free of singularities away from the origin $r = 0$. We have obtained two expressions for the vector potential, one of which, $\vec{A}'$, is non-singular along the positive z axis, and the other, $\vec{A}''$, is non-singular along the negative z axis. The singularity that each has is known as the Dirac "string singularity," since it lies along a line, or string. By making gauge transformations the location of the string can be moved around, but it can never be removed altogether.

In the discussion above, the z axis appears to have played a preferred rôle, but this is, of course, just an artefact of our gauge choices. We could equally well have chosen a different expression for $\vec{A}$, related by a gauge transformation, for which the string singularity ran along any desired line, or curve, emanating from the origin.

9.3 Dirac quantisation condition

We have seen that gauge potentials for the magnetic monopole, free of singularities on the positive and negative z axes resepctively, are given by

$$\begin{aligned} \vec{A}' &= -\frac{g}{r}\,\tan\tfrac{1}{2}\theta\,(\sin\varphi, -\cos\varphi, 0)\,, \\ \vec{A}'' &= \frac{g}{r}\,\cot\tfrac{1}{2}\theta\,(\sin\varphi, -\cos\varphi, 0)\,. \end{aligned} \tag{9.30}$$

The two are themselves related by a gauge transformation, namely

$$\vec{A}'' = \vec{A}' + \vec{\nabla}(-2g\varphi)\,. \tag{9.31}$$

Now let us consider the quantum mechanics of an electron in the background of the magnetic monopole. As we discussed in section 9.1, the Schrödinger equation for the electron is given by (9.1), where e is its charge, and m is its mass. We shall consider the Schrödinger equation in two different gauges, related as in (9.31). Denoting the corresponding electron wave-functions by ψ' and ψ'', we see from (9.6) and (9.31) that we shall have

$$\psi'' = e^{-2\mathrm{i}\,eg\varphi/\hbar}\,\psi'\,. \tag{9.32}$$

However, we have seen that the gauge transformation is not physical, but merely corresponds to shifting the string singularity of the magnetic monopole from the negative z axis to the positive z axis. Quantum mechanically, the physics will only be unchanged if the electron wave-function remains single valued under a complete 2π rotation around the z axis. This means that the phase factor in the relation (9.32) must be equal to unity, and so it must be that

$$\frac{2eg}{\hbar}\,2\pi = 2\pi\,n\,, \tag{9.33}$$

where n is an integer. Thus it must be that the product of the electric charge e on the electron, and the magnetic charge g on the magnetic monopole, must satisfy the so-called *Dirac quantisation condition*,

$$2e\,g = n\,\hbar\,. \tag{9.34}$$

It is interesting to note that although a magnetic monopole has never been observed, it would only take the existence of a single monopole, maybe somewhere in another galaxy, to imply that electric charges everywhere in the universe must quantised in units of

$$\frac{\hbar}{2g}\,, \tag{9.35}$$

where g is the magnetic charge of the lonely magnetic monopole. In fact all observed electric charges are indeed quantised; in integer multiples of the charge e on the electron, in everyday life, and in units of $\frac{1}{3}e$ in the quarks of the theory of strong interactions. It is tempting to speculate that the reason for this electric charge quantisation may be the existence of a magnetic monopole somewhere out in the vastness of space, in a galaxy far far away.