Ph106bc: Electrodynamics

Sunil Golwala

Winter/Spring 2024

Contents

Contents

Contents

Contents

Contents

Lectures

Ph106b

Lecture 1	2024/01/04
Lecture 2	2024/01/09
Lecture 3	2024/01/11
Lecture 4	2024/01/16
Lecture 5	2024/01/18
Lecture 6	2024/01/23
Lecture 7	2024/01/25
Lecture 8	2024/01/30
Lecture 9	2024/02/01
Lecture 10	2024/02/06
Lecture 11	2024/02/08
Lecture 12	2024/02/13
Lecture 13	2024/02/15
Lecture 14	2024/02/22
Lecture 15	2024/02/25
Lecture 16	2024/02/27
Lecture 17	2024/02/29
Lecture 18	2024/03/05
Lecture 19	2024/03/07

Ph106c

Lecture 20	2024/04/02
Lecture 21	2024/04/04
Lecture 22	2024/04/09
Lecture 23	2024/04/11
Lecture 24	2024/04/16
Lecture 25	2024/04/18
Lecture 26	2024/04/23
Lecture 27	2024/04/25
Lecture 28	2024/04/30
Lecture 29	2024/05/02
Lecture 30	2024/05/07
Lecture 31	2024/05/09
Lecture 32	2024/05/14
Lecture 33	2024/05/16
Lecture 34	2024/05/21
Lecture 35	2024/05/23
Lecture 36	2024/05/28
Lecture 37	2024/05/30
Lecture 38	2024/06/04
Lecture 39	N/A

Lecture 1:

Introduction to Course

Basics of Electrostatics I:
Electric Force and Field
Dirac Delta Function
Integral Form of Gauss's Law
Differential Form of Gauss's Law

Date Revised: 2024/01/09 05:45
Revised lecture break,
remove misstatement about units of electric field
Date Given: 2024/01/04

Section 1
Introduction to Course

1.1 Course Material
1.2 Notation; including Deviations from Griffiths

Course Material

Overview

This is a course on electrodynamics. It will review the basic material you learned in Ph1bc but will go beyond in both content as well as in mathematical sophistication.

The intended learning outcome of both Ph106b and Ph106c is for students to acquire the ability to calculate electric and magnetic potentials, fields, energies, and forces in a variety of basic physical configurations combined with an understanding of the underlying physical principles and calculation techniques. This outcome requires both an understanding of principles as well as the ability to apply them to do calculations!

The course will primarily use and follow *Introduction to Electrodynamics* by Griffiths (fourth edition). Supplementary material is drawn from Jackson and from Heald & Marion, both on electronic and physical reserve from the library. The material presented here will be self-contained, but past students have found it useful to obtain a copy of Jackson. It is certainly a book you will want if you continue in physics or a related field.

Prerequisites

Physics:

- Electricity and Magnetism: While Ph1bc is a formal prerequisite for the course, we will develop the material from scratch. However, review material will be covered quickly and a basic familiarity with the concepts will be assumed.
- Classical mechanics: Generally, mechanics at the level of Ph1a is sufficient for this course, though some optional material at the end of Ph106c will make use of Ph106a material.

Mathematics:

- Chapter 1 of Griffiths except for Sections 1.1.5 ("How Vectors Transform") and 1.5 ("The Dirac Delta Function"). We will review some prerequisite material as needed.
- Solutions to second-order linear ordinary differential equations with constant coefficients (*i.e.*, simple harmonic oscillator).
- Orthonormal functions/bases.
- Over the course, we will develop the following more sophisticated concepts:
  - Dirac Delta function.
  - Separation of variables to reduce second-order linear partial differential equations to ordinary differential equations.
  - Various specific types of orthonormal functions, specifically sinusoids, Legendre polynomials, and spherical harmonics.
  - Tensor formalism for relativity.
- Key point: Mathematics is the language of physics. You must be competent in the above basic mathematics to understand and use the material in this course. Intuition is crucial, but it must be formalized mathematically.

  However, mathematics is not just symbolic manipulation or brute force calculation. Make sure you understand the **meaning** *of every mathematical expression — i.e., carry along the intuition with the symbols! Only do algebra and explicit differentiation and integration as a last resort! We will demonstrate this approach regularly.*

Topics to be Covered

New topics for Ph106b not covered in Ph1bc

New topics for Ph106c not covered in Ph1bc

- Review of basic electrostatics — Coulomb's Law; Gauss's Law; electric field, potential, and potential energy; conductors, capacitors, and capacitance matrix.
- Advanced electrostatics — boundary value problems (BVP) for determining potentials and fields; Green Functions for BVP; multipole expansion of potential.
- Electrostatics in Matter — polarization, susceptibility, permittivity of matter; BVP with polarizable materials, energy and forces in matter.
- Magnetostatics — Lorentz force; Biot-Savart Law; Ampère's Law; vector potential; boundary conditions; multipole expansion of potential.
- Magnetostatics in Matter — magnetization, susceptibility, and permeability of matter; boundary conditions; ferromagnetism; BVP with magnetizable materials.
- Electrodynamics — electromotive force and electromagnetic induction; inductance and energy in magnetic fields; Maxwell's equations in vacuum and in matter; boundary conditions for Maxwell's equations.
- Conservation Laws — Continuity equation; Poynting's Theorem; electrodynamic momentum and energy.
- Electromagnetic Waves — in vacuum, in polarizable/magnetizable matter, in conductors, in transmission lines and waveguides.

- Potentials and Radiation — potential formulation; fields and potentials of moving point charges; radiated electromagnetic waves; antennas.
- Relativity and Electrodynamics — review of special relativity including relativistic kinematics and collisions, relativistic tensor notation, transformation of fields, transformation of field tensor, relativistic potentials, relativistic formulation of Maxwell's Equations, relativistic dynamics with EM fields, relativistic conservation theorems.

Notation; including Deviations from Griffiths

- We will use standard black text for material that is covered in lecture, while magenta text will be used for material that is skipped during lecture for which you remain responsible. We will skip material generally when it consists of computation or calculation that is tedious to do on the chalkboard, summarizing the results as necessary. You will need to be able to apply the skipped material as well as the techniques developed in this skipped material.
- Green text will be used to indicate supplementary material for which you will not be responsible.
- Griffiths uses boldface notation to indicate vectors and a script $\vec{r}$ to indicate the difference vector $\vec{r} - \vec{r}'$. In order to better match what can be written by hand, we use $\vec{}$ rather than boldface for vectors and we use $\vec{R}$ for the difference vector.
- Griffiths uses $\vec{r}$ to refer to the position of the test charge Q and $\vec{r}'$ to refer to the position of the source charge q. This seems unnecessarily confusing. We instead use q and $\vec{r}$ for the test charge and q' and $\vec{r}'$ for the source charge.
- Griffiths uses $\delta^3(\vec{r})$ to refer to the delta function in three spatial dimension. We use $\delta(\vec{r})$ for this for reasons that are explained after Equation 2.9.

Section 2
Review of Basics of Electrostatics

Study Guidelines

You have seen all the material in this section before in Ph1b. However, the derivations done there were not as rigorous as they could be because you were simultaneously learning vector calculus. Our goal in this section is to do more rigorous derivations to give you some practice in using the mathematical tools. We won't do any examples in lecture or the notes because they duplicate Ph1b. But you should make sure you are comfortable with the examples in Griffiths Chapter 2.

The Assumed Conditions for Electrostatics

Electrostatics is the study of electric fields, potentials, and forces under two assumptions:

- All electric charges sourcing the electric field are stationary and have been so for a sufficiently long time that all fields are static and thus the electric field can be written in terms of the source charges' current positions.
- The source charges are held fixed and cannot react to the fields from any test charges that may be stationary or moving relative to the source charges.

We will see later that, when charges are moving, it takes time for the information about the position to propagate and thus the fields at a given point depend on the configuration of the charges at earlier times.

Coulomb's Law and the Electric Field

Coulomb's Law, Electrostatic Forces, and Superposition

We begin with two empirical facts:

- **Coulomb's Law**: the empirical fact that the force on a test charge q at position $\vec{r}$ due to a source charge q' at $\vec{r}\,'$ is given by Coulomb's Law:

$$\vec{F} = \frac{1}{4\,\pi\,\epsilon_o}\,\frac{q'\,q}{R^2}\,\widehat{R} \qquad \text{with} \quad \vec{R} \equiv \vec{r} - \vec{r}\,' \tag{2.1}$$

  where $\epsilon_o = 8.85 \times 10^{-12}\ \mathrm{C}^2\ \mathrm{N}^{-1}\ \mathrm{m}^{-2}$ is the *permittivity of free space*. The force points along the line from q' to q as indicated by the sign of the definition of $\vec{R}$. The electric charge is in the units of *Coulombs* (C), which is a fundamental unit that cannot be written in terms of other fundamental units.

  Recall that: we use $\vec{\ }$ rather than boldface to indicate vectors; R where Griffiths uses a script r; and a different convention from Griffiths for the symbols for the two charges and their position vectors.

- **Superposition**: the empirical fact that Coulomb's Law obeys the principle of superposition: the force on a test charge q at $\vec{r}$ due to N charges $\{q'_i\}$ at positions $\{\vec{r}\,'_i\}$ is obtained by summing the individual vector forces:

$$\vec{F} = \sum_{i=1}^{N} \vec{F}_i = \sum_{i=1}^{N} \frac{1}{4\,\pi\,\epsilon_o}\,\frac{q'_i\,q}{R_i^2}\,\widehat{R}_i \qquad \text{with} \quad \vec{R}_i \equiv \vec{r} - \vec{r}\,'_i \tag{2.2}$$

The Electric Field

Given that any test charge q placed at the position $\vec{r}$ feels the same force per unit charge, we are motivated to abstract away the test charge and define what we call the **electric field** at that position $\vec{r}$:

$$\vec{E}(\vec{r}) = \frac{\vec{F}}{q} = \begin{cases} \frac{1}{4\,\pi\,\epsilon_o}\,\frac{q'}{R^2}\,\widehat{R} & \text{for a single source charge } q' \text{ at } \vec{r}\,' \\ \sum_{i=1}^{N}\,\frac{1}{4\,\pi\,\epsilon_o}\,\frac{q_i'}{R_i^2}\,\widehat{R}_i & \text{for } N \text{ source charges } \{q_i'\} \text{ at positions } \{\vec{r}_i\,'\} \end{cases} \tag{2.3}$$

The electric field has units of N/C.

Coulomb's Law for Continuous Charge Distributions

If a charge distribution is continuous, then the natural extension of Coulomb's Law is to *integrate* the electric field or force over the contributions from the infinitesimal charge elements dq at $\vec{r}\,'$:

$$\vec{E}(\vec{r}) = \frac{1}{4\,\pi\,\epsilon_o} \int \frac{1}{[R(\vec{r},\vec{r}\,')]^2}\,\widehat{R}(\vec{r},\vec{r}\,')\,dq(\vec{r}\,') \tag{2.4}$$

where $\vec{R}(\vec{r},\vec{r}\,') = \vec{r} - \vec{r}\,'$ varies with the location $\vec{r}\,'$ of dq as the integral is performed. dq is admittedly undefined here. However, before worrying about that, let us note that the integrand is a *vector* and so this integral requires some care: we must break up $\widehat{R}$ into its components and individually integrate each component. For example, if we use Cartesian coordinates, then $\widehat{R} = \widehat{x}\left(\widehat{R}\cdot\widehat{x}\right) + \widehat{y}\left(\widehat{R}\cdot\widehat{y}\right) + \widehat{z}\left(\widehat{R}\cdot\widehat{z}\right)$, and, since the Cartesian unit vectors do not depend on the location of the infinitesimal charge $dq(\vec{r}\,')$, we may write the integral out as follows (eliding the dependence of $\vec{R}$ on $\vec{r}$ and $\vec{r}\,'$ for brevity):

$$\vec{E}(\vec{r}) = \tag{2.5}$$
$$\frac{1}{4\,\pi\,\epsilon_o}\left[\widehat{x}\int\frac{1}{R^2}\left(\widehat{R}\cdot\widehat{x}\right)dq(\vec{r}\,') + \widehat{y}\int\frac{1}{R^2}\left(\widehat{R}\cdot\widehat{y}\right)dq(\vec{r}\,') + \widehat{z}\int\frac{1}{R^2}\left(\widehat{R}\cdot\widehat{z}\right)dq(\vec{r}\,')\right]$$

which is sum of three integrals with *scalar* integrands.

Now, consider some specific charge distributions:

- **volume charge distribution**:

$$\vec{E}(\vec{r}) = \frac{1}{4\,\pi\,\epsilon_o} \int_{\mathcal{V}} \frac{d\tau' \rho(\vec{r}')}{R^2}\,\widehat{R} \qquad (2.6)$$

with $\rho(\vec{r}')$ having units of C m^{-3}, $\vec{r}'$ running over all points in the volume distribution $\mathcal{V}$, and $d\tau'$ being the differential volume element at $\vec{r}'$ for $\mathcal{V}$

- **surface charge distribution**:

$$\vec{E}(\vec{r}) = \frac{1}{4\,\pi\,\epsilon_o} \int_{\mathcal{S}} \frac{da' \sigma(\vec{r}')}{R^2}\,\widehat{R} \qquad (2.7)$$

with $\sigma(\vec{r}')$ having units of C m^{-2}, $\vec{r}'$ running over all points in the surface distribution $\mathcal{S}$, and da' being the differential area element at $\vec{r}'$ for $\mathcal{S}$

- **line charge distribution**:

$$\vec{E}(\vec{r}) = \frac{1}{4\,\pi\,\epsilon_o}\int_{\mathcal{C}} \frac{d\ell'\,\lambda(\vec{r}\,')}{R^2}\,\widehat{R} \qquad (2.8)$$

with $\lambda(\vec{r}\,')$ having units of C m^{-1}, $\vec{r}\,'$ running over all points in the line distribution $\mathcal{C}$, and $d\ell'$ being the differential length element at $\vec{r}\,'$ for $\mathcal{C}$

Using the Dirac delta function we will define below, one can write the first two as special cases of the latter by using delta functions in the dimensions in which the charge distribution has no extent.

Aside: The Dirac Delta Function

Relating Equation 2.6 to Equation 2.3 offers us both the opportunity to rigorously connect them as well as a chance to introduce the *Dirac delta function*. (Mathematically, it is a *distribution*, not a function, but we will use the standard nomenclature.) The Dirac delta function at $\vec{r}_o$, $\delta(\vec{r} - \vec{r}_o)$, is defined by what it does when it is multiplied against an arbitrary function $f(\vec{r})$ and integrated: For any function $f(\vec{r})$ and any volume $\mathcal{V}$ containing the point $\vec{r}_o$, it holds that

$$\boxed{\int f(\vec{r}\,')\,\delta(\vec{r}\,' - \vec{r}_o)\,d\tau' = \begin{cases} f(\vec{r}_o) & \vec{r}_o \in \mathcal{V} \\ 0 & \vec{r}_o \notin \mathcal{V} \end{cases}} \tag{2.9}$$

In particular, if $f(\vec{r})$ is unity, then the right side of the above integral is unity for $\vec{r}_o \in \mathcal{V}$: the integral of a delta function over the volume containing its $\vec{r}_o$ is 1, and, conversely, the integral of a delta function over any volume not containing its $\vec{r}_o$ vanishes.

Two notes on dimensions and notation:

- In order for the units in the above equation to work out, the delta function above must have units of m^{-3}. The general rule is that the delta function's units are the inverse of those of the differential that its argument says it should be integrated with. In this case, the argument is a vector in 3D space and the differential is the differential volume element $d\tau$, and so the delta function has units of m^{-3}. The units can be subtle, though. If one considers a delta function that picks out a 2D surface in 3D space (*e.g.*, for collapsing an integral of a volume charge density to one of a surface charge density), its argument will be a 3D vector, but it should have units of m^{-1} since it eliminates only one of the three dimensions. (This would be obvious if the surface were a sphere of radius a centered at $\vec{r}_o$ because then one could instead use $\delta(|\vec{r}-\vec{r}_o|-a)$, implying units of m^{-1}. Other surfaces may not be so easily defined.)
- Griffiths refers to the above delta function as $\delta^3(\vec{r}-\vec{r}_o)$. He does this because one can think of this delta function in terms of 1D delta functions

$$\delta^3(\vec{r}-\vec{r}_o) = \delta(x-x_o)\delta(y-y_o)\delta(z-z_o) \quad \text{where} \quad \begin{array}{l} \vec{r} = x\,\hat{x} + y\,\hat{y} + z\,\hat{z} \\ \vec{r}_o = x_o\,\hat{x} + y_o\,\hat{y} + z_o\,\hat{z} \end{array} \tag{2.10}$$

We drop the ^{3} because it is unnecessary: the dimension of the delta function is implied by its argument, the fact that it picks a single point out of 3D space. Moreover, the ^{3} notation is misleading and confusing because it suggests that δ^3 is the cube of something that has $\vec{r}-\vec{r}_o$ as its argument. It is not!

With the above, if we define the charge distribution for a set of point charges $\{q_i'\}$ at positions $\{\vec{r}_i'\}$ to be

$$\rho(\vec{r}) = \sum_{i=1}^{N} q_i' \, \delta(\vec{r} - \vec{r}_i') \tag{2.11}$$

then, when we do the integral in Equation 2.6 over any volume $\mathcal{V}$ containing all N charges, we recover the discrete version of the expression for the electric field, Equation 2.3:

$$\begin{aligned}
\vec{E}(\vec{r}) &= \frac{1}{4\,\pi\,\epsilon_o} \int_{\mathcal{V}} \sum_{i=1}^{N} \frac{d\tau' q_i' \, \delta(\vec{r}' - \vec{r}_i')}{|\vec{r} - \vec{r}'|^2} \, \frac{\vec{r} - \vec{r}'}{|\vec{r} - \vec{r}'|} \\
&= \frac{1}{4\,\pi\,\epsilon_o} \sum_{i=1}^{N} \int_{\mathcal{V}} d\tau' q_i' \, \delta(\vec{r}' - \vec{r}_i') \frac{\vec{r} - \vec{r}'}{|\vec{r} - \vec{r}'|^3} \\
&= \frac{1}{4\,\pi\,\epsilon_o} \sum_{i=1}^{N} q_i' \frac{\vec{r} - \vec{r}_i'}{|\vec{r} - \vec{r}_i'|^3} = \frac{1}{4\,\pi\,\epsilon_o} \sum_{i=1}^{N} \frac{q_i'}{R_i^2} \, \widehat{R}_i
\end{aligned} \tag{2.12}$$

Gauss's Law

Statement of Gauss's Law

The *flux* of the electric field through a surface is the integral of the component of the electric field normal to the surface over the surface:

$$\mathcal{F}_{\mathcal{S}} = \int_{\mathcal{S}} \vec{E} \cdot \hat{n}(\vec{r})\, da \tag{2.13}$$

where $\vec{r}$ lies on the surface $\mathcal{S}$ and $\hat{n}(\vec{r})$ is the surface normal at that point $\vec{r}$. Note that the flux has a sign based on the choice of the orientation of $\hat{n}$.

Gauss's Law relates the flux of the electric field through any *closed* surface to the total charge enclosed by that surface:

$$\boxed{\mathcal{F}_{\mathcal{S}} = \oint_{\mathcal{S}} \vec{E} \cdot \hat{n}(\vec{r})\, da = \frac{1}{\epsilon_o} \int_{\mathcal{V}(\mathcal{S})} d\tau\, \rho(\vec{r})} \tag{2.14}$$

where $\mathcal{V}(\mathcal{S})$ is the surface enclosed by $\mathcal{S}$ and $\oint$ indicates the integral over a closed surface. Our derivation below will take the surface normal direction to be outward from the closed volume.

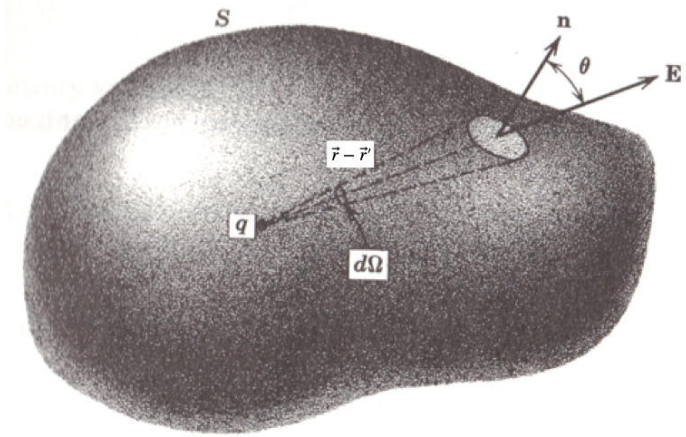
Utility of Gauss's Law

Gauss's Law has three uses:

- For charge distributions having some amount of geometrical symmetry, it provides a way to calculate the electric field that is much easier than brute-force integration of Coulomb's Law.
- We will see that it will enable us to relate the electric field's *boundary conditions* at an interface between two volumes (the conditions relating the electric field components on the two sides of the interface) to the amount of charge at that interface.
- We can obtain a differential version of it, relating spatial derivatives of the electric field to the charge density locally. Doing so directly from Coulomb's Law is difficult (though not impossible, given what we will prove about the divergence of Coulomb's Law!).

Proof of Gauss's Law

The proof offered in Griffiths' is unnecessarily unrigorous; we follow Jackson §1.3.



© 1999 Jackson, *Classical Electrodynamics*

First, consider a charge distribution $\rho(\vec{r})$ that lies completely inside an arbitrarily shaped closed surface $\mathcal{S}$. What is the infinitesimal flux through an infinitesimal portion da of $\mathcal{S}$ at a point $\vec{r}$ due to the infinitesimal amount of charge in the infinitesimal volume $d\tau'$ at the location $\vec{r}'$? It is

$$d^2\mathcal{F}_{\mathcal{S}}(\vec{r},\vec{r}') = \frac{1}{4\,\pi\,\epsilon_o}\,\frac{d\tau'\rho(\vec{r}')}{|\vec{r}-\vec{r}'|^3}\,(\vec{r}-\vec{r}')\cdot\hat{n}(\vec{r})\,da \tag{2.15}$$

The left side is a double differential because the right side is. If one considers the geometry (see diagram above), one sees that the quantity $(\vec{r}-\vec{r}')\cdot\hat{n}(\vec{r})\,da/|\vec{r}-\vec{r}'|$ is the *projected area* of the area element da normal to the unit vector $(\vec{r}-\vec{r}')/|\vec{r}-\vec{r}'|$ from $\vec{r}'$ to $\vec{r}$. Since $|\vec{r}-\vec{r}'|^2$ is the square of the distance from $\vec{r}'$ to $\vec{r}$, then the quantity $(\vec{r}-\vec{r}')\cdot\hat{n}(\vec{r})\,da/|\vec{r}-\vec{r}'|^3$ is the *solid angle* $d\Omega(\vec{r},\vec{r}')$ subtended by da at $\vec{r}$ as viewed from $\vec{r}'$.

The corresponding mathematical formula is

$$d^2\mathcal{F}_{\mathcal{S}}(\vec{r},\vec{r}\,') = \frac{1}{4\,\pi\,\epsilon_o}\,d\tau'\rho(\vec{r}\,')\,d\Omega(\vec{r},\vec{r}\,') \tag{2.16}$$

We know that if we integrate the solid angle over the entire closed surface $\mathcal{S}$ surrounding our source charge point $\vec{r}\,'$, we recover $4\,\pi$, so:

$$d\mathcal{F}_{\mathcal{S}}(\vec{r}\,') = \frac{1}{4\,\pi\,\epsilon_o}\oint_{\mathcal{S}} d\tau'\rho(\vec{r}\,')\,d\Omega(\vec{r},\vec{r}\,') = \frac{1}{\epsilon_o}\,d\tau'\rho(\vec{r}\,') \tag{2.17}$$

That is, for any infinitesimal volume element $d\tau'$ at $\vec{r}\,'$, Coulomb's Law implies that the flux of the electric field due to that element through any surface $\mathcal{S}$ enclosing it is equal to the charge in that infinitesimal volume divided by ϵ_o.

We expect that, due to superposition, if the above is true for the flux due to an infinitesimal volume of charge, then it holds for the whole distribution of charge enclosed by $\mathcal{S}$. We can prove this by calculating the flux through $\mathcal{S}$ due to the entire charge distribution, using the fact that the distribution is fully contained inside $\mathcal{S}$ (one of our starting assumptions):

$$\begin{aligned}\mathcal{F}_{\mathcal{S}} &= \oint_{\mathcal{S}} \vec{E}_{\mathcal{S}}(\vec{r}) \cdot \hat{n}(\vec{r})\, da = \frac{1}{4\,\pi\,\epsilon_o} \oint_{\mathcal{S}} \int_{\mathcal{V}(\mathcal{S})} \frac{d\tau' \rho(\vec{r}\,')}{|\vec{r} - \vec{r}\,'|^3} \left(\vec{r} - \vec{r}\,'\right) \cdot \hat{n}(\vec{r})\, da \\ &= \frac{1}{4\,\pi\,\epsilon_o} \oint_{\mathcal{S}} \int_{\mathcal{V}(\mathcal{S})} d\tau' \rho(\vec{r}\,')\, d\Omega(\vec{r}, \vec{r}\,') = \frac{1}{4\,\pi\,\epsilon_o} \oint_{\mathcal{S}} \int_{\mathcal{V}(\mathcal{S})} d^2\mathcal{F}_{\mathcal{S}}(\vec{r}, \vec{r}\,') \qquad (2.18)\end{aligned}$$

where $\vec{E}_{\mathcal{S}}(\vec{r})$ is the electric field at $\vec{r}$ due to all the charge contained by $\mathcal{S}$. Note that we implicitly used superposition in the above via the formula relating $\vec{E}_{\mathcal{S}}(\vec{r})$ to the charge distribution. Exchanging the order of integration,

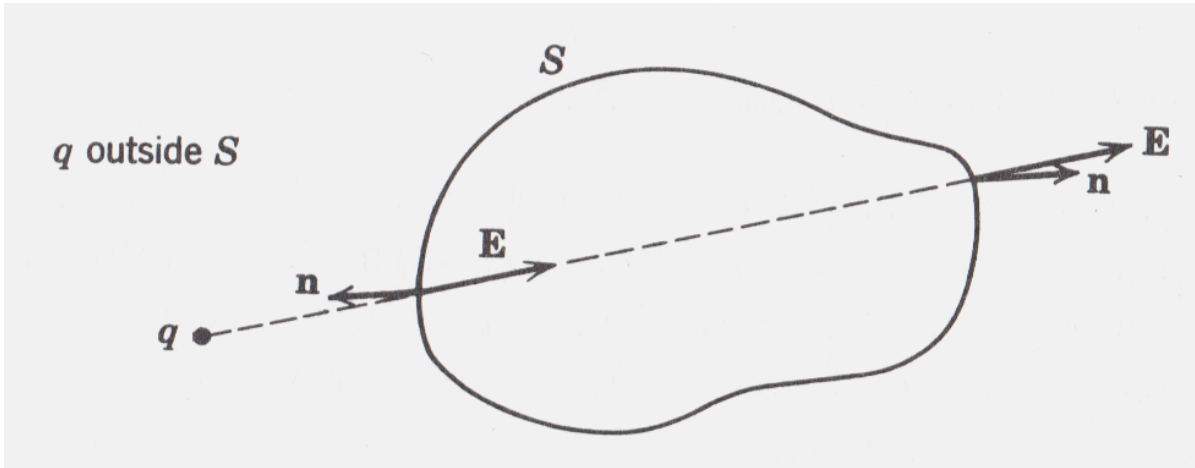
$$\mathcal{F}_{\mathcal{S}} = \frac{1}{4\,\pi\,\epsilon_o} \int_{\mathcal{V}(\mathcal{S})} \oint_{\mathcal{S}} d^2\mathcal{F}_{\mathcal{S}}(\vec{r}, \vec{r}\,') = \frac{1}{\epsilon_o} \int_{\mathcal{V}(\mathcal{S})} d\mathcal{F}_{\mathcal{S}}(\vec{r}\,') = \frac{1}{\epsilon_o} \int_{\mathcal{V}(\mathcal{S})} d\tau' \rho(\vec{r}\,') \quad (2.19)$$

which is Gauss's Law.

Note how the proof depended on the $1/r^2$ dependence and superposition property of Coulomb's Law. The proof could be done in the opposite direction: Gauss's Law implies Coulomb's Law. In general, for any force, there is a simple Gauss's Law if and only if the force has a $1/r^2$ dependence. Another example is gravity, as you learned in Ph1a and Ph106a.

But we are not quite done yet, as we assumed at the start that the charge distribution vanishes outside of $\mathcal{S}$. Does the result generalize to the case where there is some charge outside of $\mathcal{S}$ so that $\vec{E}_{\mathcal{S}}$ receives contributions from that charge? Yes, it does.

Returning to $d^2\mathcal{F}_{\mathcal{S}}(\vec{r},\vec{r}\,')$ (Equation 2.16), suppose we consider a source charge at a point $\vec{r}\,'$ that lies outside of $\mathcal{S}$. (See diagram below.) Then, for a given point $\vec{r}$ on $\mathcal{S}$ and the solid angle it subtends $d\Omega(\vec{r},\vec{r}\,')$ as viewed from the source charge point $\vec{r}\,'$, there will be second point on $\mathcal{S}$ that has the same unit vector to the source charge point $\vec{r}\,'$ and subtends the same solid angle. But, because the direction of $\widehat{n}(\vec{r})$ enters the expression for $d^2\mathcal{F}_{\mathcal{S}}(\vec{r},\vec{r}\,')$, and the two points subtending the same solid angle will have opposite signs of $\widehat{n}$, their two contributions cancel. Thus, the integral over $\mathcal{S}$ that yields $d\mathcal{F}_{\mathcal{S}}(\vec{r}\,')$ vanishes for $\vec{r}\,'$ outside of $\mathcal{S}$, and, therefore, the charge distribution at points outside of $\mathcal{S}$ do not contribute to the flux through $\mathcal{S}$, and so our derivation remains valid.



The Divergence of $\vec{E}$ and the Differential Version of Gauss's Law

You learned about the divergence theorem (Gauss's theorem) in Ma1abc. Applied to $\vec{E}$, the divergence theorem says

$$\int_{\mathcal{V}(\mathcal{S})} \vec{\nabla} \cdot \vec{E}(\vec{r})\, d\tau = \oint_{\mathcal{S}} \vec{E}(\vec{r}) \cdot \hat{n}(\vec{r})\, da \tag{2.20}$$

Gauss's Law tells us

$$\frac{1}{\epsilon_o} \int_{\mathcal{V}(\mathcal{S})} d\tau\, \rho(\vec{r}) = \oint_{\mathcal{S}} \vec{E}(\vec{r}) \cdot \hat{n}(\vec{r})\, da \tag{2.21}$$

Combining the two, we have

$$\int_{\mathcal{V}(\mathcal{S})} \vec{\nabla} \cdot \vec{E}(\vec{r})\, d\tau = \frac{1}{\epsilon_o} \int_{\mathcal{V}(\mathcal{S})} d\tau\, \rho(\vec{r}) \tag{2.22}$$

Since the above holds for any volume $\mathcal{V}$, the integrands must be equal, giving us the differential version of Gauss's Law:

$$\boxed{\vec{\nabla} \cdot \vec{E}(\vec{r}) = \frac{1}{\epsilon_o} \rho(\vec{r})} \tag{2.23}$$

The differential version states that a particular combination of the spatial derivatives of the electric field at a point is related to the charge density at that point. *We will frequently employ this technique of using an equality between two integrals over an arbitrary volume or surface to conclude their integrands are equal*.

Direct Proof of Differential Version of Gauss's Law

We can prove the above differential version by simply calculating the divergence of $\vec{E}$ using Coulomb's Law, also. This is of course not independent of Gauss's Law because Gauss's Law is proven using Coulomb's Law, but it provides some exercise in vector calculus and leads us to the Dirac delta function. We take the divergence of Coulomb's Law for $\vec{E}$, noting explicitly that the divergence is a set of derivatives in $\vec{r}$:

$$\vec{\nabla}_{\vec{r}} \cdot \vec{E}(\vec{r}) = \vec{\nabla}_{\vec{r}} \cdot \int_{\mathcal{V}'} \frac{1}{4\,\pi\,\epsilon_o} \frac{d\tau' \rho(\vec{r}')}{|\vec{r} - \vec{r}'|^3} (\vec{r} - \vec{r}') \tag{2.24}$$

The integral is over $\vec{r}'$ over the volume $\mathcal{V}'$, but the divergence is calculated in the $\vec{r}$ coordinate, so we can bring the divergence inside the integral. Note that it does not act on ρ because ρ is a function of $\vec{r}'$, not $\vec{r}$. Thus, we have

$$\vec{\nabla}_{\vec{r}} \cdot \vec{E}(\vec{r}) = \frac{1}{4\,\pi\,\epsilon_o} \int_{\mathcal{V}'} d\tau' \rho(\vec{r}')\, \vec{\nabla}_{\vec{r}} \cdot \frac{\vec{r} - \vec{r}'}{|\vec{r} - \vec{r}'|^3} \tag{2.25}$$

One could calculate the above divergence explicitly in any particular coordinate system. But it is both more rigorous and more instructive to do it using the divergence theorem, which requires us to apply a nonintuitive technique: we integrate both sides of the above equation in order to evaluate it.

We can calculate the integral of the above divergence over some arbitrary volume $\mathcal{V}$ (with surface $\mathcal{S}$, with neither $\mathcal{V}$ nor $\mathcal{S}$ necessarily related to $\mathcal{V}'$ and $\mathcal{S}'$), as we need to do for Gauss's Law, by exchanging the order of integration (no prohibition on doing so because we don't move $\vec{\nabla}_{\vec{r}}$ around) and converting the volume integral over $\vec{r}$ to an easier-to-do surface integral using the divergence theorem:

$$\begin{aligned}\int_{\mathcal{V}} d\tau\, \vec{\nabla}_{\vec{r}} \cdot \vec{E}(\vec{r}) &= \int_{\mathcal{V}} d\tau\, \frac{1}{4\,\pi\,\epsilon_o} \int_{\mathcal{V}'} d\tau'\rho(\vec{r}\,')\, \vec{\nabla}_{\vec{r}} \cdot \frac{\vec{r}-\vec{r}\,'}{|\vec{r}-\vec{r}\,'|^3} \\ &= \frac{1}{4\,\pi\,\epsilon_o} \int_{\mathcal{V}'} d\tau'\rho(\vec{r}\,') \int_{\mathcal{V}} d\tau\, \vec{\nabla}_{\vec{r}} \cdot \frac{\vec{r}-\vec{r}\,'}{|\vec{r}-\vec{r}\,'|^3} \\ &= \frac{1}{4\,\pi\,\epsilon_o} \int_{\mathcal{V}'} d\tau'\rho(\vec{r}\,') \oint_{\mathcal{S}(\mathcal{V})} da\, \widehat{n}(\vec{r}) \cdot \frac{\vec{r}-\vec{r}\,'}{|\vec{r}-\vec{r}\,'|^3} \end{aligned} \tag{2.26}$$

We can apply to the surface integral the same argument about solid angles that we used in proving Gauss's Law. The integrand above is just the solid angle subtended by the area element da at $\vec{r}$ as viewed from $\vec{r}\,'$. As before, if $\vec{r}\,'$ is inside $\mathcal{V}$, then the above integral yields the total solid angle, $4\,\pi$. If $\vec{r}\,'$ is not inside of $\mathcal{V}$, then, for every area element da at $\vec{r}$, there is an area element with an equal and opposite contribution, making the integral vanish. That is, as in the proof of Gauss's Law,

$$\oint_{\mathcal{S}(\mathcal{V})} da\, \widehat{n}(\vec{r}) \cdot \frac{\vec{r}-\vec{r}\,'}{|\vec{r}-\vec{r}\,'|^3} = \begin{cases} 4\,\pi & \text{if } \vec{r}\,' \text{ is inside } \mathcal{V} \\ 0 & \text{if } \vec{r}\,' \text{ is outside } \mathcal{V} \end{cases} \tag{2.27}$$

The above statement says that the integral over $\mathcal{V}$ vanishes if $\vec{r}'$ is not inside $\mathcal{V}$ and yields $4\,\pi$ if it is inside $\mathcal{V}$. This turns the double integral over $\mathcal{V}$ and $\mathcal{V}'$ into a single integral over $\mathcal{V} \cap \mathcal{V}'$:

$$\int_{\mathcal{V}} d\tau\, \vec{\nabla}_{\vec{r}} \cdot \vec{E}(\vec{r}) = \frac{1}{4\,\pi\,\epsilon_o} \int_{\mathcal{V}\cap\mathcal{V}'} d\tau'\, 4\,\pi\,\rho(\vec{r}') = \frac{1}{\epsilon_o} \int_{\mathcal{V}\cap\mathcal{V}'} d\tau'\, \rho(\vec{r}') \tag{2.28}$$

Now, consider points in $\mathcal{V}$ but outside $\mathcal{V} \cap \mathcal{V}'$. Because $\mathcal{V}'$ is the entire volume containing charge (by Coulomb's Law), the charge density vanishes in $\mathcal{V} - \mathcal{V} \cap \mathcal{V}'$. We can thus add the volume $\mathcal{V} - \mathcal{V} \cap \mathcal{V}'$ without changing the integral of the charge density because the contribution from the added volume vanishes. This changes the volume of integration from $\mathcal{V} \cap \mathcal{V}'$ to $\mathcal{V}$. Therefore,

$$\int_{\mathcal{V}} d\tau\, \vec{\nabla}_{\vec{r}} \cdot \vec{E}(\vec{r}) = \frac{1}{\epsilon_o} \int_{\mathcal{V}} d\tau'\, \rho(\vec{r}') \tag{2.29}$$

The volume $\mathcal{V}$ is arbitrary, so the integrands must be equal:

$$\vec{\nabla}_{\vec{r}} \cdot \vec{E}(\vec{r}) = \frac{1}{\epsilon_o}\, \rho(\vec{r}) \tag{2.30}$$

which is again the differential version of Gauss's Law.

Note the use of two nonintuitive techniques: Using the equality of integrals over an arbitrary volume to show their integrands are equal, and integrating an expression to determine what it is equal to.

Lecture 2:

Basics of Electrostatics II:
Dirac Delta Function Redux
$\vec{\nabla} \times \vec{E} = 0$
Electric Potential
Poisson's and Laplace's Equations

Date Revised: 2024/01/09 05:45
Date Given: 2024/01/09

Aside: Relation of the Dirac Delta Function to a Divergence, Invariance under Inversion of its Argument

Let's return to a geometric theorem we proved during the above manipulations to prove a property of the Dirac delta function. We showed above that:

$$\int_{\mathcal{V}} d\tau \, \vec{\nabla}_{\vec{r}} \cdot \frac{\vec{r}-\vec{r}\,'}{|\vec{r}-\vec{r}\,'|^3} = \oint_{\mathcal{S}(\mathcal{V})} da \, \hat{n}(\vec{r}) \cdot \frac{\vec{r}-\vec{r}\,'}{|\vec{r}-\vec{r}\,'|^3} = \begin{cases} 4\pi & \text{if } \vec{r}\,' \text{ is inside } \mathcal{V} \\ 0 & \text{if } \vec{r}\,' \text{ is outside } \mathcal{V} \end{cases} \tag{2.31}$$

The far right side is proportional to the integral of the Dirac delta function, yielding

$$\int_{\mathcal{V}} d\tau \, \vec{\nabla}_{\vec{r}} \cdot \frac{\vec{r}-\vec{r}\,'}{|\vec{r}-\vec{r}\,'|^3} = 4\pi \int_{\mathcal{V}} d\tau \, \delta(\vec{r}-\vec{r}\,') \tag{2.32}$$

(Note the ordering of $\vec{r}$ and $\vec{r}\,'$ in the argument of the delta function! $\vec{r}\,'$ is the equivalent of $\vec{r}_o$ in Equation 2.9.) Since these integrals are equal for an arbitrary volume $\mathcal{V}$, the integrands are equal:

$$\boxed{\vec{\nabla}_{\vec{r}} \cdot \frac{\vec{r}-\vec{r}\,'}{|\vec{r}-\vec{r}\,'|^3} = 4\pi \, \delta(\vec{r}-\vec{r}\,')} \tag{2.33}$$

The delta function is the divergence of the $1/r^2$ law. We will find this very useful!

Since the delta function picks out the point where its argument vanishes, it doesn't matter what the sign of the argument is. One can prove this explicitly using change of variables: when the sign of the argument changes, the sign of the differential and of the limits of integration change also. Those two sign flips cancel each other. Thus

$$\boxed{\delta(\vec{r}-\vec{r}\,') = \delta(\vec{r}\,'-\vec{r})} \tag{2.34}$$

In order to introduce a new technique we will use soon, let's show this more explicitly:

$$4\pi\delta(\vec{r}-\vec{r}') = \vec{\nabla}_{\vec{r}}\cdot\frac{\vec{r}-\vec{r}'}{|\vec{r}-\vec{r}'|^3} = \vec{\nabla}_{\vec{r}}\cdot-\frac{\vec{r}'-\vec{r}}{|\vec{r}'-\vec{r}|^3} = -\vec{\nabla}_{\vec{r}}\cdot\frac{\vec{r}'-\vec{r}}{|\vec{r}'-\vec{r}|^3} \stackrel{?}{=} 4\pi\delta(\vec{r}'-\vec{r}) \qquad (2.35)$$

The last equality is not obviously true because it is not just a symbolic replacement of Equation 2.33. We can prove it by making a few more steps, though:

$$\vec{\nabla}_{\vec{r}}\cdot\frac{\vec{r}-\vec{r}'}{|\vec{r}-\vec{r}'|^3} = \vec{\nabla}_{\vec{r}-\vec{r}'}\cdot\frac{\vec{r}-\vec{r}'}{|\vec{r}-\vec{r}'|^3} = \left(-\vec{\nabla}_{\vec{r}'-\vec{r}}\right)\cdot\left(-\frac{\vec{r}'-\vec{r}}{|\vec{r}'-\vec{r}|^3}\right) = \vec{\nabla}_{\vec{r}'}\cdot\frac{\vec{r}'-\vec{r}}{|\vec{r}'-\vec{r}|^3} \qquad (2.36)$$

We used the following: 1) the divergence with respect to $\vec{r}-\vec{r}'$ is the same as the divergence with respect to $\vec{r}$ because $\vec{r}-\vec{r}'$ just offsets $\vec{r}$; 2) when we flip the sign on $\vec{r}-\vec{r}'$, we can get the argument of the divergence to match this sign flip by flipping the sign on the divergence overall (the 1D analogue would be to flip the sign on d/dx when mirroring through the origin, $x \to -x$); and, 3) $\vec{r}$ acts like an offset for $\vec{r}'$, and so the divergence with respect to $\vec{r}'-\vec{r}$ is the same as with respect to $\vec{r}'$. Therefore:

$$4\pi\delta(\vec{r}-\vec{r}') = \vec{\nabla}_{\vec{r}}\cdot\frac{\vec{r}-\vec{r}'}{|\vec{r}-\vec{r}'|^3} = \vec{\nabla}_{\vec{r}'}\cdot\frac{\vec{r}'-\vec{r}}{|\vec{r}'-\vec{r}|^3} = 4\pi\delta(\vec{r}'-\vec{r}) \qquad (2.37)$$

Note this technique of applying an offset; we will use it again.

The above derivation is a situation in which it is very important to remember what coordinate $\vec{\nabla}$ is acting on/with respect to. If we had not subscripted $\vec{\nabla}$ in the manipulation above, we would have written:

$$4\,\pi\,\delta(\vec{r}-\vec{r}\,') = \vec{\nabla}\cdot\frac{\vec{r}-\vec{r}\,'}{|\vec{r}-\vec{r}\,'|^3} = \vec{\nabla}\cdot-\frac{\vec{r}\,'-\vec{r}}{|\vec{r}\,'-\vec{r}|^3} = -\vec{\nabla}\cdot\frac{\vec{r}\,'-\vec{r}}{|\vec{r}\,'-\vec{r}|^3} \stackrel{?}{=} -4\,\pi\,\delta(\vec{r}\,'-\vec{r}) \qquad (2.38)$$

Rather than just yielding a last equality that was not obviously true, we have obtained an equality that seems to contradict the symmetry property! From the more detailed derivation on the previous page, we see that the seeming contradiction vanishes when we take care with the meaning of symbols by explicitly subscripting $\vec{\nabla}$. Errors of the above type are easy to make and not self-evident! Mathematics in physics is not just symbol manipulation: there is meaning that must be understood in order to be sure those manipulations are justified.

The Electric Field has Vanishing Curl

Calculating the Curl of the Electric Field

The curl of $\vec{E}$ can be shown to vanish simply by calculating it for an arbitrary charge distribution:

$$\begin{aligned}\vec{\nabla} \times \vec{E}(\vec{r}) &= \frac{1}{4\,\pi\,\epsilon_o}\,\vec{\nabla} \times \int_{\mathcal{V}} d\tau'\,\rho(\vec{r}\,')\,\frac{\vec{r}-\vec{r}\,'}{|\vec{r}-\vec{r}\,'|^3} \\ &= \frac{1}{4\,\pi\,\epsilon_o}\int_{\mathcal{V}} d\tau'\,\rho(\vec{r}\,')\,\vec{\nabla} \times \frac{\vec{r}-\vec{r}\,'}{|\vec{r}-\vec{r}\,'|^3} \qquad (2.39)\end{aligned}$$

We could brute-force calculate the curl in the integrand in Cartesian or spherical coordinates, but that would be painful because the function on which the curl is acting has no symmetry in the $\vec{r}$ coordinate system.

Let's take a simpler, more geometric, and more intuitive approach. As we saw above, $\vec{r}'$ is just an offset to $\vec{r}$, thus

$$\vec{\nabla}_{\vec{r}} \times \frac{\vec{r}-\vec{r}'}{|\vec{r}-\vec{r}'|^3} = \vec{\nabla}_{\vec{r}-\vec{r}'} \times \frac{\vec{r}-\vec{r}'}{|\vec{r}-\vec{r}'|^3} \tag{2.40}$$

Note that, in doing this offset, the curl will be expressed in terms of the components of $\vec{r}-\vec{r}'$. This does not change the bounds of integration, but it may make the expression look complicated because the variable of integration is still $\vec{r}'$. Since we will show this expression, the integrand, vanishes, this bookkeeping complication is not important. If we define $\vec{s} = \vec{r}-\vec{r}'$, then we have

$$\vec{\nabla}_{\vec{r}} \times \frac{\vec{r}-\vec{r}'}{|\vec{r}-\vec{r}'|^3} = \vec{\nabla}_{\vec{s}} \times \frac{\vec{s}}{s^3} \tag{2.41}$$

Now, the function on which the curl is acting has symmetry in the coordinate system in which the curl is acting, and hence the calculation will be simplified. You can probably see intuitively that the above curl vanishes, but let's prove it. (Note also that the change of variables would require a change to the limits of integration, but, again, because we will prove the integrand will vanish, this bookkeeping complication will not be important.)

With the above form, we can trivially apply the formula for the curl in spherical coordinates, which is listed in Griffiths. For the sake of being explicit, that formula is (with r replaced by s as the radial coordinate to avoid confusion)

$$\vec{\nabla} \times \vec{v} = \frac{1}{s}\frac{1}{\sin\theta}\left[\frac{\partial}{\partial\theta}\left(v_\phi \sin\theta\right) - \frac{\partial v_\theta}{\partial\phi}\right]\widehat{s} + \frac{1}{s}\left[\frac{1}{\sin\theta}\frac{\partial v_s}{\partial\phi} - \frac{\partial}{\partial s}\left(s\, v_\phi\right)\right]\widehat{\theta} + \frac{1}{s}\left[\frac{\partial}{\partial s}\left(s\, v_\theta\right) - \frac{\partial v_s}{\partial\theta}\right]\widehat{\phi} \tag{2.42}$$

In the case considered here, s is the radial variable and the radial component of $\vec{s}/s^3$ is $1/s^2$. Thus, $\vec{v}$ has only a radial component and that radial component depends only on the radial distance from the origin. All the derivatives involving the θ and ϕ components of $\vec{v}$ vanish because the components themselves vanish, and the derivatives involving the radial component vanish because those derivatives are with respect to θ and ϕ. (Don't be confused: $\vec{v}$ itself depends on θ and ϕ because the direction of $\vec{v}$ depends on them; but the curl formula takes care of that dependence.)

Thus, we have $\vec{\nabla}_{\vec{s}} \times (\vec{s}/s^3) = 0$ and the integrand in Equation 2.39 vanishes. So:

$$\boxed{\vec{\nabla} \times \vec{E}(\vec{r}) = 0} \tag{2.43}$$

Note again that we did not brute-force differentiate, but rather we thought about how to simplify the calculational aspect (via origin offset) and then saw that made the result both geometrically/intuitively obvious and easier to demonstrate via calculation.

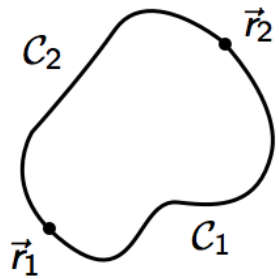
The Line Integral of the Electric Field

Stokes' Theorem (a mathematical theorem we will not prove here but that you saw in Ma1abc) then tells us that, for any surface $\mathcal{S}$ with boundary $\mathcal{C}(\mathcal{S})$,

$$\oint_{\mathcal{C}(\mathcal{S})} d\vec{\ell} \cdot \vec{E}(\vec{r}) = \int_{\mathcal{S}} da\, \hat{n}(\vec{r}) \cdot \left[\vec{\nabla} \times \vec{E}(\vec{r})\right] = 0 \tag{2.44}$$

The Electric Potential

Electric Potential Definition using Line Integral



We used above the fact that the line integral of the electric field around any closed loop $\mathcal{C}$ vanishes. If we consider two points along the loop $\vec{r}_1$ and $\vec{r}_2$, $\mathcal{C}$ defines two paths along the loop from $\vec{r}_1$ to $\vec{r}_2$, $\mathcal{C}_1$ and $\mathcal{C}_2$. Let's difference the line integrals along these two paths, using the vanishing of the line integral around the loop to see that the difference vanishes:

$$\int_{\mathcal{C}_1,\vec{r}_1}^{\vec{r}_2} d\vec{\ell}\cdot\vec{E}(\vec{r}) - \int_{\mathcal{C}_2,\vec{r}_1}^{\vec{r}_2} d\vec{\ell}\cdot\vec{E}(\vec{r}) = \int_{\mathcal{C}_1,\vec{r}_1}^{\vec{r}_2} d\vec{\ell}\cdot\vec{E}(\vec{r}) + \int_{\mathcal{C}_2,\vec{r}_2}^{\vec{r}_1} d\vec{\ell}\cdot\vec{E}(\vec{r})$$
$$= \oint_{\mathcal{C}} d\vec{\ell}\cdot\vec{E}(\vec{r}) = 0 \qquad (2.45)$$

(Be careful about the endpoint ordering and signs of the two terms! The differential $d\vec{\ell}$ in a line integral has no intrinsic polarity; the polarity is set by the ordering of the endpoints. The sign of $d\vec{\ell}$ and of the endpoint ordering do not multiply; they are redundant.) Therefore,

$$\int_{\mathcal{C}_1,\vec{r}_1}^{\vec{r}_2} d\vec{\ell}\cdot\vec{E}(\vec{r}) = \int_{\mathcal{C}_2,\vec{r}_1}^{\vec{r}_2} d\vec{\ell}\cdot\vec{E}(\vec{r}) \qquad (2.46)$$

The above relation tells us that the value of the above line integral depends only on the location of its endpoints, not on the path taken. Thus, we can construct a function, *the electric potential*, $V(\vec{r})$, defining it via its differences between points:

$$\boxed{V(\vec{r}_2) - V(\vec{r}_1) \equiv -\int_{\vec{r}_1}^{\vec{r}_2} d\vec{\ell} \cdot \vec{E}(\vec{r})} \tag{2.47}$$

The fundamental theorem of calculus for line integrals in multiple dimensions implies

$$V(\vec{r}_2) - V(\vec{r}_1) = \int_{\vec{r}_1}^{\vec{r}_2} d\vec{\ell} \cdot \vec{\nabla} V(\vec{r}) \tag{2.48}$$

where $\vec{\nabla} V(\vec{r})$ is the *gradient* of the electric potential. The above two formulae hold regardless of choice of endpoints and path, so the integrands are equal and we have

$$\boxed{\vec{E}(\vec{r}) = -\vec{\nabla} V(\vec{r})} \tag{2.49}$$

which can be viewed as an alternate definition of the potential. The offset of $V(\vec{r})$ is not defined because it has no influence on $\vec{E}(\vec{r})$, which is the quantity we began with from Coulomb's Law.

The electric potential has units of (N m/C) = (J/C), which we call the *volt*, V. (The appearance of J will be important when we discuss electric potential energy.) The electric field is frequently written in units of V/m instead of N/C.

It will be useful in homework to remember the mathematical definition of a gradient (another case of keeping present in your mind what the mathematical symbols mean):

$$\lim_{|\delta\vec{r}|\to 0} V(\vec{r}+\delta\vec{r}) - V(\vec{r}) = \lim_{|\delta\vec{r}|\to 0} \vec{\nabla} V(\vec{r}) \cdot \delta\vec{r} = -\lim_{|\delta\vec{r}|\to 0} \vec{E}(\vec{r}) \cdot \delta\vec{r} \qquad (2.50)$$

The above expression also serves to remind you of how to do a Taylor expansion in multiple dimensions: the dot product of the gradient and the vector differential replaces the product of the one-dimensional derivative and the scalar differential.

Relation of the Electric Potential to the Charge Distribution

We know two things now:

$$\vec{E}(\vec{r}) = \frac{1}{4\,\pi\,\epsilon_o}\int_{\mathcal{V}} d\tau'\,\rho(\vec{r}\,')\,\frac{\vec{r}-\vec{r}\,'}{|\vec{r}-\vec{r}\,'|^3} \qquad \text{and} \qquad V(\vec{r}_2) - V(\vec{r}_1) \equiv -\int_{\vec{r}_1}^{\vec{r}_2} d\vec{\ell}\cdot\vec{E}(\vec{r})$$

We can use these to obtain an explicit expression for the potential in terms of the charge distribution. In practice, trying to do the line integral explicitly using the definition of $\vec{E}$ is tedious and not illuminating.

Instead, let us use our understanding of the *meaning* of the mathematical expression $\vec{E}(\vec{r}) = -\vec{\nabla}V(\vec{r})$ to make an Ansatz. If we have a point charge at the origin, then the electric field points radially outward and falls off as $1/r^2$. What function's derivative gives that dependence? $V(\vec{r}) = 1/r$. This suggests to us

$$\boxed{V(\vec{r}) = \frac{1}{4\,\pi\,\epsilon_o}\int_{\mathcal{V}} d\tau'\,\rho(\vec{r}\,')\,\frac{1}{|\vec{r}-\vec{r}\,'|}} \tag{2.51}$$

We may then prove explicitly this form is correct by taking the gradient.

First, pass $\vec{\nabla}$ inside the integral because it is $\vec{\nabla}_{\vec{r}}$ while the variable of integration is $\vec{r}\,'$:

$$-\vec{\nabla}_{\vec{r}} V(\vec{r}) = -\frac{1}{4\,\pi\,\epsilon_o}\int_{\mathcal{V}} d\tau'\,\rho(\vec{r}\,')\,\vec{\nabla}_{\vec{r}}\left(\frac{1}{|\vec{r}-\vec{r}\,'|}\right) \tag{2.52}$$

As we did earlier when calculating $\vec{\nabla}\times\vec{E}$, we change variables to $\vec{s}=\vec{r}-\vec{r}\,'$ to evaluate the gradient:

$$\vec{\nabla}_{\vec{r}}\left(\frac{1}{|\vec{r}-\vec{r}\,'|}\right) = \vec{\nabla}_{\vec{r}-\vec{r}\,'}\left(\frac{1}{|\vec{r}-\vec{r}\,'|}\right) = \vec{\nabla}_{\vec{s}}\frac{1}{s} = -\frac{\widehat{s}}{s^2} = -\frac{\vec{r}-\vec{r}\,'}{|\vec{r}-\vec{r}\,'|^3} \tag{2.53}$$

where we used the formula for the gradient in spherical coordinates from Griffiths (again, with r replaced by s as the radial coordinate to avoid confusion):

$$\vec{\nabla}_{\vec{s}} T(\vec{s}) = \frac{\partial T}{\partial s}\,\widehat{s} + \frac{1}{s}\,\frac{\partial T}{\partial \theta}\,\widehat{\theta} + \frac{1}{s}\,\frac{1}{\sin\theta}\,\frac{\partial T}{\partial \phi}\,\widehat{\phi} \tag{2.54}$$

We see that our form for $V(\vec{r})$ yields the correct electric field (dropping the $\vec{r}$ subscript on $\vec{\nabla}$ because, on the left-hand side, it is obvious what coordinate $\vec{\nabla}$ acts on):

$$-\vec{\nabla} V(\vec{r}) = \frac{1}{4\,\pi\,\epsilon_o}\int_{\mathcal{V}} d\tau'\,\rho(\vec{r}\,')\,\frac{\vec{r}-\vec{r}\,'}{|\vec{r}-\vec{r}\,'|^3} = \vec{E}(\vec{r}) \tag{2.55}$$

Comments on the Electric Potential

- **The electric potential obeys superposition**
  This is trivial consequence of superposition for the electric field: because the electric potential is a linear function of the electric field, and integration is a linear operation, superposition for the electric field transfers to superposition for the electric potential. One can also see it from Equation 2.51, where the potential is a linear function of the charge density.
- **Definition of potential offset**
  There are two typical choices. When the charge distribution is confined to a finite volume, the electric field vanishes at infinity, which suggests one should define the electric potential to vanish at infinity too. When the charge distribution is not confined (*e.g.*, a uniform electric field over all of space), it is typical to choose the origin to be the point at which the potential vanishes. Any other point would work, too, but will generally make the explicit functional form of $V(\vec{r})$ unnecessarily complicated if one is interested in using the above integral expression. There will be situations, however, where such a choice is the most convenient.
- **Utility of the electric potential**
  The electric potential is scalar, not a vector, function, and thus applying superposition to calculate the potential due to a charge distribution, followed by taking the gradient to find the electric field, is usually much simpler than explicitly calculating the electric field.

Aside on Techniques

It is important to recognize how we almost uniformly avoided brute-force calculations of divergences, curls, and gradients so far. The only times we did those calculations explicitly were when we had rendered the calculations trivial. A key part of doing E&M successfully and with minimal pain is avoiding algebra and calculus whenever possible and instead making use of clever arguments of the type we used above. Only do algebra and calculus as a last resort! There are two reasons for this.

First, the kinds of arguments we used are more physical and help you develop intuition. For example, in proving the differential version of Gauss's Law, at no point did we explicitly take derivatives of $\vec{E}$! Incredible, right? Instead, we proved that the divergence of the $1/r^2$ law is the delta function (again, not explicitly, but by referring to the geometric proof we made for the integral version of Gauss's Law) and used that fact. We could have done the brute-force calculation in Cartesian coordinates, and it would have given the same result. But you would have derived no intuition from it.

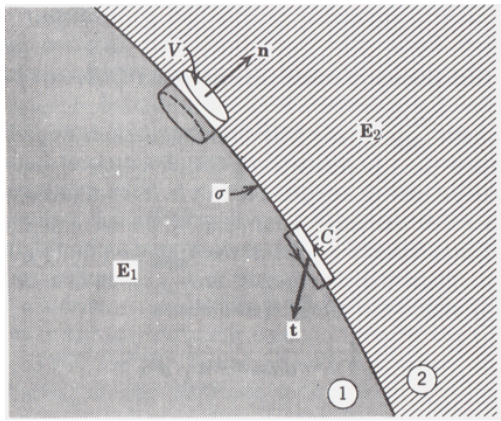
Second, brute-force calculations are prone to oversights — like the one about the sign flip on $\vec{\nabla}$ in the delta-function symmetry derivation — as well as bookkeeping mistakes — algebraic sign flips, misapplications of the product and chain rules, etc. Doing brute-force calculations does not help you understand physics, or even mathematics. Of course, sometimes brute-force calculations are needed, but try to avoid them, and keep your wits and intuition about you as you do them!

It takes time to learn how to work this way, but we do derivations (rather than just quote results) so you can learn these techniques.

Boundary Conditions on the Electric Field and Potential

While Gauss's Law makes it possible to determine the electric field for charge distributions with sufficient symmetry, the more important application of Gauss's Law and the vanishing of $\vec{\nabla} \times \vec{E}$ is to obtain generic information on the behavior of the electric field and potential across an interface between two regions.

Boundary Condition on the Normal Component of the Electric Field



Construct a Gaussian cylinder of infinitesimal height dz whose axis is normal to the interface under question at the point of interest. Let $\hat{n}$ be the surface normal at $\vec{r}$, with orientation from region 1 to region 2. Let's calculate the flux through the cylinder's (non-infinitesimal) faces $\mathcal{S}_1$ and $\mathcal{S}_2$:

$$\mathcal{F} = \int_{\mathcal{S}_1} da\,(-\hat{n}(\vec{r})) \cdot \vec{E}_1(\vec{r}) + \int_{\mathcal{S}_2} da\,\hat{n}(\vec{r}) \cdot \vec{E}_2(\vec{r}) \qquad (2.56)$$

where $\vec{E}$ is evaluated over the two faces. We neglect the flux through the cylindrical wall because we will let dz vanish in the end and so its area will vanish and it will contribute no flux. We momentarily make the assumption that there is no charge density that is singular in the direction parallel to the interface — *i.e.*, point charges or a line charge density — so that we don't have to worry about possible singularities in the electric field that might complicate the flux calculation. We allow only a surface charge density, which is delta-function singular in the z dimension but not in the dimensions parallel to the interface.

For Gauss's Law, the volume integral of the charge density enclosed has two contributions: from the non-delta-function-like *volume* charge density in the half-cylinders and from any delta-function-like *surface* charge density on the surface. The contribution of the former will vanish as we let $dz \to 0$. The latter converts the volume integral to a surface integral:

$$\mathcal{F} = \frac{1}{\epsilon_o} \int_{\mathcal{V}} d\tau \, \delta(\vec{r} - \mathcal{S}) \, \sigma(\vec{r}) = \frac{1}{\epsilon_o} \int_{\mathcal{S}} da \, \sigma(\vec{r}) \tag{2.57}$$

where $\mathcal{S}$ is the area at the interface intersected by the cylinder. (Note that this is a case where the delta function's argument requires some interpretation to understand the delta function's units. It is the $\mathcal{S}$ in the argument that implies the function has units of m^{-1} rather than m^{-3}: it is picking out a surface rather than a point and thus changing the units by one power of distance, not three.) Equating the two expressions for $\mathcal{F}$, letting $dz \to 0$, and seeing that $\mathcal{S}_1, \mathcal{S}_2 \to \mathcal{S}$ as $dz \to 0$ in the flux integral yields

$$\int_{\mathcal{S}} da \, \hat{n}(\vec{r}) \cdot \left[\vec{E}_2(\vec{r}) - \vec{E}_1(\vec{r}) \right] = \frac{1}{\epsilon_o} \int_{\mathcal{S}} da \, \sigma(\vec{r}) \tag{2.58}$$

This holds for any choice of cylinder and thus any $\mathcal{S}$, so the integrands must be equal:

$$\boxed{\hat{n}(\vec{r}) \cdot \left[\vec{E}_2(\vec{r}) - \vec{E}_1(\vec{r}) \right] = \frac{1}{\epsilon_o} \sigma(\vec{r})} \tag{2.59}$$

That is, the change in the normal component of the electric field across the interface is proportional to the surface charge density at the interface. If there is no surface charge at the interface, this component of the electric field must be continuous.

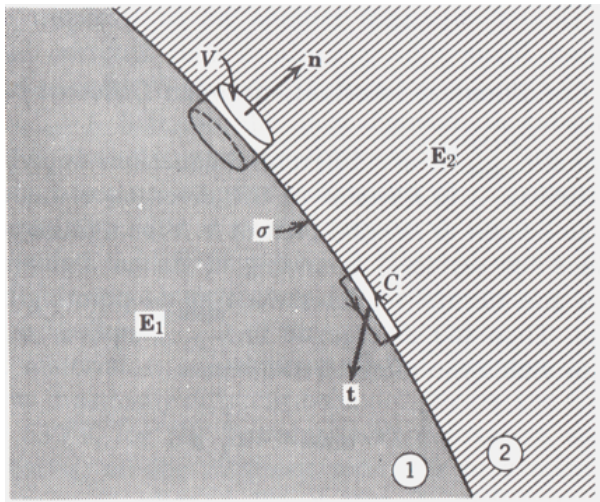
Let's now reconsider the condition we placed at the start of the derivation, that there be no charge density at the intersection of the cylinder and interface that is singular in the dimension parallel to the interface, which could consist of a set of point charges and/or a line charge density.

If that charge density is not at the contour $\mathcal{C}$ consisting of the intersection of the cylinder's wall and the interface, then the flux of its field remains entirely calculable. It may cause $\sigma(\vec{r})$ to have a delta-function singularity in one or two dimensions parallel to the interface, but no part of the derivation fails. We simply allow that type of $\sigma(\vec{r})$ in Equation 2.59.

If the parallel-dimension-singular charge density is on $\mathcal{C}$, then things are bit more complicated. If we consider the flux through the cylindrical wall anywhere but on the charge density, that flux vanishes because the field of the charge density is always parallel to the cylindrical wall as $dz \to 0$. What about *on* the charge density? Answering this question in a mathematically *explicit* fashion — *i.e.*, by calculation — is difficult, as the field not only becomes singular at this point but the direction of the singular field depends on the direction from which one approaches the charge. One can, however, conclude from this indeterminancy that there cannot be a contribution to the flux, as it would imply that the field direction is *not* indeterminate so that $\widehat{n} \cdot \vec{E}$ can be nonzero. This is a mathematically valid proof by contradiction. Thus, such charge distributions do not affect the derivation and Equation 2.59 continues to hold.

We will see the above expectation confirmed in practice when we compare the potential for the point charge near the grounded sphere derived by method of images (which does not rely on Equation 2.59 in the case of such a singular charge distribution) and by separation of variables (which does).

Boundary Condition on the Tangential Component of the Electric Field



© 1999 Jackson, *Classical Electrodynamics*

Construct a rectangular loop $\mathcal{C}$ with two legs normal to the interface of interest (*i.e.*, along $\hat{n}(\vec{r})$ at positions $\vec{r}_a$ and $\vec{r}_b$) having infinitesimal length dz and two (non-infinitesimal) legs parallel to the interface $\mathcal{C}_1$ and $\mathcal{C}_2$. Let $\hat{t}(\vec{r})$ denote the normal to the loop area (so $\hat{n}(\vec{r}) \cdot \hat{t}(\vec{r}) = 0$). $\hat{t}$ will set the orientation of the line integral we will do around the loop following the right-hand rule. The loop legs $\mathcal{C}_1$ and $\mathcal{C}_2$ parallel to the interface are parallel to the vector $\hat{s}(\vec{r}) = \hat{t}(\vec{r}) \times \hat{n}(\vec{r})$. Let's calculate the line integral of $\vec{E}$ along this loop (referencing the diagram: $\vec{r}_a$ at the lower right, $\vec{r}_b$ at the upper left):

$$\oint_{\mathcal{C}} d\vec{\ell} \cdot \vec{E}(\vec{r}) = \int_{\mathcal{C}_1, \vec{r}_b - \hat{n}(\vec{r})\frac{dz}{2}}^{\vec{r}_a - \hat{n}(\vec{r})\frac{dz}{2}} \vec{E}_1(\vec{r}) \cdot d\vec{\ell} + \int_{\mathcal{C}_2, \vec{r}_a + \hat{n}(\vec{r})\frac{dz}{2}}^{\vec{r}_b + \hat{n}(\vec{r})\frac{dz}{2}} \vec{E}_2(\vec{r}) \cdot d\vec{\ell} \qquad (2.60)$$

where we neglect the contributions from the infinitesimal legs because they will vanish as $dz \to 0$. (We may apply arguments similar those just used in the derivation of the normal field boundary condition to show that these legs contribute nothing even in the case of a charge density at the interface with a delta-function singularity in the dimension parallel to the interface.)

Be careful about the signs of the integrals: $d\vec{\ell}$ for an open contour acquires its orientiation from the the ordering of the endpoints; it has no intrinsic orientation until this ordering is specified. Therefore, the sign of $d\vec{\ell}$ and of the endpoint ordering do not multiply; they are redundant. Specifically, in this case, the endpoints imply that $d\vec{\ell}$ points along $+\widehat{s}$ for the second term and $-\widehat{s}$ for the first term and thus that the integrands have opposite sign. Do not then think that the opposite polarity of the endpoint ordering of the two terms implies another relative sign between the two integrals, with the two relative signs canceling!

The vanishing of the curl of the electric field implies the left side of the equation is zero.

We can combine the two terms on the right side by changing the endpoint ordering on the first term and recognizing that $\mathcal{C}_1 \to \mathcal{C}_2$ as $dz \to 0$ (remember: $\mathcal{C}_1$ and $\mathcal{C}_2$ themselves have no orientation: the orientation of the line integrals is set by the ordering of the endpoints). Thus, we have

$$0 = -\int_{\mathcal{C}_1, \vec{r}_a - \widehat{n}(\vec{r})\frac{dz}{2}}^{\vec{r}_b - \widehat{n}(\vec{r})\frac{dz}{2}} \vec{E}_1(\vec{r}) \cdot d\vec{\ell} + \int_{\mathcal{C}_2, \vec{r}_a + \widehat{n}(\vec{r})\frac{dz}{2}}^{\vec{r}_b + \widehat{n}(\vec{r})\frac{dz}{2}} \vec{E}_2(\vec{r}) \cdot d\vec{\ell} \overset{dz \to 0}{\longrightarrow} \int_{\mathcal{C}_2, \vec{r}_a}^{\vec{r}_b} \left[\vec{E}_2(\vec{r}) - \vec{E}_1(\vec{r})\right] \cdot d\vec{\ell}$$

With this ordering of the endpoints, we may identify $d\vec{\ell} = \widehat{s}(\vec{r})\, ds$. Since the contour $\mathcal{C}_2$ is arbitrary, the integrand must vanish, yielding

$$\boxed{\widehat{s}(\vec{r}) \cdot \left[\vec{E}_2(\vec{r}) - \vec{E}_1(\vec{r})\right] = 0} \tag{2.61}$$

This expression holds for any $\widehat{t}$ and thus $\widehat{s}$ parallel to the surface, so it tells us that the tangential component of the electric field is continuous across any boundary (regardless of whether there is surface charge present).

Boundary Conditions on the Electric Potential

From our definition of the electric potential as the line integral of the electric field, and the corollary $\vec{E} = \vec{\nabla} V$, we can derive boundary conditions on the electric potential:

- **Continuity of the electric potential**
  The electric potential is the line integral of the electric field. If we think about calculating the discontinuity in V by integrating $\vec{E} \cdot \hat{n}\, d\ell$ across the boundary, we recognize that, as the length of the path goes to zero, the only way to prevent the integral from vanishing is if $\vec{E} \cdot \hat{n}$ is not only nonzero but delta-function singular. The only place that can conceivably happen is at a point where a charge density becomes singular in at least one dimension (point charge or linear or surface charge density). In the same way as we argued in the derivation of the normal field boundary condition, Equation 2.59, we may also argue here that this quantity $\vec{E} \cdot \hat{n}$ still vanishes and thus V is always continuous.

  We do note that, while V could become infinite near these charge densities, it must approach infinity from both sides of the boundary in the same way, and thus it remains continuous. We will see this in the example of the point charge near the grounded sphere when we do separation of variables in spherical coordinates.

- **Change in the normal gradient**
  This is just a direct rewriting of the boundary condition on the normal component of the field, Equation 2.59:

$$\frac{1}{\epsilon_o}\sigma(\vec{r}) = \hat{n}(\vec{r}) \cdot \left[\vec{E}_2(\vec{r}) - \vec{E}_1(\vec{r})\right] = \hat{n}(\vec{r}) \cdot \left[-\vec{\nabla} V_2(\vec{r}) + \vec{\nabla} V_1(\vec{r})\right]$$

$$\Longrightarrow \boxed{\hat{n}(\vec{r}) \cdot \left[\vec{\nabla} V_2(\vec{r}) - \vec{\nabla} V_1(\vec{r})\right] = -\frac{1}{\epsilon_o}\sigma(\vec{r})} \tag{2.62}$$

  Note the sign!

- **Continuity of the tangential gradient**
  Again, this follows directly from the continuity of the tangential component of the electric field, Equation 2.61:

$$0 = \hat{s}(\vec{r}) \cdot \left[\vec{E}_2(\vec{r}) - \vec{E}_1(\vec{r})\right] = \hat{s}(\vec{r} \cdot \left[-\vec{\nabla} V_2(\vec{r}) + \vec{\nabla} V_1(\vec{r})\right]$$

$$\Longrightarrow \boxed{\hat{s}(\vec{r}) \cdot \left[\vec{\nabla} V_2(\vec{r}) - \vec{\nabla} V_1(\vec{r})\right] = 0} \tag{2.63}$$

Poisson's and Laplace's Equations

It is natural to rewrite Gauss's Law in terms of the electric potential:

$$\frac{1}{\epsilon_o}\rho(\vec{r}) = \vec{\nabla}\cdot\vec{E}(\vec{r}) = -\nabla^2 V(\vec{r}) \tag{2.64}$$

Rewritten more cleanly:

$$\boxed{\nabla^2 V(\vec{r}) = -\frac{1}{\epsilon_o}\rho(\vec{r})} \tag{2.65}$$

This is known as *Poisson's Equation*.

Poisson's Equation is a partial differential equation. You know from basic calculus that a differential equation alone is not sufficient to obtain a full solution $V(\vec{r})$: constants of integration are required. For partial differential equations in multiple dimensions, the constants of integration are given by specifying *boundary conditions*, conditions for how the solution or its derivatives must behave on the boundary of the volume in which we are specifying $\rho(\vec{r})$ and would like to determine $V(\vec{r})$.

Our expression for the potential in terms of the charge distribution, Equation 2.51, is the explicit solution to this equation *for a particular boundary condition*, $V(\vec{r}) \to 0$ as $r \to \infty$. Section 3.11 will develop the concept of a Green Function, which is the generic tool for solving Poisson's Equation for arbitrary boundary conditions.

When there is no charge and the right side vanishes, Equation 2.65 is known as *Laplace's Equation*. The importance of this equation is that it implies that, in a region where there is no charge, the second derivative vanishes everywhere, which implies there can be no local maxima or minima (they would require a positive or negative second derivative). We will prove this explicitly in Section 3.1.

For completeness, let's also rewrite the curl-freeness of the electric field in terms of the electric potential. There is a mathematical theorem that the curl of a gradient always vanishes:

$$\vec{\nabla} \times (-\vec{\nabla} V) = 0 \tag{2.66}$$

This is not surprising, as the vanishing of the curl of $\vec{E}$ is the mathematical property of $\vec{E}$ that allowed us to define the potential as a line integral, which then allowed us to write $\vec{E}$ as the gradient of the potential. The above must be true for self-consistency.

Lecture 3:

Basics of Electrostatics III:
Electric Potential Energy
Conductors

Date Revised: 2024/01/16 06:00
Revised lecture break
Date Given: 2024/01/11

Electrostatic Energy

Electric Potential Energy of a Point Charge in an Electric Field

Consider moving a point charge from $\vec{r}_1$ to $\vec{r}_2$ along a contour $\mathcal{C}$. The work done *on* the charge by the *mechanical* force pushing it in this direction is given by doing the line integral of the *negative* of the electric force along the path because that is the mechanical force that has to be exerted to move the charge against the electric force $\vec{F}_e$:

$$W_{12} = -\int_{\mathcal{C},\vec{r}_1}^{\vec{r}_2} d\vec{\ell} \cdot \vec{F}_e(\vec{r}) \tag{2.67}$$

The force is related to the electric field, and so we have

$$W_{12} = -q \int_{\mathcal{C},\vec{r}_1}^{\vec{r}_2} d\vec{\ell} \cdot \vec{E}(\vec{r}) = q\,[V(\vec{r}_2) - V(\vec{r}_1)] \tag{2.68}$$

That is, the work done on the charge by the mechanical force in going from $\vec{r}_1$ to $\vec{r}_2$ is given by the charge times the change in electric potential between the two positions. Note the sign: if the potential is higher at the end point, then the *mechanical* work done was positive.

Of course, this lets us to define the *electric potential energy* by

$$\boxed{U(\vec{r}_2) - U(\vec{r}_1) = q\,[V(\vec{r}_2) - V(\vec{r}_1)]} \tag{2.69}$$

That is, the electric potential energy of the charge and the electric potential of the field are simply related. Since it was defined in terms of work done against a force, electric potential energy obviously has units of Joules (J). That is explicit in the above form, which is C (N m/C) = (N m) = J.

Note that the electric field can also do work on the charge. In this case, the sign in the above line integral for the work is flipped and work is done as the charge loses potential energy. In this case, the work done by the electric field on a charge is what gives it the kinetic energy it has at the end: the electric potential energy is converted to mechanical kinetic energy.

We are not going to write an expression for the potential energy of a continuous distribution of charge quite yet because it is difficult to distinguish two pieces: the change in potential energy that arises from moving the charge distribution in an external potential, and the change in potential energy due simply to a reconfiguration of the charge distribution in its own potential, and we will discuss the latter next.

Electric Potential Energy of a Charge Distribution

How much work must be done to assemble a distribution of charge? This energy is most easily understood by first considering the assembly of a set of point charges one-by-one by bringing them in from infinity. When the ith charge is brought in, work must be done against the electric field of the first $i-1$ charges. Put another way, the ith charge starts with zero potential energy and ends with potential energy

$$U_i = \sum_{j=1}^{i-1} q_i \, \frac{1}{4\,\pi\,\epsilon_o} \, \frac{q_j}{|\vec{r}_i - \vec{r}_j|} \tag{2.70}$$

Thus, the total potential energy is

$$\boxed{U = \frac{1}{4\,\pi\,\epsilon_o} \sum_{i=1}^{N} \sum_{j=1}^{i-1} \frac{q_i q_j}{|\vec{r}_i - \vec{r}_j|} = \frac{1}{8\,\pi\,\epsilon_o} \sum_{i,j=1, i\neq j}^{N} \frac{q_i q_j}{|\vec{r}_i - \vec{r}_j|}} \tag{2.71}$$

where the factor of $1/2$ was introduced to allow i and j to both run from 1 to N. Generalizing this to a continuous charge distribution, we have

$$\boxed{U = \frac{1}{8\,\pi\,\epsilon_o} \int_{\mathcal{V}} d\tau \int_{\mathcal{V}} d\tau' \, \frac{\rho(\vec{r})\,\rho(\vec{r}')}{|\vec{r} - \vec{r}'|}} \tag{2.72}$$

In principle, the above expression is valid for any configuration of charge and potential. But sometimes we do not specify the charge distribution sourcing part of the potential — e.g., a uniform electric field over all of space — and, instead, we just state what the potential is. We can then generalize the above expression to:

$$\boxed{U = \frac{1}{8\,\pi\,\epsilon_o}\int_{\mathcal{V}} d\tau \int_{\mathcal{V}} d\tau'\,\frac{\rho(\vec{r})\,\rho(\vec{r}\,')}{|\vec{r}-\vec{r}\,'|} + \int_{\mathcal{V}} d\tau\,\rho(\vec{r})\,V_{ext}(\vec{r})} \qquad (2.73)$$

where $V_{ext}(\vec{r})$ is the potential sourced by charges not explicitly indicated.

One has to be a bit careful with such situations, though, for two reasons. First, the potential energy provided is not the total electric potential energy of the system because it does not include the energy needed to put $V_{ext}(\vec{r})$ in place, only the energy needed to put $\rho(\vec{r})$ together and to put it in $V_{ext}(\vec{r})$. More importantly, because the charges sourcing $V_{ext}(\vec{r})$ are at infinity, it is not possible to start $\rho(\vec{r})$ out at infinity with zero potential energy there and then bring it into the charge configuration. We define the zero point for $V_{ext}(\vec{r})$ for convenience, *e.g.*, at the origin, but only changes in the second term above are meaningful; the actual value of the second term is not.

Electric Potential Energy in Terms of the Electric Field

We can use the relations between potential, field, and charge density (Equations 2.6, 2.51, and 2.65) and the divergence theorem (Equation 2.20) to obtain an alternate expression for the electric potential energy in terms of the electric field as follows (now disallowing any "external potentials" $V_{ext}(\vec{r})$ of the type mentioned above, requiring us to have knowledge of all charge distributions sourcing potentials):

$$U = \frac{1}{8\,\pi\,\epsilon_o}\int_{\mathcal{V}} d\tau \int_{\mathcal{V}} d\tau'\,\frac{\rho(\vec{r})\,\rho(\vec{r}\,')}{|\vec{r}-\vec{r}\,'|} = \frac{1}{2}\int_{\mathcal{V}} d\tau\rho(\vec{r})\,V(\vec{r}) = -\frac{\epsilon_o}{2}\int_{\mathcal{V}} d\tau\left[\nabla^2 V(\vec{r})\right]V(\vec{r})$$

$$\overset{ibp}{=} -\frac{\epsilon_o}{2}\int_{\mathcal{V}} d\tau\,\vec{\nabla}\cdot\left[V(\vec{r})\,\vec{\nabla}V(\vec{r})\right] + \frac{\epsilon_o}{2}\int_{\mathcal{V}}|\vec{\nabla}V(\vec{r})|^2 \quad \text{with } \overset{ibp}{=} \equiv \text{integration by parts}$$

$$\overset{\substack{divergence\\theorem}}{=} \frac{\epsilon_o}{2}\int_{\mathcal{S}(\mathcal{V})} da\,\widehat{n}\cdot\left[V(\vec{r})\,\vec{E}(\vec{r})\right] + \frac{\epsilon_o}{2}\int_{\mathcal{V}}|\vec{\nabla}V(\vec{r})|^2 \tag{2.74}$$

In the last line, the first term is an integral of the product of the potential and the field at the surface of the volume. In order to get the full energy of the charge distribution, $\mathcal{V}$ must include all the charge. If we assume the charge distribution is restricted to some finite volume, then $\mathcal{V}$ is naturally the volume containing the charge distribution. But we can add volume that does not contain charge because it contributes nothing to the initial expression for the electric potential energy. (This requirement of restriction of $\rho(\vec{r})$ to finite volume is exactly the requirement of no external potentials.)

Therefore, we replace $\mathcal{V}$ with all of space and let $\mathcal{S}$ go to infinity:

$$U = \frac{\epsilon_o}{2} \int_{r\to\infty} da\, \hat{n} \cdot \left[V(\vec{r})\, \vec{E}(\vec{r}) \right] + \frac{\epsilon_o}{2} \int_{all\ \ space} |\vec{\nabla} V(\vec{r})|^2 \tag{2.75}$$

Because the charge distribution is restricted to the finite volume $\mathcal{V}$ and thus looks like a point charge as $r \to \infty$, the field and potential fall off like $1/r^2$ and $1/r$. The surface area of $\mathcal{S}$ only grows as r^2, so the integral goes like $1/r$ and thus vanishes as $r \to \infty$. (If the charge distribution is not restricted to a finite volume, the surface term may not vanish, requiring one to either keep the surface term or use the initial expression.)

It may seem strange that we can make this choice of $\mathcal{S}$, as changing $\mathcal{V}$ and $\mathcal{S}$ affects both integrals in the last expression. The explanation is that the choice of $\mathcal{S}$ changes the two integrals but leaves their sum constant, and taking $\mathcal{S}$ to infinity simply zeros out the first integral, leaving only the contribution of the second integral.

We thus find

$$\boxed{U = \frac{\epsilon_o}{2} \int |\vec{E}(\vec{r})|^2} \tag{2.76}$$

where the integral is over all of space. Correspondingly, the quantity $u = \frac{\epsilon_o}{2} |\vec{E}|^2$ is an energy density. We interpret this form as indicating that the potential energy created by assembling the charge distribution is stored in the field: less charge implies a smaller field and therefore less potential energy.

Superposition and Electric Potential Energy

Because the electric potential energy is a quadratic function of the charge distribution or the electric field,

electric potential energy does not obey superposition

The energy of a sum of fields is more than just the sum of the energies of the individual fields because there are cross terms due to the potential energy of the charges in one another's fields.

Self-energy and Point Charges vs. Continuous Charge Distributions

We were slightly cavalier in going from Equation 2.71 to Equation 2.72 in that the "self-energy" term $i = j$ that was not included in the former did get included in the latter. In the point-charge version, this term is infinite because the denominator vanishes. In the continuous distribution version, $\rho(\vec{r})\,\rho(\vec{r}\,')\,d\tau \to 0$ as $|\vec{r} - \vec{r}\,'| \to 0$ as long as ρ remains finite over all space, and thus there is no infinite contribution. (If ρ included a delta function, as would be necessary to represent a point charge, then it would produce an infinite contribution because the integral would yield $\delta(\vec{0})/0$.) Thus, we must be careful and choose the appropriate formula depending on the situation.

The infinite self-energy of a point charge reflects the fact that we do not know how to assemble a point charge. In fundamental particle physics, the existence of point charges such as the electron is an assumption, not a consequence, of the theory. In fact, there is scheme, called "renormalization," by which the infinite self-energy one calculates for such a charge from Equation 2.76 is "subtracted off" in a self-consistent fashion across all situations. While this practice is accepted and applied carefully, it is not understood. String theory, which postulates that all particles are actually vibrating string-like objects with finite extent, may offer a solution, but string theory currently is not complete — it does not offer a way to calculate the Standard Model — and there is no explicit proof it is correct.

Electric Conductors

Definition and Behavior of a Conductor

We now talk about *electric conductors*, both because they are interesting and because they provide a first opportunity to use boundary conditions to determine properties of the charge distribution, field, and potential. Notice that we derive these properties without explicit calculations!

An *electric conductor* is defined to be a material in which charge is able to flow completely freely in response to an external electric field. It is assumed, *a priori*, to contain equal and opposite amounts of positive and negative electric charge that perfectly cancel everywhere in the absence of an electric field ($\rho = 0$) but that can separate in response to an electric field. One can add charge to a conductor explicitly.

Without any calculation, we know what the response of the conductor will be to an externally applied electric field: If there is any field present in the conductor, positive and negative charge densities will separate in response to the field. That separation results in an additional field whose direction is opposite the applied field because of the directions the two polarities of charge move in response to the applied field. This movement occurs until the sum field vanishes, at which point there is no further force on the charges and the system becomes static. Therefore, $\vec{E} = 0$ inside any conductor. Note the lack of distinction between the applied field and the field created by the charges: each charge is only sensitive to the *total* field, so it is the *total* field that must vanish inside the conductor. The charges arrange themselves so their contribution to the total field cancels that of the applied field.

Derived Properties of a Conductor

We may derive the following conductor properties from the fact that $\vec{E} = 0$ inside a conductor everywhere:

- **ρ also vanishes inside a conductor**
  This follows directly from Gauss's Law: because $\vec{E} = 0$ everywhere in the interior, then $\vec{\nabla} \cdot \vec{E} = \rho/\epsilon_o$ also vanishes.

  Another way of seeing this, at least for a conductor with no net charge, is that, if there were a nonzero ρ, then there must be an equal and opposite amount of charge elsewhere in the conductor because the conductor is neutral overall. An electric field would appear between the oppositely signed charge distributions, contradicting the $\vec{E} = 0$ condition. Alternatively, the opposite charge will be attracted to the nonzero ρ by the field and move to cancel it until the field vanishes.

- **Any net charge or induced charge resides on the surface**
  The picture we described before, of charge separation being induced by the external field, does imply that there may be such *induced charge* on the surface. This does not violate Gauss's Law because $\vec{E}$ may be nonzero outside the conductor and thus one has to be careful in calculating $\vec{\nabla} \cdot \vec{E}$ at the conductor boundary (we must resort to the boundary conditions we derived, Equations 2.59 and 2.61).

  Also, if we intentionally add charge to a conductor, it must also move to the surface by the same Gauss's Law argument. An alternative, microscopic way of seeing this is that, if we add charge to a neutral conductor, which has no electric field or charge density in its interior, the added charge repels itself, pushing itself to the exterior (as far as it can go without leaving the conductor). Or, equivalently, the added charge attracts charge from the surface to cancel it, leaving net charge on the surface. Regardless, the added charge that now appears on the surface arranges itself so there is no net field in the interior.

  Aside: As Griffiths notes in a footnote, this property can be interpreted to be a consequence of the fact that the electric field obeys the Coulomb's Law $1/r^2$ dependence in three dimensions (from which we derived Gauss's Law, which we used above in the proof). In a different number of dimensions, or with a different dependence on r, we would not have been able to derive Gauss's Law! There will be a homework problem considering conductors when Coulomb's Law is modified.

- **A conductor has the same electric potential everywhere**
  That is, a conductor is an *equipotential*. This occurs because $\vec{E}$ vanishes everywhere in the conductor: any line integral of $\vec{E}$ between two points must therefore also vanish. The conductor may have a nonzero electric potential, but the value is the same everywhere.

  One can see this using the gradient, too. If V were not constant in the conductor, there would be a nonzero $\vec{E} = -\vec{\nabla}V$, which we said above is not allowed.

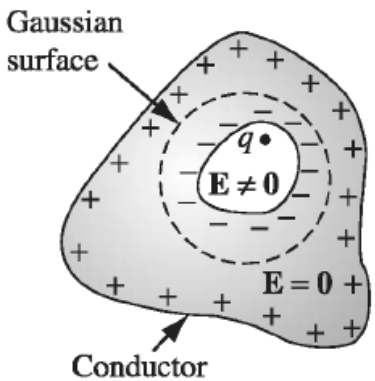
- **The electric field just outside a conductor is always normal to its surface**
  This arises from the boundary conditions we derived, Equations 2.59 and 2.61. Since $\vec{E}$ vanishes inside the conductor, and the tangential component of $\vec{E}$ is continuous across any interface, the tangential component must vanish just outside the conductor, too. There is no such condition on the normal component because there may be an induced or net surface charge density σ on the surface.

  Another way of looking at this is is that an electric field tangential to the surface would cause charge to move along the surface until that tangential component vanished. No such argument applies to the normal component because the charge is no longer free to move normal to the surface when it sits at the surface — it cannot leave the conductor.

Conductors with Cavities

The mental image we have so far is of a conductor that has no cavities inside of it. What additional properties can we derive for a conductor with cavities?

- **A charge q inside a cavity in a conductor results in an equal induced charge q on the surface of the conductor**



© 2013 Griffiths, *Introduction to Electrodynamics*

To see this, construct a surface $\mathcal{S}$ that lies inside the conductor but also contains the cavity. The electric field vanishes on $\mathcal{S}$ because it is in the conductor, so the net charge enclosed must vanish. Since a charge q is inside the cavity, there must be a canceling charge $-q$ inside $\mathcal{S}$. Since $\mathcal{S}$ can be shrunk to be arbitrarily close to the inner surface without changing this statement, the induced charge must lie on the inner surface of the cavity.

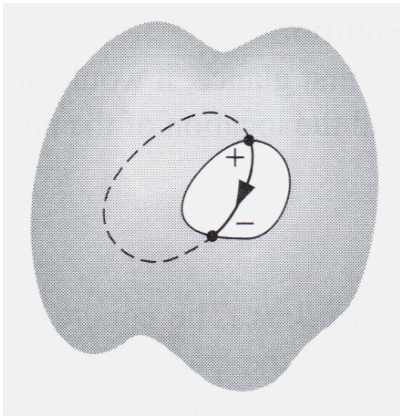
Since $-q$ has appeared on the inner surface, we know, by neutrality of the conductor, there must be a charge $+q$ elsewhere on the conductor. If we now expand $\mathcal{S}$ to approach the outer surface, the above statement about $-q$ inside $\mathcal{S}$ continues to hold, so the only place $+q$ can be is on the outer surface.

The exact distribution of q on the surface depends on the geometry. For cases with some symmetry, we may be able to guess the solution easily.

Consider a conductor with a spherical outer surface. Since there are no field lines inside the conductor, there is no way the charge in the cavity or on the inner surface of the conductor can influence the distribution of charge on the outer surface, even if the inner cavity is non-spherical and/or the charge is not placed at the center of the cavity. Thus, the charge must distribute itself on the outer surface of the conductor in the same way as it would if charge $+q$ were added to a spherical conductor with no cavity. By symmetry, that distribution is uniform with surface charge density $\sigma = q/4\,\pi\,r^2$.

Note, however, that, in general, the charge on the inner surface of the conductor will **not** be distributed uniformly. It will only be uniform if the inner surface is spherical and the charge in the cavity is at the center of the cavity, as this situation has symmetry. (Note that the shape of the outer surface and the inner cavity's location with respect to the outer surface have no impact, for the same reasons as the inner cavity does not affect the distribution of charge on the outer surface.) In any other case, the field lines from the charge in the cavity will exhibit no symmetry as they terminate on the cavity wall and therefore the surface charge required to cancel those field lines in the conductor will have no symmetry.

- **If there is no net charge inside a cavity in a conductor, the electric field inside the cavity vanishes, independent of the external field applied to or net charge added to the conductor**



© 2013 Griffiths, *Introduction to Electrodynamics*

We use proof by contradiction. Assume there is a nonzero electric field in the cavity. Since there is no charge in the cavity, the field lines must start and end on charges on the surface of the cavity. Therefore, there is a path through the cavity with $\int d\vec{\ell} \cdot \vec{E} \neq 0$. Now close the path with a segment inside the conductor. This portion of the now-closed loop $\mathcal{C}$ contributes nothing to the line integral $\oint_{\mathcal{C}} d\vec{\ell} \cdot \vec{E}$ over the entire loop because $\vec{E} = 0$ inside the conductor. Since $\oint_{\mathcal{C}} d\vec{\ell} \cdot \vec{E} = 0$, the contribution from inside the cavity must vanish also. Contradiction. So the assumption $\vec{E} \neq 0$ in the cavity must be false.

Aside 1: Note the technique of proof by contradiction, which we will use again in E&M.

Aside 2: This fact is used for shielding of experiments from external electric fields (and also electromagnetic waves) and is called a *Faraday cage*. Note that the conductor can have some net charge on it (and correspondingly sit at some nonzero electric potential with respect to infinity) and this property still holds. As we will see later, it also holds in the presence of external electromagnetic waves, which is the more typical and important application.

Surface Charge and the Force on the Surface of a Conductor

Our boundary condition for the normal component of the electric field combined with the fact that the electric field vanishes inside a conductor tells us that the electric field infinitesimally above the surface of the conductor is

$$\vec{E} = \frac{\sigma}{\epsilon_o}\,\widehat{n} \tag{2.77}$$

where $\widehat{n}$ points from the inside to the outside of the conductor.

There is a charge density σ at this point, and an electric field above it, so is there a force on the charge? Yes, but the calculation is subtle. The thing to recognize is that the small element of charge $\sigma\, da$ in an infinitesimal area da cannot exert a force on itself. The field to which this element of charge is subject is the field of the charge distribution excluding it. We find this field by finding the field of this charge element and subtracting it from the total field. This is an example of one of the indirect approaches we must apply in E&M: a brute-force approach will not be successful or generic.

We know (Griffiths Example 2.5) that the electric field of a charge sheet in the xy plane is $\vec{E} = \pm(\sigma/2\,\epsilon_o)\,\hat{z}$ where the sign applies depending on whether $z > 0$ or $z < 0$. While the small patch we are considering is not an infinite sheet, it looks like one if we are infinitesimally close to it. We also know $\vec{E}_{other}$ must be continuous at the charge element because, in the absence of that charge element, there is no charge at the boundary and thus no surface charge density to cause a discontinuity in the normal component. (Note that we do not claim we know $\vec{E}_{other}$, only that we know that it has this continuity property!) Thus, we may write the equations

$$\vec{E}_{outside} = \vec{E}_{other} + \frac{\sigma}{2\,\epsilon_o}\,\hat{n} \qquad \vec{E}_{inside} = \vec{E}_{other} - \frac{\sigma}{2\,\epsilon_o}\,\hat{n} \tag{2.78}$$

where $\vec{E}_{other}$ is the field due to the rest of the charge distribution excepting da and, because of its continuity, the same expression for $\vec{E}_{other}$ appears in both equations. (Note this technique, which you learned doing story problems in middle-school pre-algebra, of writing down an equation in which the knowns are not segregated on one side yet.) Using $\vec{E}_{outside} = (\sigma/\epsilon_o)\,\hat{n}$ and/or $\vec{E}_{inside} = 0$, we find $\vec{E}_{other} = (\sigma/2\,\epsilon_o)\,\hat{n}$. This is the field that acts on the charge $\sigma\,da$ in da. Therefore, the force per unit area is

$$\vec{f} = \frac{\vec{F}}{da} = \frac{\sigma\,da\,\vec{E}_{other}}{da} = \sigma\,\frac{\sigma}{2\,\epsilon_o}\,\hat{n} = \frac{\sigma^2}{2\,\epsilon_o}\,\hat{n} \tag{2.79}$$

Writing the force per unit area in terms of the field at the surface $\vec{E} = (\sigma/\epsilon_o)\,\hat{n}$:

$$\boxed{\vec{f} = \frac{\sigma^2}{2\,\epsilon_o}\,\hat{n} = \frac{\epsilon_o}{2}\,E^2\,\hat{n}} \tag{2.80}$$

That is, the surface of a conductor always feels an outward force. Consider what would happen if you put charge on a balloon with a metallized surface.

Note the force per unit area, which has units of energy density, is actually equal to the energy density just above the conductor. We could have in fact used the energy density to derive the force: the force per unit area is the gradient of the energy per unit area, and moving the conductor surface in or out by an infinitesimal distance dz would have changed the total energy per unit area by $u\,dz$.

Note the indirect technique of proof. Again, we did no integral and we did not use Coulomb's Law explicitly.

Lecture 4:

Basic Electrostatics IV:
Capacitance

Date Revised: 2024/01/18 06:45
Revised lecture break
Date Given: 2024/01/16

Capacitance

Consider two conductors (of arbitrary shapes) and suppose we put equal and opposite charges Q and $-Q$ on them. The potential difference ΔV between the two is of course given by the line integral of the electric field from any point on the surface of one to any point on the surface of the other. How does ΔV scale with the charges?

The linear dependence of $\vec{E}$ on the charge density ρ ensures that ΔV is linear in Q. Therefore, we may define the *capacitance*

$$\boxed{C = \frac{Q}{\Delta V}} \tag{2.81}$$

Capacitance is a *purely geometric quantity*: it does not depend on the amount of charge on the two conductors (as long as equal and opposite charges are given to each, a caveat we will remove soon). It does depend on the shapes of the conductors and their *relative* position and orientation because those determine the shape of the electric field (while Q varies its normalization). The unit of capacitance is Coulombs/volt, which we define to be the Farad, F.

One can talk about the capacitance of a single conductor with charge Q by implicitly assuming there is another conductor at infinity that has charge $-Q$ and is defined to be at $V = 0$.

Now departing from Griffiths and instead following Jackson §1.11, we can generalize capacitance to include multiple conductors by simply assuming a generalized linear relationship between potentials, which we also call voltages, and charges as we argued above must be true:

$$V_i = \sum_{j=1}^{N} D_{ij} Q_j \qquad \text{or} \qquad \underline{V} = \underline{\underline{D}}\,\underline{Q} \tag{2.82}$$

where $\underline{V}$ and $\underline{Q}$ are N-element column matrices for the voltages and charges on the N conductors and $\underline{\underline{D}}$ is a $N \times N$ matrix that connects the two. It is explicit that any voltage depends linearly on all the charges. The capacitance matrix is then $\underline{\underline{C}} = \underline{\underline{D}}^{-1}$, with

$$\boxed{Q_i = \sum_{j=1}^{N} C_{ij} V_j \qquad \text{or} \qquad \underline{Q} = \underline{\underline{C}}\,\underline{V}} \tag{2.83}$$

This form serves to make it clear that the capacitance is not just a single quantity between two conductors, but is more general. According to Jackson, the diagonal element C_{ii} is the "capacitance" of electrode i, and the C_{ij} are termed the "coefficients of induction" to convey that they indicate the charge induced on electrode i when a voltage is placed on electrode j. We will show below that neither of these is what one would consider the capacitance of a pair of conductors as we discussed initially.

In all of this, there is an implicit assumption that $V(r \to \infty) = 0$. Without this assumption, we would always need to explicitly include the electrode at ∞ (with an additional index in $\underline{\underline{C}}$ and $\underline{\underline{D}}$) in order to get the right offset for V.

To calculate the capacitance or the capacitance matrix, one clearly needs to determine, given a set of charges $\{Q_i\}$, what the voltages $\{V_i\}$ are. To do this trivially, there typically must be a symmetry or approximation that allows one to guess the charge distributions on the conductors (*e.g.*, uniform as for an infinite parallel plate capacitor) and to calculate the field using Gauss's Law and from the field the potential. For more complex geometries, the boundary-value problem techniques we will develop may be sufficient. The total charge on each electrode normalizes the voltage.

For the simple case of two *mirror-symmetric* electrodes with equal and opposite charges $\pm Q$ and voltages $\pm V$, we can relate the elements of the capacitance matrix to the *pair capacitance*, which is what we usually call the capacitance (*e.g.*, in Ph1b). We can assume the following form for the capacitance matrix:

$$\underline{\underline{C}} = \begin{bmatrix} C_s & -C_m \\ -C_m & C_s \end{bmatrix} \tag{2.84}$$

Why could we assume the above form? The symmetry of the system implies $C_{11} = C_{22}$. We shall see below that all capacitance matrices are symmetric matrices, so $C_{12} = C_{21}$. We chose the negative sign on $C_{12} = -C_m$ with some foreknowledge of the result, but that's a choice and doesn't affect the value of C_{12}.

The defining condition of the *pair capacitance* is that equal and opposite charges are placed on the two conductors. By symmetry, we can conclude that the conductors carry equal and opposite voltages (not true for a non-mirror-symmetric configuration). Thus

$$Q_1 = C_s V_1 - C_m V_2 = C_s V - C_m(-V) = (C_s + C_m)\, V \tag{2.85}$$

$$Q_2 = -C_m V_1 + C_s V_2 = -C_m V + C_s(-V) = -(C_s + C_m)\, V \tag{2.86}$$

which yields $Q_2 = -Q_1 = -Q$ as assumed. Thus, the capacitance of the pair is

$$C = \frac{Q}{\Delta V} = \frac{(C_s + C_m)\, V}{2\, V} = \frac{C_s + C_m}{2} \tag{2.87}$$

After we have discussed energy, we will return to this system for a more detailed analysis of what one can say about C_s and C_m.

Capacitance and Field Lines

Let's also think about capacitance in terms of field lines. The diagonal element D_{ii} tells us the potential of electrode i if we put charge on it and no other electrodes. That potential is the line integral of the field from infinity to the electrode, so it is telling us about the field lines going from the charge on that electrode to infinity (or to/from if the charge is negative). The off-diagonal elements D_{ji} tell us how the potential of electrode j changes when charge is put on electrode i. This makes sense, as that charge on i will change the overall field configuration, also due to the addition of the field lines that must start from or end on its charge, and that change will affect V_j.

The elements of $\underline{\underline{C}}$ are interpreted differently. When we put one electrode i at a voltage while holding the others fixed (possibly at zero), charge must be added to that electrode. The diagonal element C_{ii} tells us how much charge must go onto the electrode, and that charge sources field lines. The off-diagonal elements C_{ji} then tell us how much charge must appear on the other electrodes so their voltages V_j remain fixed. This reflects the fact that some of the new field lines starting (or ending) on electrode i due to the new charge on it must end (start) on some of the other electrodes j, and in fact tells us how much charge must be added to those other electrodes to terminate those new field lines.

As a corollary, an off-diagonal element of $\underline{\underline{D}}$ or $\underline{\underline{C}}$ can only vanish if there is no mutual influence of the two electrodes. For simply connected electrodes (*i.e.*, none of the electrodes have cavities inside them), it is hard to see how this could happen unless they are infinitely far apart!

Electric Potential Energy of a Capacitor

In a simple two-electrode, mirror-symmetric capacitor with charges $\pm q$ on the electrodes and a voltage difference $\Delta V = q/C$ between the two electrodes, the amount of work required to change the charge from q to $q + dq$ is given by the amount of work required to move a charge dq from the negative electrode (which has charge $-q$ and voltage $-\Delta V(q)/2$) to the positive electrode (which has charge $+q$ and voltage $+\Delta V(q)/2$):

$$dU = dq \left[\frac{\Delta V(q)}{2} - \left(-\frac{\Delta V(q)}{2} \right) \right] = \Delta V(q)\, dq = \frac{q}{C}\, dq \tag{2.88}$$

Note that ΔV is a function of q here: the voltage is not held fixed while the charge is moved; rather, the voltage and charge increase together (linearly).

We integrate this expression from 0 to the final charge Q to find

$$U = \frac{1}{C}\int_0^Q q\,dq = \frac{1}{2}\,\frac{Q^2}{C} \tag{2.89}$$

Alternatively, using $Q = C\,\Delta V$,

$$\boxed{U = \frac{1}{2}\,\frac{Q^2}{C} = \frac{1}{2}\,C\,(\Delta V)^2} \tag{2.90}$$

We could have modeled the above process differently. Our transferral of dq from one electrode to the other is the equivalent of taking charge dq from the negative voltage electrode, carrying it out to infinity (where we set $V = 0$), and bringing it back and putting it on the positive voltage electrode. The equivalence is because the voltage difference between two points is path-independent. This process is, then, equivalent to bringing charges dq and $-dq$ in from infinity and putting them on the positive and negative voltage electrodes, respectively. And the last process is equivalent to bringing the charges in consecutively rather than simultaneously because we proved earlier the potential energy does not depend on the order of assembly of the charge distribution.

The above picture is what we need for considering a multi-electrode system: we build up the charge on each conductor by bringing in charge from infinity and calculating the work done. Consider bringing charge dq_i in from infinity and adding it to electrode i. The change in the electric potential energy of the system due to adding this charge is

$$dU_i = V_i\, dq_i = \sum_{j=1}^{N} D_{ij}\, q_j\, dq_i \tag{2.91}$$

There are two possible double-countings we must avoid: 1) This infinitesimal element of charge dq_i is moved from $V = 0$ at infinity to $V = V_i$ on the ith electrode, so the voltages of the other electrodes are irrelevant during this infinitesimal charge transfer and we should not bring them into the equation; 2) Because the charges on all the other electrodes $j \neq i$ are physically immobile as dq_i is brought in, no work is done on them, and so there are no other contributions to include (as strange as it may seem given that their voltages change by $dV_j = D_{ji} dq_i$; remember, a force must be exerted over a distance for it to do work).

Now, let's integrate over dq_i. We will later do a sum over i. The ordering of the two steps does not matter because we proved earlier that the electric potential energy does not depend on the order of assembly. But we do need to worry about the order of how we have brought in the charges because we should not calculate cross-terms for charges that do not yet exist. Let's assume that, if we are integrating the ith charge, then the first $i-1$ charges have already been integrated to their full values $\{Q_j\}$, $j = \{1, \ldots, i-1\}$, and the remaining $N-i$ electrodes $j = \{i+1, \ldots, N\}$ have no charge on them yet. Thus, the voltage $V_i(q_i; \{Q_j\}_{j<i})$ is given by

$$V_i(q_i; \{Q_j\}_{j<i}) = \sum_{j=1}^{N} D_{ij}\, q_j = D_{ii}\, q_i + \sum_{j=1}^{i-1} D_{ij}\, Q_j \tag{2.92}$$

because $q_j = Q_j$ has already been achieved for $j = \{1, \ldots, i-1\}$, $q_j = 0$ for $j = \{i+1, \ldots, N\}$, and $q_i \neq Q_i$ is still being changed. Therefore,

$$\begin{aligned} U_i &= \int_0^{Q_i} V_i(q_i; \{Q_j\}_{j<i})\, dq_i = \int_0^{Q_i} \left[D_{ii}\, q_i\, dq_i + \sum_{j=1}^{i-1} D_{ij}\, Q_j\, dq_i \right] \\ &= \frac{1}{2}\, D_{ii} Q_i^2 + \sum_{j=1}^{i-1} D_{ij} Q_j Q_i \end{aligned} \tag{2.93}$$

Next, we need to sum over i to account for the charging up of all the electrodes:

$$U = \frac{1}{2}\sum_{i=1}^{N} D_{ii} Q_i^2 + \sum_{i=1}^{N}\sum_{j=1}^{i-1} D_{ij} Q_i Q_j \tag{2.94}$$

Modifying the second sum to be symmetric (assuming $\underline{\underline{D}}$ is symmetric, which we will prove below) and including a factor of $1/2$ to correct for double-counting, we have

$$U = \frac{1}{2}\sum_{i=1}^{N} D_{ii} Q_i^2 + \frac{1}{2}\sum_{i,j=1, i\neq j}^{N} D_{ij} Q_i Q_j = \frac{1}{2}\sum_{i,j=1}^{N} D_{ij} Q_i Q_j$$

We can write this more succinctly as

$$\boxed{U = \frac{1}{2}\,\underline{Q}^T\,\underline{\underline{D}}\,\underline{Q} = \frac{1}{2}\,\underline{Q}^T\,\underline{\underline{C}}^{-1}\,\underline{Q}} \tag{2.95}$$

Using $\underline{Q} = \underline{\underline{C}}\,\underline{V}$, we can rewrite as

$$\boxed{U = \frac{1}{2}\,\underline{V}^T\,\underline{\underline{C}}\,\underline{V}} \tag{2.96}$$

Let's check that this gives the correct result for an elementary capacitor with two mirror-symmetric electrodes having equal and opposite charges $\pm Q$ and voltages $\pm V$. Using the capacitance matrix we derived earlier (recall, $C_{11} = C_{22} = C_s$, $C_{12} = C_{21} = -C_m$, and $C = (C_s + C_m)/2$),

$$\begin{aligned} U &= \frac{1}{2}\left[C_{11}(+V)^2 + C_{22}(-V)^2 + C_{12}(+V)(-V) + C_{21}(-V)(+V)\right] \\ &= \frac{1}{2}\,V^2\left[C_s + C_s + C_m + C_m\right] = 2\,C\,V^2 = \frac{1}{2}\,C(\Delta V)^2 \end{aligned} \tag{2.97}$$

as expected.

Properties of the Capacitance Matrix and Its Inverse

We can derive a number of useful properties:

- **Both $\underline{\underline{C}}$ and $\underline{\underline{D}}$ are symmetric.**
  Let's consider two electrodes, i and j with $i \neq j$. From Equation 2.93, their contribution to the potential energy, assuming j has been charged up before i, is

$$U_{ij} = \frac{1}{2}\left(D_{ii}Q_i^2 + D_{jj}Q_j^2\right) + D_{ij}Q_i\,Q_j \tag{2.98}$$

  What happens if we reverse the charging order? Then we get

$$U_{ji} = \frac{1}{2}\left(D_{ii}Q_i^2 + D_{jj}Q_j^2\right) + D_{ji}Q_i\,Q_j \tag{2.99}$$

  In our initial discussion of the electric potential energy, we argued that the charging order does not matter. So we may equate the two, $U_{ij} = U_{ji}$. Recognizing that Q_i and Q_j are arbitrary then implies

$$D_{ij} = D_{ji} \iff \underline{\underline{D}}^T = \underline{\underline{D}} \iff \boxed{\underline{\underline{C}}^T = \underline{\underline{C}}} \tag{2.100}$$

- **The self-capacitances C_{ii} are positive.**
  We need only consider the energy in the case that all other electrodes are held at zero potential. Then the energy is

$$U_i(\text{all others grounded}) = \frac{1}{2} C_{ii} V_i^2 \tag{2.101}$$

  Since the energy should be positive (it takes work to add charge dq_i in the presence of the same-sign charge q_i, as is done when charging up the electrode), C_{ii} must be positive.

- **The diagonal elements of the inverse capacitance matrix, $C_{ii}^{-1} = D_{ii}$ are positive.**
  Now, we consider the energy in the case that all other electrodes are kept neutral. Then the energy is

$$U_i(\text{all others neutral}) = \frac{1}{2} D_{ii} Q_i^2 \tag{2.102}$$

  Again, since the energy should be positive, D_{ii} must be positive.

- **The off-diagonal elements of the inverse capacitance matrix $C_{ij}^{-1} = D_{ij}$ are positive.**
  Now, let's consider two electrodes i, j in a multi-electrode configuration, with all the other electrodes uncharged. Let's suppose electrode i is already raised to its final charge, and now we want to consider the work needed to increment electrode j's charge:

$$dU_{ij} = D_{jj}\, q_j\, dq_j + D_{ij}\, Q_i\, dq_j \tag{2.103}$$

  (The self-terms and cross-terms vanish for all the electrodes $k \neq i, j$ because they have $Q_k = 0$.) If we consider the case of Q_i, q_j positive, and if we bring in more positive charge dq_j, it is obvious that both the change in the jth self-energy and the energy cross-term should be positive: we are bringing positive charges in proximity to existing positive charges. (While the existing charge might move around on the electrodes, those electrodes are conductive and so are equipotentials: no work is done.) We already know the self-energy terms are positive. In order for the energy cross-term to be positive, D_{ij} must be positive. In the mirror-symmetric electrode case, we would see via explicit inversion of $\underline{\underline{C}}$ that $\underline{\underline{D}}$'s off-diagonal elements are positive.

  Another way to see that the cross-terms must be positive is to recall that the entire expression must be consistent with our original expression for the electric potential energy, Equation 2.72. That expression could be broken down into three integrals, one for each self-energy term and one for the cross-term. When the charge density is positive, all contributions to that expression are manifestly positive.

- **The off-diagonal elements of the capacitance matrix C_{ij} are negative.** Let's consider the same multi-electrode system with electrodes $k \neq i, j$ grounded (*i.e.*, $V_k = 0$), electrode i at its final positive voltage V_i, and electrode j's voltage being incremented from v_j to $v_j + dv_j$, both positive. The change in energy is

$$dU_{ij} = C_{jj}\, v_j\, dv_j + C_{ij}\, V_i\, dv_j \tag{2.104}$$

We already know the first term is positive. The second term is more challenging. If we want to increment a positive voltage v_j by a positive amount dv_j, we need to put positive charge on it. This positive charge will draw negative charge out of the battery holding V_i constant: some of the field lines of that new charge on electrode j have to terminate on electrode i if C_{ij} is non-zero. Again, from Equation 2.72, we know that contribution to the electric potential energy must be negative even if V_i is positive. Thus, the energy cross-term must be negative, which requires C_{ij} to be negative. (If V_i is negative, that implies Q_i is negative. It takes positive work to add negative charge to an electrode that already has negative charge on it, so $C_{ij} < 0$ ensures the cross-term becomes positive, as it should.)

- $|\sum_{i\neq j} C_{ij}| \leq |C_{jj}|$: **for a given electrode, the sum of the off-diagonal elements of the capacitance matrix is no larger in magnitude than the corresponding diagonal element.**
  Just consider the same situation as just considered. The change in the charge on the jth electrode is $dq_j = C_{jj}\, dv_j$. The field lines from those added charges will terminate either on other electrodes or infinity, so the total negative charge added to all the other electrodes can be no larger in magnitude than $|dq_j|$. Therefore,

$$\left|\sum_{i\neq j} dq_i = \sum_{i\neq j} C_{ij}\, dv_j\right| \leq \left|dq_j = C_{jj}\, dv_j\right| \qquad \Longrightarrow \qquad \left|\sum_{i\neq j} C_{ij}\right| \leq \left|C_{jj}\right| \quad (2.105)$$

Capacitance Matrix of a Mirror-Symmetric Configuration Revisited

Considering again a mirror-symmetric two-electrode configuration, we now know $C_s > C_m > 0$, and we know the pair capacitance we are familiar with is related to them by $C = (C_s + C_m)/2$, but can we determine C_s and C_m explicitly?

If we consider the case $V_1 = V$ and $V_2 = 0$, we find $Q_1 = C_s V$ and $Q_2 = -C_m V$, so we can determine C_s and C_m if we know the full field configuration, with the boundary condition $V = 0$ at infinity: we obtain the surface charge density from the normal component of the field at the electrode surfaces and integrate it to get Q_1 and Q_2 and thus C_s and C_m. (Remember, if $V \neq 0$ at infinity, we need to include infinity explicitly as an electrode of the system.)

Maybe we can then do this for the one mirror-symmetric case whose full electric field configuration we can calculate trivially, the infinite parallel-plate capacitor? No! The infinite parallel-plate capacitor violates the condition $V = 0$ at infinity because, if either plate has non-zero potential, that plate's non-zero equipotential surface extends off to infinity in the transverse direction. We violate the assumption that allowed us to ignore the electrode at infinity. Moreover, infinity is no longer even an equipotential surface in this configuration! On the equipotentials defined by the two electrodes (at, *e.g.*, $z = \pm d/2$), the potential at infinity is the potential of the corresponding electrode. If the two plates have equal and opposite potentials, then the field outside the plates vanishes and the potential on the surface of that volume at infinity is zero. The potential on the line $z = 0$ is also zero. And then, for $0 < |z| < d/2$ and $x, y \to \infty$, the potential is the same linear function of z that it would be at $x, y = 0$. Clearly, our assumptions are violated!

We note that, formally, C is infinite for this mirror-symmetric configuration, anyways: the mirror-symmetric potential configuration requires infinite charge on each electrode! The pair capacitance per unit area, however, is finite and trivially calculated.

So, we are stymied. In order for the $V = 0$ at infinity condition to be satisfied, our electrodes must be finite in extent. But, for electrodes finite in extent, we cannot calculate the potential in a trivial fashion, so we cannot determine C_s and C_m, or even C, trivially. We need to develop the full machininery for solving Poisson's and Laplace''s Equations, which we will begin to do soon.

Section 3
Advanced Electrostatics

Lecture 5:

Advanced Electrostatics I:
Laplace's Equation
Uniqueness Theorem
Method of Images

Date Revised: 2024/01/23 07:00
Revised lecture break
Date Given: 2024/01/18

Intuitive Approach to Laplace's Equation

As we mentioned earlier, the integral forms for the electric field or the potential

$$\vec{E}(\vec{r}) = \frac{1}{4\,\pi\,\epsilon_o}\int_{\mathcal{V}} d\tau'\,\rho(\vec{r}\,')\,\frac{\vec{r}-\vec{r}\,'}{|\vec{r}-\vec{r}\,'|^3} \qquad \text{and} \qquad V(\vec{r}) = \frac{1}{4\,\pi\,\epsilon_o}\int_{\mathcal{V}} d\tau'\,\frac{\rho(\vec{r}\,')}{|\vec{r}-\vec{r}\,'|} \quad (3.1)$$

are always correct but can be difficult to deal with in practice. Most systems will not have symmetries that make the integrals easily doable (or avoidable via Gauss's Law). Moreover, and this is the greater problem, it is rare that one completely specifies $\rho(\vec{r})$ in setting up a problem. Experimentally, what we can easily control are the shapes, positions, and potentials (voltages) of conductors. We do not control how the charge arranges itself on the conductors. Thus, we need to seek alternate ways to solve for the potential and field over all of space. Laplace's and Poisson's Equations are the key.

Laplace's Equation in One Dimension

In one dimension, Laplace's Equation takes the simple form

$$\frac{d^2V}{dx^2} = 0 \tag{3.2}$$

We can solve this by direct integration to obtain

$$V(x) = mx + b \tag{3.3}$$

where m and b are two constants of integration. We determine m and b by *boundary conditions*: specification of V or dV/dx at specific point(s). In the one dimensional case, there are two options for how to specify the boundary conditions:

- Specify V at two points.
- Specify V at one point and dV/dx at one point (possibly the same point).

Note that these are the only choices in one dimension. Specifying dV/dx at two points either yields a contradiction (if two different values of dV/dx are given) or insufficient information (if the same value is given). There are no other quantities to specify: all higher derivatives vanish thanks to Laplace's Equation.

Let us note two important characteristics of the solutions of Laplace's Equation:

- **Averaging Property:** $V(x)$ is equal to the average of any pair of points $V(x+a)$ and $V(x-a)$ for any a such that $x \pm a$ belong to the region being considered:

$$\begin{aligned}\frac{1}{2}\left[V(x+a)+V(x-a)\right] &= \frac{1}{2}\left[(m(x+a)+b)+(m(x-a)+b)\right] \\ &= mx+b = V(x)\end{aligned} \tag{3.4}$$

- $V(x)$ **has no nontrivial local maxima or minima**
  We already mentioned this property for the three-dimensional Laplace's Equation. The proof is straightforward in one dimension. Suppose x_0 is a local maximum or minimum. Then we have $dV/dx = 0$ at this point x_0. Then, for any other point x_1:

$$\left.\frac{dV}{dx}\right|_{x_1} = \left.\frac{dV}{dx}\right|_{x_0} + \int_{x_0}^{x_1} \frac{d^2V}{dx^2}\,dx = 0+0=0 \tag{3.5}$$

  Therefore, if dV/dx vanishes anywhere, then dV/dx vanishes *everywhere* and thus $V(x)$ is a constant. This is a trivial local maximum/minimum. If dV/dx vanishes nowhere, then the endpoints of the region give the maximum and minimum of $V(x)$ or, if there are no endpoints, there are no maxima or minima at all. Consider, for example, a uniform electric field $\vec{E}_0$ over all of space.

Laplace's Equation in Multiple Dimensions

We quote the analogues of the above two properties for arbitrary numbers of dimensions and prove them for three dimensions:

- **Averaging Property:** The value $V(\vec{r})$ of a solution to Laplace's Equation at any point is equal to the average of its value on any sphere centered on that point in the region of interest:

$$V(\vec{r}) = \langle V(\vec{r}) \rangle_a \equiv \frac{\int_{\mathcal{S}_a(\vec{r})} da'\, V(\vec{r}\,')}{\int_{\mathcal{S}_a(\vec{r})} da'} \tag{3.6}$$

where $\mathcal{S}_a(\vec{r})$ is the sphere of radius a centered on $\vec{r}$. This is straightforward to show (Griffiths Problem 3.37). Let's integrate Laplace's Equation over the volume enclosed by $\mathcal{S}_a(\vec{r})$, $\mathcal{V}_a(\vec{r})$, and use the divergence theorem:

$$0 \overset{\substack{\text{Laplace's} \\ \text{Equation}}}{=} \int_{\mathcal{V}_a(\vec{r})} d\tau'\, \nabla^2_{\vec{r}\,'} V(\vec{r}\,') \tag{3.7}$$

$$\overset{\substack{\text{divergence} \\ \text{theorem}}}{=} \int_{\mathcal{S}_a(\vec{r})} da'\, \hat{n}(\vec{r}\,') \cdot \vec{\nabla}_{\vec{r}\,'} V(\vec{r}\,') \quad = \int_{\mathcal{S}_a(\vec{r})} da'\, \hat{n}(\vec{r}\,') \cdot \vec{\nabla}_{\vec{r}\,'-\vec{r}} V(\vec{r}\,')$$

In the last step, we have used the fact that $\vec{\nabla}$ does not care about the location of the origin (since it is just an offset).

Now, we can define $\vec{s} = \vec{r}\,' - \vec{r}$. In this coordinate system, where $\vec{r}$ is at the origin, $\widehat{n}(\vec{r}\,' \in S_a(\vec{r})) = \widehat{s}$, the radial unit vector in the $\vec{s}$ coordinate system. So, we have (inserting a factor $1/4\,\pi\,a^2$):

$$0 = \frac{1}{4\,\pi\,a^2} \int_{\mathcal{S}_a(\vec{s}=\vec{0})} s^2 d\Omega_s \left.\frac{\partial V}{\partial s}\right|_{s=a} \tag{3.8}$$

where $\mathcal{S}_a(\vec{s} = \vec{0})$ is the sphere of radius a centered on the origin of the $\vec{s}$ system (*i.e.*, the same as $S_a(\vec{r})$, the sphere of radius a centered on $\vec{r}$ in the $\vec{r}\,'$ coordinate system). Because the integration is over a sphere of radius a, $s = a$ is fixed in the integrand. This permits us to both pull a factor $s^2 = a^2$ outside the integral, canceling the factor of a^2 in the prefactor, and to pull the radial derivative outside the integral and turn it into a derivative with respect to a. Thus:

$$0 = \frac{1}{4\,\pi\,a^2}\,a^2 \int_{\mathcal{S}_a(\vec{s}=\vec{0})} d\Omega_s \left.\frac{\partial V}{\partial s}\right|_{s=a} = \frac{1}{4\,\pi}\frac{\partial}{\partial a}\int_{\mathcal{S}_a(\vec{s}=\vec{0})} d\Omega_s\, V(\vec{s}) \tag{3.9}$$

Note that $\partial/\partial s$ becomes $\partial/\partial a$ when we move it outside the integral because the limits of integration, $S_a(\vec{r})$, imply s should be evaluated at a when the integral is done: s no longer exists once the integral is done.

Thus, the integral must be a constant

$$C = \frac{1}{4\,\pi} \int_{\mathcal{S}_a(\vec{s}=\vec{0})} d\Omega_s\, V(\vec{s}) = \frac{1}{4\,\pi\,a^2} \int_{\mathcal{S}_a(\vec{r})} da'\, V(\vec{r}\,') \tag{3.10}$$

where we switched the variable of integration back to $\vec{r}\,'$ and we reinserted a^2. The right side is just the average of V over the sphere of radius a centered at $\vec{r}$. Since this holds for any a, it must hold as $a \to 0$, which tells us $C = V(\vec{r})$. So, we have

$$V(\vec{r}) = \frac{1}{4\,\pi\,a^2} \int_{\mathcal{S}_a(\vec{r})} da'\, V(\vec{r}\,') \tag{3.11}$$

- **V can have no local maxima or minima in the region of interest**
  The averaging property makes the proof of this property trivial: if there were such a candidate maximum (minimum), simply draw a sphere around it. Because the point is a maximum (minimum) there must be some radius of the sphere for which the values of all the points on the sphere are less than (greater than) the value at the candidate maximum (minimum). The average over this sphere is therefore less than (greater than) the value at the candidate maximum (minimum). This contradicts the above averaging property.

  One could also prove this by a technique similar to the 1D case, calculating $\vec{\nabla} V$ at any point $\vec{r}\,'$ in the region by doing a line integral of Laplace's Equation from the candidate extremum $\vec{r}$ to that point. Since $\vec{\nabla} V$ vanishes at the candidate extremum (because it is an extremum of V), and the integrand ($\nabla^2 V$) of the line integral vanishes by Laplace's Equation, $\vec{\nabla} V$ vanishes at $\vec{r}\,'$.

Uniqueness Theorem

Before obtaining a solution of Laplace's and Poisson's Equations, we prove some uniqueness theorems we will need. This section draws from Jackson §1.8 and §1.9.

Green's Identities and Theorem

First, some mathematical preliminaries. Let us apply the divergence theorem to the function $\phi\vec{\nabla}\psi$ where $\phi(\vec{r})$ and $\psi(\vec{r})$ are arbitrary functions:

$$\oint_{\mathcal{S}} da\,\hat{n}\cdot\left(\phi\vec{\nabla}\psi\right) = \int_{\mathcal{V}(\mathcal{S})} d\tau\,\vec{\nabla}\cdot\left(\phi\vec{\nabla}\psi\right)$$

This yields *Green's First Identity*:

$$\boxed{\oint_{\mathcal{S}} da\,\phi\,\hat{n}\cdot\vec{\nabla}\psi = \int_{\mathcal{V}(\mathcal{S})} d\tau\left[\phi\nabla^2\psi + \vec{\nabla}\phi\cdot\vec{\nabla}\psi\right]} \tag{3.12}$$

The function $\hat{n}\cdot\vec{\nabla}\psi$ is the *normal gradient* of ψ because it is the projection of the gradient of ψ along the direction normal to the surface. If we exchange ϕ and ψ and then difference the two versions, we have *Green's Second Identity* or *Green's Theorem*:

$$\boxed{\oint_{\mathcal{S}} da\left[\phi\,\hat{n}\cdot\vec{\nabla}\psi - \psi\,\hat{n}\cdot\vec{\nabla}\phi\right] = \int_{\mathcal{V}(\mathcal{S})} d\tau\left[\phi\nabla^2\psi - \psi\nabla^2\phi\right]} \tag{3.13}$$

Types of Boundary Conditions

We shall see in the proof of the Uniqueness Theorem that three types of boundary conditions are permitted:

- *Dirichlet boundary condition*
  In this case, the value of the potential $V(\vec{r})$ is specified on all bounding surfaces. This is the most typical experimentally realized situation, where we attach a number of conductors to voltage sources to set their voltages.
- *Neumann boundary condition*
  In this case, the value of the normal derivative of the potential, $\hat{n} \cdot \vec{\nabla} V(\vec{r})$, is specified on the boundary. An example of such a condition is specification of the electric field (or, equivalently, the surface charge density) at the surfaces of a set of conductors; since the tangential electric field vanishes at these surfaces, the normal electric field fully defines the electric field at the conductors.
- *Mixed boundary conditions*
  Dirichlet in some places, Neumann in others, is allowed as long as both are not specified at the same place.

If the volume under consideration is not bounded by a surface on which we specify the boundary conditions, then we must also specify a boundary condition at infinity.

The proof of the Uniqueness Theorem will not show why only one of these types of boundary conditions may be specified. That proof will be provided soon, in §3.4.1.

Generic Uniqueness Proof for Poisson's Equation

We will use proof by contradiction.

Suppose we have specified one of the above three types of boundary conditions. Assume that, for a particular given charge distribution $\rho(\vec{r})$, there are two independent solutions $V_1(\vec{r})$ and $V_2(\vec{r})$ of Poisson's Equation that both satisfy the boundary condition. Let $V_3 = V_1 - V_2$. Since the charge distribution is the same, $\nabla^2 V_1 = -\rho/\epsilon_o = \nabla^2 V_2$ and thus $\nabla^2 V_3 = 0$: V_3 satisfies Laplace's Equation. By a similar differencing argument, V_3 either satisfies the Dirichlet boundary condition $V_3(\vec{r} \in \mathcal{S}) = 0$, the Neumann boundary condition $\hat{n} \cdot \vec{\nabla} V_3(\vec{r} \in \mathcal{S}) = 0$, or a mixed boundary condition of these types. If we apply Green's first identity with $\phi = \psi = V_3$, we have

$$\oint_{\mathcal{S}} da\, V_3\, \hat{n} \cdot \vec{\nabla} V_3 = \int_{\mathcal{V}(\mathcal{S})} d\tau \left(V_3 \nabla^2 V_3 + \vec{\nabla} V_3 \cdot \vec{\nabla} V_3 \right) \tag{3.14}$$

The left side vanishes because of the boundary condition (any type). The first term on the right side vanishes by Laplace's Equation. Thus, we have

$$\int_{\mathcal{V}(\mathcal{S})} d\tau\, |\vec{\nabla} V_3|^2 = 0 \quad \Longrightarrow \quad \vec{\nabla} V_3(\vec{r}) = 0 \quad \Longrightarrow \quad V_3 = \text{constant} \tag{3.15}$$

where we take the second step because the integrand is nonnegative. This result implies that our two candidate solutions $V_1(\vec{r})$ and $V_2(\vec{r})$ differ by at most a constant. Hence, uniqueness is proven.

Special Cases of Uniqueness Theorem

Given the above, we may state/prove three special cases of the uniqueness theorem, the ones given in Griffiths:

- *The solution to Laplace's Equation in some volume $\mathcal{V}$ is uniquely specified if V is specified on the boundary surface $\mathcal{S}(\mathcal{V})$.*
  This is the above uniqueness theorem with $\rho = 0$ in $\mathcal{V}$ and a Dirichlet boundary condition on $\mathcal{S}(\mathcal{V})$.

- *The solution to Poisson's Equation in some volume $\mathcal{V}$ is uniquely specified if $\rho(\vec{r})$ is specified throughout the region and V is specified on the boundary surface $\mathcal{S}(\mathcal{V})$.*
  This is the above uniqueness theorem with arbitrary $\rho(\vec{r})$ in $\mathcal{V}$ and a Dirichlet boundary condition on $\mathcal{S}(\mathcal{V})$.

- *In a volume $\mathcal{V}$ surrounded by conductors at the surface(s) $\mathcal{S}(\mathcal{V})$ and containing a specified charge density $\rho(\vec{r})$, the electric field is uniquely determined if the total charge on each conductor is specified.*
  This one is not as obvious, but we can show that this BC yields the same input to the Uniqueness Theorem derivation as the other BCs we have specified.

Let each conductor i have surface $\mathcal{S}_i$ and charge Q_i. Since we know the surface charge density on each conductor is related to the normal component of the electric field at that conductor, we may see

$$\oint_{\mathcal{S}_i} da\, \hat{n}(\vec{r}) \cdot \vec{E}(\hat{r}) = \frac{1}{\epsilon_o} \oint_{\mathcal{S}_i} da\, \sigma(\vec{r}) = \frac{1}{\epsilon_o} Q_i \tag{3.16}$$

Now, as before, let's assume that there are two different solutions $V_1(\vec{r})$ and $V_2(\vec{r})$ and their difference is $V_3 = V_2 - V_1$. Let's evaluate the left-hand side of Equation 3.14 for the BC we are specifying here:

$$\oint_{\mathcal{S}} da\, V_3\, \hat{n} \cdot \vec{\nabla} V_3 = -\sum_i \oint_{\mathcal{S}_i} da\, V_3\, \hat{n} \cdot \vec{E}_3 = -\sum_i V_{3,i} \oint_{\mathcal{S}_i} da\, \hat{n} \cdot \vec{E}_3 \tag{3.17}$$

where we were able to pull V_3 out of the integrals because V_1 and V_2 have equipotentials on each surface and so therefore does V_3 (with values $V_{3,i}$, which we do not need to know). The surface integral of the normal component of $\vec{E}_3$ over each $\mathcal{S}_i$ vanishes because, as we indicated above, specifying Q_i specifies this surface integral to be the same for $\vec{E}_1$ and $\vec{E}_2$, so the surface integral vanishes for $\vec{E}_3 = \vec{E}_2 - \vec{E}_1$. Thus, the LHS of Equation 3.14 also vanishes for this BC, and so the remainder of the proof of uniqueness carries through.

Note how this proof relied on the boundary surfaces being conductors! Knowing the total charges on nonconducting boundary surfaces would not be sufficient.

Method of Images

Overview: The Basic Idea of Method of Images

The method of images uses the concept of uniqueness of solutions to Poisson's Equation. Basically, given a physical setup involving a true charge distribution $\rho(\vec{r})$ and Dirichlet boundary conditions for some volume $\mathcal{V}$, one tries to replace the region *outside* of $\mathcal{V}$ with an *image charge* distribution $\rho_{image}(\vec{r})$ such that, when the image charge's potential is summed with that of $\rho(\vec{r})$, the potential on the boundary is the same as that specified by the Dirichlet BC.

The technique works because of the uniqueness theorem: since the potential due to the image and original charges matches the boundary conditions and satisfies Poisson's Equation with the same source term inside $\mathcal{V}$, it is *the* solution to Poisson's Equation for that source term, that volume $\mathcal{V}$, and that choice of boundary conditions.

The imagined charge distribution is called "image charge" because, at least in the example of the boundary condition being imposed by the presence of a conductor, the image charge is a (possibly distorted) mirror image, through the boundary, of the original charge distribution $\rho(\vec{r})$. "Image charge" is also used (somewhat erroneously) to refer to the surface charge induced on a conducting boundary that sources the potential that one models as due to the image charge.

Note that the image charge must be placed outside the volume $\mathcal{V}$ because we may not change $\rho(\vec{r})$ inside $\mathcal{V}$; that would change the problem we are trying to solve.

We will see later how the potential due to the image charge distribution (the induced surface charge) is a component of the particular problem's Green Function.

A Point Charge near a Grounded Infinite Conducting Plane

For a system with the point charge q at $d\,\hat{z}$ above a conducting plane at $z = 0$ with $V = 0$, and considering the volume $\mathcal{V}$ consisting of the $z > 0$ half-space, the appropriate image charge is $-q$ at $-d\,\hat{z}$. By symmetry, the (Dirichlet) boundary condition $V = 0$ at $z = 0$ is met. Thus, the solution for $V(\vec{r})$ for $\vec{r} \in \mathcal{V}$ (the $z > 0$ half-space) is

$$V(\vec{r}) = \frac{1}{4\,\pi\,\epsilon_o}\left[\frac{q}{\sqrt{x^2+y^2+(z-d)^2}} - \frac{q}{\sqrt{x^2+y^2+(z+d)^2}}\right] \tag{3.18}$$

The potential clearly satisfies $V(z = 0) = 0$ (and $V(r \to \infty) \to 0$). Let's use this solution to do some other calculations:

- *Induced surface charge*
  This we can calculate by recognizing that it is given by the change in the normal component of the electric field at the conducting boundary. Since $\vec{E} = -\vec{\nabla} V$,

$$\begin{aligned}\sigma &= -\,\epsilon_o \left.\frac{\partial V}{\partial z}\right|_{z=0} = \frac{q}{4\,\pi}\left[\frac{z-d}{(x^2+y^2+(z-d)^2)^{3/2}} - \frac{z+d}{(x^2+y^2+(z+d)^2)^{3/2}}\right]\Bigg|_{z=0} \\ &= -\frac{q}{2\,\pi}\frac{d}{(x^2+y^2+d^2)^{3/2}}\end{aligned} \tag{3.19}$$

  We will treat the surface charge density and the normal component of the electric field (the normal gradient of the potential) as almost equivalent going forward.

We can calculate the total induced surface charge:

$$Q_{ind} = \int_0^\infty r\,dr \int_0^{2\pi} d\phi \frac{-q\,d}{2\pi} \frac{1}{(r^2+d^2)^{3/2}} = q\,d \left. \frac{1}{\sqrt{r^2+d^2}} \right|_0^\infty = -q \quad (3.20)$$

This is an example of an important general theorem: *The total induced surface charge is equal to the image charge, or to the negative of the real charge, or to some combination of the two, depending on the geometry, by Gauss's Law*. Because of the mirror symmetry of this problem, the two cases are degenerate, so this is not a particularly illustrative example of the theorem. Furthermore, because the volumes and surfaces one must integrate over are infinite, Gauss's Law cannot be applied to such a geometry. We'll return to this theorem in our next example where there is no such issue.

- *Force on the point charge*
  The induced charge is opposite in sign to the real charge, so the two are attracted to each other. We can calculate the force by taking the gradient of the potential due to the *image charge only* (because the real charge does not feel a force due to its own potential). Since the image charge's potential is just that of a point charge, calculating the force is straightforward:

$$\vec{F} = q\,\vec{E}_{image\ charge}(d\,\hat{z}) = -\frac{1}{4\pi\epsilon_o}\frac{q^2}{(2d)^2}\hat{z} \quad (3.21)$$

This is equivalent to just calculating the force on the real charge exerted by the image charge, which is in general a valid approach. Whether to calculate the image charge potential and take the gradient or calculate the image charge force is a matter of choice and convenience.

- *Electric potential energy*
  Here we have to be more careful because potential energy is not linear in charge, and, moreover, because the induced charge depends on the original point charge. Let's figure this out by calculating the work one would have to do *against* the electric force (*i.e.*, the mechanical force F_m doing the work is opposite in sign to the attractive electric force F_e) to bring q from $z = d$ to $z = \infty$.

$$U = -\int_d^\infty (-F_e(z))\, dz = -\frac{1}{4\,\pi\,\epsilon_o}\frac{q^2}{4}\int_d^\infty \frac{dz}{z^2} = -\frac{1}{4\,\pi\,\epsilon_o}\frac{q^2}{4\,d} \tag{3.22}$$

Note that this result is half what one would get for the potential energy of two equal and opposite point charges separated by a distance $2d$:

$$U_{alt} = -\frac{1}{4\,\pi\,\epsilon_o}\frac{q^2}{2d} \tag{3.23}$$

There are two ways to understand this. The first is to recognize that, unlike in the case of two point charges, no energy is gained or lost in moving the negative charge because it is in the conductor, where $V = 0$ and thus $q\,V = 0$ everywhere. The second is to recognize that the above expression is the energy stored in all of space in the field of two point charges, but, in this case, the field is only real in the $z > 0$ half-space and so the integrated energy is reduced by a factor of 2.

Lecture 6:

Advanced Electrostatics II:
Method of Images (cont.)
Green Functions for Poisson's Equation

Date Revised: 2024/01/25 05:30
Revised lecture break
Date Given: 2024/01/23

A Point Charge near a Grounded, Conducting Sphere

Consider a conducting sphere of radius R centered on the origin and held at $V = 0$. Place a point charge at $a\,\hat{z}$ with $a > R$ so the point charge is outside the sphere. We would like to know the potential in the volume $\mathcal{V}$ *outside* the conducting sphere, which is the volume in which the point charge sits.

By symmetry, the appropriate image charge must be on the z axis. Let its value be q' and its position be $b\,\hat{z}$, where b may be positive or negative. We can find q' and b by requiring that $V = 0$ at $\vec{r} = \pm R\,\hat{z}$:

$$0 = V(+R\,\hat{z}) = \frac{1}{4\,\pi\,\epsilon_o}\left[\frac{q}{a-R} + \frac{q'}{R-b}\right]$$

$$0 = V(-R\,\hat{z}) = \frac{1}{4\,\pi\,\epsilon_o}\left[\frac{q}{a+R} + \frac{q'}{R+b}\right]$$

$$\Longrightarrow \qquad q' = -q\,\frac{R}{a} \neq -q \qquad\qquad b = \frac{R^2}{a} \tag{3.24}$$

(This is an example of how one does not always need to consider the generic case; these special cases at the two poles give us the information we need.) We see that both values are always physically reasonable because $R < a$. In particular, $b < R$ so the image charge remains outside $\mathcal{V}$ (*i.e.*, inside the sphere), as we expect. Note that $q' \neq -q$!

The potential at a point $(r \geq R, \theta, \phi)$ is found by summing the potentials of the real charge and the image charge:

$$V(r \geq R, \theta, \phi) = \frac{q}{4\,\pi\,\epsilon_o}\left[\frac{1}{|\vec{r} - a\,\widehat{z}|} - \frac{R/a}{|\vec{r} - \frac{R^2}{a}\,\widehat{z}|}\right] \tag{3.25}$$

$$= \frac{q}{4\,\pi\,\epsilon_o}\left[\frac{1}{\sqrt{r^2\sin^2\theta + (a - r\,\cos\theta)^2}} - \frac{R/a}{\sqrt{r^2\sin^2\theta + (\frac{R^2}{a} - r\,\cos\theta)^2}}\right] \tag{3.26}$$

We can use the above expression to see that the boundary condition $V(r = R) = 0$ is satisfied in full generality:

$$V(r = R, \theta, \phi) = \frac{q}{4\,\pi\,\epsilon_o}\left[\frac{1}{\sqrt{R^2 sin^2\theta + (a - R\,\cos\theta)^2}} - \frac{R/a}{\sqrt{R^2 sin^2\theta + (\frac{R^2}{a} - R\,\cos\theta)^2}}\right]$$

$$= \frac{q}{4\,\pi\,\epsilon_o}\left[\frac{1}{\sqrt{R^2\sin^2\theta + (a - R\cos\theta)^2}} - \frac{1}{\sqrt{a^2\sin^2\theta + (R - a\cos\theta)^2}}\right]$$

$$= 0 \tag{3.27}$$

Let's calculate the induced surface charge from $\hat{n} \cdot \vec{\nabla} V = \partial V / \partial r$:

$$\sigma = -\epsilon_o \left. \frac{\partial V}{\partial r} \right|_{r=R} \tag{3.28}$$

$$= \frac{q}{4\pi} \left[\frac{R \sin^2\theta - (a - R\cos\theta)\cos\theta}{(R^2 sin^2\theta + (a - R\cos\theta)^2)^{3/2}} - \frac{R}{a} \frac{R\sin^2\theta - (\frac{R^2}{a} - R\cos\theta)\cos\theta}{\left(R^2 sin^2\theta + (\frac{R^2}{a} - R\cos\theta)^2\right)^{3/2}} \right]$$

$$= \frac{q}{4\pi} \left[\frac{R - a\cos\theta}{(R^2 + a^2 - 2aR\cos\theta)^{3/2}} - \frac{a^2}{R^2} \frac{R - \frac{R^2}{a}\cos\theta}{(a^2 + R^2 - 2aR\cos\theta)^{3/2}} \right]$$

$$= \frac{q}{4\pi} \frac{R(1 - \frac{a^2}{R^2})}{(R^2 + a^2 - 2aR\cos\theta)^{3/2}} = -\frac{q}{4\pi R^2} \frac{R}{a} \frac{1 - \frac{R^2}{a^2}}{\left(1 + \frac{R^2}{a^2} - 2\frac{R}{a}\cos\theta\right)^{3/2}}$$

One can show by integration that the total induced charge is q'. In this geometry, this makes sense because the volume enclosed by a surface integral of electric field flux at the boundary *encloses* the volume containing the image charge. This example illustrates one case of the theorem stated earlier; in this case, the total induced surface charge is equal to the image charge. We will see other cases illustrated in the next example.

The force on the point charge and the electric potential energy can be calculated in a manner similar to that used for the conducting plane.

Calculate the force by taking gradient of electric field:

$$\vec{F} = q\,\vec{E}_{image\ \ charge}(z=a) = -q\,\vec{\nabla}\,\frac{q}{4\,\pi\,\epsilon_o}\,\frac{-R/a}{\left|\vec{r}-\frac{R^2}{a}\,\hat{z}\right|}\Bigg|_{\vec{r}=a\,\hat{z}}$$

$$= \frac{q^2}{4\,\pi\,\epsilon_o}\,\hat{z}\,\frac{\partial}{\partial z}\,\frac{R/a}{\left[x^2+y^2+\left(z-\frac{R^2}{a}\right)^2\right]^{1/2}}\Bigg|_{z=a}$$

$$= -\frac{q^2}{4\,\pi\,\epsilon_o}\,\frac{R}{a^3}\,\frac{1}{\left[1-\frac{R^2}{a^2}\right]^2}\,\hat{z}$$

Again, this is the same we would have obtained by directly calculating the force on the real charge from the image charge:

$$\vec{F} = \frac{q\left(-q\,\frac{R}{a}\right)}{4\,\pi\,\epsilon_o}\,\frac{1}{\left[a-\frac{R^2}{a}\right]^2}\,\hat{z} = -\frac{q^2}{4\,\pi\,\epsilon_o}\,\frac{R}{a^3}\,\frac{1}{\left[1-\frac{R^2}{a^2}\right]^2}\,\hat{z}$$

Both of these match Jackson Equation 2.6.

We calculate the potential energy by line integral of the force:

$$U = \int_{\infty}^{a\,\widehat{z}} d\vec{\ell}\cdot\vec{F}(\vec{r}) = \int_{\infty}^{a}(-dz)\,\frac{-q^2}{4\,\pi\,\epsilon_o}\,\frac{R}{z^3}\,\frac{1}{\left[1-\frac{R^2}{z^2}\right]^2}$$

$$= \frac{q^2}{4\,\pi\,\epsilon_o}\,R\int_{\infty}^{a}\frac{z\,dz}{\left[z^2-R^2\right]^2} = -\frac{q^2}{8\,\pi\,\epsilon_o}\,\frac{R}{a^2-R^2}$$

Note that, if we calculate the potential energy from the image charge and real charge, we get the same factor of two error we saw above for the point charge and the plane:

$$U_{alt} = \frac{q\left(-q\,\frac{R}{a}\right)}{4\,\pi\,\epsilon_o}\,\frac{1}{\left|a\,\widehat{z}-\frac{R^2}{a}\,\widehat{z}\right|} = -\frac{q^2}{4\,\pi\,\epsilon_o}\,\frac{R}{a^2-R^2} = 2\,U \qquad (3.29)$$

which is incorrect for the same reasons as given before.

Some Related Examples

These are drawn from Jackson Chapter 2.

Example 3.1: Point charge inside a spherical volume with a conducting boundary

The geometry of this problem is like the last one, except the point charge is *inside* the spherical boundary, $a < R$, and everything outside the boundary is conductor. One can show that the solution is identical: same formula for image charge value and position, same induced surface charge density. *However,* strangely enough, the total surface charge is now just $-q$!

Mathematically, this is because the evaluation of the integral depends on whether $R < a$ or $R > a$. (There is a power series expansion involved, which must be done differently in the two cases.)

Physically, this is because the calculation of the total induced surface charge via Gauss's Law must be done differently. One method is to use a spherical surface just outside the boundary, so it is in the conducting volume where the field vanishes. This implies that the sum of the real and induced charge vanishes, so the induced charge is the negative of the real charge.

The other method is to put the surface just inside the boundary. Now, the charge enclosed is only the real charge. As the surface approaches the boundary, though, the flux integral is equal to the negative of the integral of the surface charge density (up to ϵ_o) because the electric field near a conductor is σ/ϵ_o (with the negative because the field is pointed inward). So this tells us the total induced surface charge is the negative of the real charge too.

Thus, we see illustrated another case of the theorem we stated earlier, that the total induced surface charge is the image charge, the negative of the real charge, or some combination of the two. Which one depends on the geometry: is the boundary outside the volume of interest, inside, or some combination of the two?

In the case of the point charge outside the conducting sphere, we noted that the Gauss's Law calculation, with the Gaussian sphere just inside the volume $\mathcal{V}$ (i.e., having radius infinitesimally larger than a), yields $q' \neq -q$. The distinction is whether the volume $\mathcal{V}$ of interest is "outside" the boundary (neglecting the boundary at infinity) as in the previous case or "inside" the boundary as in this case.

(In the previous case, the Gauss's Law calculation outside $\mathcal{V}$ (i.e., using a Gaussian sphere of radius less than a) yields no useful information because the sphere doesn't contain the induced surface charge. The flux through such a sphere vanishes because the field is zero inside the conductor, which just tells us that all the induced surface charge resides, well, on the surface.)

Example 3.2: Point charge in the presence of a conducting sphere at fixed potential V_0

We can treat this by superposition. Consider first bringing the sphere up to the desired potential in the absence of the point charge, then bringing the point charge in from infinity to its final position $a\,\hat{z}$. We can use the grounded-case solution for the latter part because it has $V = 0$ on the sphere and $V \to 0$ at infinity, so the sum of it and the solution for the $V \neq 0$ sphere alone satisfies the boundary condition of the problem of the point charge near the $V \neq 0$ sphere, and thus it must be the correct solution. (Note the use of the principle of superposition for the potential.)

What is the solution for the $V \neq 0$ sphere on its own? Certainly, the sphere is an equipotential with the desired value V_0. By symmetry (remember, the point charge is not present for this problem), the charge is uniformly distributed on the surface. Thus, we can apply Gauss's Law to the problem, which tells us that the potential of the sphere, for $r > R$, is identical to that of a point charge at the origin. To figure out the value of the point charge, we require that the point charge's potential match the boundary condition:

$$\frac{q_0}{4\,\pi\,\epsilon_o\,R} = V_0 \quad \Longrightarrow \quad q_0 = 4\,\pi\,\epsilon_o\,V_0\,R \quad \Longrightarrow \quad V(r) = V_0\,\frac{R}{|\vec{r}|} \tag{3.30}$$

Finally, we add the two solutions together:

$$V(r \geq R, \theta, \phi) = \frac{q}{4\,\pi\,\epsilon_o}\left[\frac{1}{|\vec{r} - a\,\hat{z}|} - \frac{R/a}{|\vec{r} - \frac{R^2}{a}\,\hat{z}|}\right] + V_0\,\frac{R}{|\vec{r}|} \tag{3.31}$$

Example 3.3: Point charge in the presence of a charged, insulated, conducting sphere

We can solve this using the solution we just calculated along with the principle of superposition (again!). Suppose we want to have a charge Q on the sphere. This is the same as first bringing the point charge q in while the sphere is grounded, disconnecting the grounding wire, adding $Q - q'$ ($> Q$ for $q > 0$), which causes the sphere to float to some nonzero voltage, and then connecting to a voltage source with that voltage. This situation is identical to the situation we just studied if we require

$$q_0 = Q - q' \qquad \Longrightarrow \qquad V_0 = \frac{q_0}{4\,\pi\,\epsilon_o\,R} = \frac{Q - q'}{4\,\pi\,\epsilon_o\,R} = \frac{Q + q\,\frac{R}{a}}{4\,\pi\,\epsilon_o\,R} \tag{3.32}$$

Plugging this into solution for the sphere held at V_0 gives

$$V(r \geq R, \theta, \phi) = \frac{q}{4\,\pi\,\epsilon_o}\left[\frac{1}{|\vec{r} - a\,\widehat{z}|} - \frac{R/a}{|\vec{r} - \frac{R^2}{a}\,\widehat{z}|}\right] + \frac{Q + q\frac{R}{a}}{4\,\pi\,\epsilon_o\,|\vec{r}|} \tag{3.33}$$

Notice that this reduces to our original point charge near a sphere solution not when $Q = 0$ but rather when $Q = q' = -q\,R/a$, which is the charge that must flow onto the sphere for it to stay at $V = 0$ (*i.e.*, grounded).

Formal Solution to Poisson's Equation: Green Functions

The remaining material in this section of the notes is based on Jackson §1.10.

Integral Equation for the Electric Potential

Can we solve Poisson's Equation? Sort of. We can convert it from a differential equation for V in terms of ρ (with boundary conditions separately specified) to an integral equation for V in terms of ρ with the need for the boundary conditions quite explicit. It is still not a closed-form solution for V in terms of ρ and the boundary conditions, but it helps us to frame the problem of finding solutions for V in a different manner that is helpful.

We obtain this equation by applying Green's Theorem (Equation 3.13) with $\phi(\vec{r}\,') = V(\vec{r}\,')$ and $\psi(\vec{r}\,') = |\vec{r} - \vec{r}\,'|^{-1}$. Note that $\vec{r}\,'$ is the variable we integrate over; $\vec{r}$ is considered a constant for the purposes of the Green's Theorem integrals.

$$\int_{\mathcal{V}(\mathcal{S})} d\tau' \left[V(\vec{r}\,') \nabla^2_{\vec{r}\,'} \frac{1}{|\vec{r} - \vec{r}\,'|} - \frac{1}{|\vec{r} - \vec{r}\,'|} \nabla^2_{\vec{r}\,'} V(\vec{r}\,') \right]$$
$$= \oint_{\mathcal{S}} da \left[V(\vec{r}\,') \, \widehat{n}(\vec{r}\,') \cdot \vec{\nabla}_{\vec{r}\,'} \frac{1}{|\vec{r} - \vec{r}\,'|} - \frac{1}{|\vec{r} - \vec{r}\,'|} \, \widehat{n}(\vec{r}\,') \cdot \vec{\nabla}_{\vec{r}\,'} V(\vec{r}\,') \right] \quad (3.34)$$

We reduce this by making use of the very important relation

$$\boxed{\nabla^2_{\vec{r}\,'} \frac{1}{|\vec{r}\,' - \vec{r}|} = -4\,\pi\,\delta(\vec{r} - \vec{r}\,')} \tag{3.35}$$

which is seen by combining Equations 2.53 and 2.33 with $\vec{r} \leftrightarrow \vec{r}\,'$:

$$\vec{\nabla}_{\vec{r}\,'} \frac{1}{|\vec{r}\,' - \vec{r}|} = -\frac{\vec{r}\,' - \vec{r}}{|\vec{r}\,' - \vec{r}|^3} \quad \text{and} \quad \vec{\nabla}_{\vec{r}\,'} \cdot \frac{\vec{r}\,' - \vec{r}}{|\vec{r}\,' - \vec{r}|^3} = 4\,\pi\,\delta(\vec{r}\,' - \vec{r}) = 4\,\pi\,\delta(\vec{r} - \vec{r}\,')$$

Using the above expression for the Laplacian of $|\vec{r} - \vec{r}\,'|^{-1}$, doing the integral over the delta function, applying Poisson's Equation, moving the second term on the right side to the left side, and multiplying everything by $-\frac{1}{4\,\pi}$ yields, *now only for* $\vec{r} \in \mathcal{V}(\mathcal{S})$:

$$V(\vec{r} \in \mathcal{V}(\mathcal{S})) = \frac{1}{4\,\pi\,\epsilon_o} \int_{\mathcal{V}(\mathcal{S})} d\tau' \frac{\rho(\vec{r}\,')}{|\vec{r} - \vec{r}\,'|} + \frac{1}{4\,\pi} \oint_{\mathcal{S}} da' \left[\frac{1}{|\vec{r} - \vec{r}\,'|}\, \widehat{n}(\vec{r}\,') \cdot \vec{\nabla}_{\vec{r}\,'} V(\vec{r}\,') - V(\vec{r}\,')\, \widehat{n}(\vec{r}\,') \cdot \vec{\nabla}_{\vec{r}\,'} \frac{1}{|\vec{r} - \vec{r}\,'|} \right] \tag{3.36}$$

(The left side vanishes for $\vec{r} \notin \mathcal{V}(\mathcal{S})$ because the integral was over $\vec{r}\,' \in \mathcal{V}(\mathcal{S})$).

This is a formal equation for the electric potential. The boundary conditions are present on the right side: in the case of Dirichlet, we specify $V(\vec{r}')$ for $\vec{r}' \in \mathcal{S}$, while in the case of Neumann, we specify $\hat{n}(\vec{r}') \cdot \vec{\nabla}_{\vec{r}'} V(\vec{r}')$ for $\vec{r}' \in \mathcal{S}$. Our Uniqueness Theorem says we should only need to specify one or the other at any given point on the boundary. In fact, since the Uniqueness Theorem says that knowing one specifies the other (knowing one gives the full solution, which determines the other), we don't have the freedom to specify both independently! Knowing both essentially requires knowing the solution to the problem. For example, if we consider the simplest possible case of specifying an equipotential on the boundary, then knowing the other boundary term requires knowing the normal gradient of the potential at the boundary, which is equivalent to knowing the surface charge density on the boundary. We would not be able to guess this except in cases with sufficient symmetry.

Therefore, this is not a closed-form solution but rather an integral equation for $V(\vec{r}')$ for $\vec{r}' \in \mathcal{V}(\mathcal{S}) \cup \mathcal{S}$: the boundary condition does not provide everything on the right side, but, if we know the solution, it will satisfy the equation.

Note that, in the limit of $\mathcal{S} \to \infty$ and $V(r \to \infty) \propto 1/r \to 0$, the integrand of the surface integral falls off as r^{-3} and so the surface term vanishes and we recover the usual Coulomb's Law expression for $V(\vec{r})$, Equation 2.51. That is, in a situation where we know the behavior of both surface terms is trivial, the equation does provide a closed-form expression for $V(\vec{r})$ in terms of $\rho(\vec{r})$.

So far, however, this integral equation is not very useful. Once we have introduced the concept of Green Functions, we will see its utility.

The Concept of Green Functions

Suppose we have the generalization of Poisson's Equation, the linear partial differential equation

$$O_{\vec{r}} f(\vec{r}) = g(\vec{r}) \tag{3.37}$$

where $O_{\vec{r}}$ is a linear partial differential operator taking derivatives with respect to the coordinate $\vec{r}$, f is a generalized potential, and g is a generalized source function. Poisson's Equation is an example, with $O_{\vec{r}} = -\epsilon_o \nabla^2$, $f(\vec{r}) = V(\vec{r})$, and $g(\vec{r}) = \rho(\vec{r})$. Is there a general approach for finding f given g?

Yes, there is, it is called the Green Function approach. The basic idea is to find the "impulse" response function for the differential equation: the generalized potential one gets if one has a point-like source. Given the impulse response function, and the linearity of $O_{\vec{r}}$, one can obtain the generalized potential for an arbitrary source function by convolving the impulse response function with that source function.

Mathematically, the impulse response function, or Green Function, is the function $G(\vec{r}, \vec{r}')$ that solves the equation

$$O_{\vec{r}}\, G(\vec{r}, \vec{r}') = \delta(\vec{r} - \vec{r}') \tag{3.38}$$

meaning that $G(\vec{r}, \vec{r}')$ calculates the generalized potential at the point $\vec{r}$ for a point source of size $q = 1$ at the position $\vec{r}'$ (*i.e.*, the total source charge recovered by integrating over the source function is 1). If such a G exists, then, for an arbitrary source function $g(\vec{r})$, G gives us the following solution $f(\vec{r})$ to the generalized linear partial differential equation, Equation 3.37:

$$f(\vec{r}) = \int d\tau'\, G(\vec{r}, \vec{r}')\, g(\vec{r}') \tag{3.39}$$

We can check that Equation 3.37 is satisfied by this solution by applying the operator:

$$O_{\vec{r}}\, f(\vec{r}) = O_{\vec{r}} \int d\tau'\, G(\vec{r}, \vec{r}')\, g(\vec{r}') = \int d\tau' \left[O_{\vec{r}}\, G(\vec{r}, \vec{r}') \right] g(\vec{r}') \tag{3.40}$$

$$= \int d\tau' \delta(\vec{r} - \vec{r}')\, g(\vec{r}') = g(\vec{r}) \tag{3.41}$$

Note how this check relied on the linearity of $O_{\vec{r}}$, which allowed us to bring it inside the integral. Assuming solutions to the generalized linear partial differential equation are unique (true for Poisson's Equation), the Green Function is the *only* solution we need to find.

General Discussion of Green Functions for Poisson's Equation

Let's consider the simplest possible case, that in which there is no bounding surface and the potential vanishes at infinity. We can read the Green Function off by rewriting our usual expression for the potential for this boundary condition, Equation 2.51, in the same form as Equation 3.39:

$$V(\vec{r}) = \frac{1}{4\,\pi\,\epsilon_o}\int_{\mathcal{V}} d\tau'\,\frac{\rho(\vec{r}\,')}{|\vec{r}-\vec{r}\,'|} = \int_{\mathcal{V}} d\tau'\,G(\vec{r},\vec{r}\,')\,\rho(\vec{r}\,') \tag{3.42}$$

Therefore, the Green Function for Poisson's Equation is

$$G(\vec{r},\vec{r}\,') = \frac{1}{4\,\pi\,\epsilon_o}\,\frac{1}{|\vec{r}-\vec{r}\,'|} \qquad \text{if} \qquad \mathcal{V} = \text{all space}, V(r\to\infty)\to 0 \tag{3.43}$$

More generally — *i.e.*, for a more complex boundary condition — Poisson's Equation implies that its Green Function must decompose into the form

$$G(\vec{r},\vec{r}\,') = \frac{1}{4\,\pi\,\epsilon_o}\,\frac{1}{|\vec{r}-\vec{r}\,'|} + F(\vec{r},\vec{r}\,') \qquad \text{with} \qquad \nabla^2_{\vec{r}} F(\vec{r},\vec{r}\,') = 0 \tag{3.44}$$

where the first term provides the right side of Poisson's Equation but the second term is not only allowed by Poisson's Equation but, we will see, is crucial for satisfying the boundary conditions for any situation except the trivial one noted above, that of the potential vanishing at infinity. The F term plays multiple roles, depending on the type of boundary condition, and we will explain those roles later. Finding G thus consists of finding F.

We note that both G and F are symmetric in their arguments, $G(\vec{r}\,',\vec{r}) = G(\vec{r},\vec{r}\,')$ and $F(\vec{r}\,',\vec{r}) = F(\vec{r},\vec{r}\,')$, for reasons we will explain later.

Lecture 7:

Advanced Electrostatics III:
Green Functions for Poisson's Equation (cont.)
Obtaining Green Functions from the Method of Images
Separation of Variables: General Considerations
Separation of Variables in Cartesian Coordinates

Date Revised: 2024/01/30 09:30
Revised lecture break
Date Given: 2024/01/25

Green Functions for Poisson's Equation with Dirichlet or Neumann Boundary Conditions

To apply the concept of Green Functions to Poisson's Equation, we start by taking $\phi(\vec{r}\,') = V(\vec{r}\,')$ and $\psi(\vec{r}\,') = -\epsilon_o G(\vec{r}, \vec{r}\,')$ in Green's Theorem (Equation 3.13) and assuming

$$-\epsilon_o \nabla^2_{\vec{r}\,'} G(\vec{r}, \vec{r}\,') = \delta(\vec{r} - \vec{r}\,') \tag{3.45}$$

Note that this equation does not match Equation 3.38, which had the Laplacian acting on $\vec{r}$, not $\vec{r}\,'$. We will recover Equation 3.38 later. We then apply the same kinds of manipulations we did to obtain the integral equation for the potential, Equation 3.36 (these manipulations rely on Equation 3.45), giving

$$V(\vec{r}) = \int_{\mathcal{V}} d\tau'\, \rho(\vec{r}\,')\, G(\vec{r}, \vec{r}\,') + \epsilon_o \oint_{\mathcal{S}(\mathcal{V})} da' \left[G(\vec{r}, \vec{r}\,')\, \hat{n}(\vec{r}\,') \cdot \vec{\nabla}_{\vec{r}\,'} V(\vec{r}\,') - V(\vec{r}\,')\, \hat{n}(\vec{r}\,') \cdot \vec{\nabla}_{\vec{r}\,'} G(\vec{r}, \vec{r}\,') \right] \tag{3.46}$$

As noted earlier, a differential equation is not alone sufficient; we need boundary conditions to make G unique. For a particular type of boundary condition on V, we can make the choice to impose a condition on G such that the integrand involving the other type of boundary condition on V vanishes. If we do so, then our integral equation for V reduces to an integration over the source distribution with the Green Function and over the boundary condition with the Green Function (Neumann) or its normal gradient (Dirichlet).

We also see that, even though we assumed Equation 3.45 instead of Equation 3.38 for the equation defining the Green Function, the result we obtain above is consistent with Equation 3.39, which states that the source function $\rho(\vec{r}')$ should be convolved with the Green Function, integrating over its second argument, to obtain the potential function in its first argument. We will resolve this apparent inconsistency shortly.

Note that the equation we obtain for $V(\vec{r})$ is different from the integral equation for $V(\vec{r})$, Equation 3.36, because there we could not impose such a condition on $V(\vec{r})$, since it is set by the situation under consideration, or on $|\vec{r}-\vec{r}'|^{-1}$ (obviously). $G(\vec{r},\vec{r}')$ is, on the other hand, our tool for solving that integral equation, so we may design the tool — by choosing boundary conditions for it — to do its job as long as it respects its defining equation. (Again, the differential equation is insufficient — we need to impose a boundary condition on G in order to make G's defining equation solvable.)

We can be more specific about what we mean by "forcing the other BC term to vanish" by picking a type of boundary condition:

- *Dirichlet boundary condition*

  In this case, $V(\vec{r})$ is specified for $\vec{r}\in\mathcal{S}$. Therefore, $\hat{n}(\vec{r})\cdot\vec{\nabla}_{\vec{r}}V(\vec{r})$ should be left unspecified — it should be determined by the solution itself — so we need for it to not appear in the integral equation. We can eliminate the term containing this normal derivative if we require the Dirichlet Green Function, $G_D(\vec{r},\vec{r}')$, to satisfy the boundary condition (in $\vec{r}'$, the variable for the defining PDE)

$$G_D(\vec{r},\vec{r}')=0 \qquad \text{for} \quad \vec{r}'\in\mathcal{S},\vec{r}\in\mathcal{V},\mathcal{S} \tag{3.47}$$

The boundary condition must be defined for all valid $\vec{r}$, which consists of $\vec{r} \in \mathcal{S}$ *and* $\vec{r} \in \mathcal{V}$. This has the benefit of making Equation 3.46 usable for calculating $V(\vec{r} \in \mathcal{S})$ to check the solution is consistent with the boundary condition.

Using the interpretation implied by the convolution of the charge density with the Green Function in Equation 3.46 (admittedly, an interpretation not obviously consistent with the defining equation, Equation 3.45), the above condition is equivalent to requiring that charge on the boundary ($\vec{r}' \in \mathcal{S}$), given by the normal gradient of V on the boundary, yield no contribution to the potential elsewhere on the boundary ($\vec{r} \in \mathcal{S}$) or in the volume ($\vec{r} \in \mathcal{V}$). In one sense, this is what we expect, as the Dirichlet boundary condition specifies $V(\vec{r})$ on the boundary, so any charge that appears on the boundary to enforce that boundary condition had better do so in a way that does not modify the boundary condition.

However, in another sense, it is the opposite of what we expect: how can the induced surface charge on the boundary not affect the potential on the surface or in the volume? Wasn't that the whole idea behind the method of images, that one calculates the additional potential of the induced surface charge on the boundary by replacing it with an image charge? We resolve this confusion below.

With the above condition, the solution for $V(\vec{r})$ reduces to

$$\boxed{V(\vec{r}) = \int_{\mathcal{V}} d\tau' \, \rho(\vec{r}') \, G_D(\vec{r}, \vec{r}') - \epsilon_o \oint_{\mathcal{S}(\mathcal{V})} da' \, V(\vec{r}') \, \hat{n}(\vec{r}') \cdot \vec{\nabla}_{\vec{r}'} \, G_D(\vec{r}, \vec{r}')} \qquad (3.48)$$

This form allows us to resolve our confusion above:

- The first term calculates the potential due to the real charge, including the potential due to the "image" charge induced by it on the boundary. (We'll start being sloppy about the use of the word "image" and drop the quotes.) The latter contribution must come from this term (*and not the surface term*) because the image charge and its potential ought to be linear in the real charge density: there is no image charge without real charge. The defining condition does not contradict this: $G_D(\vec{r},\vec{r}') \neq 0$ is allowed for $\vec{r},\vec{r}' \in \mathcal{V}$, $G_D(\vec{r},\vec{r}') = 0$ is only required for $\vec{r}' \in \mathcal{S}$ (and $\vec{r} \in \mathcal{V},\mathcal{S}$).

- The second term adds a contribution to the potential for surface charge that appears on the boundary in order for the boundary to sit at the nonzero potential given by the boundary condition. This is not image charge because it is not induced by real charge and it appears even if there is no real charge in $\mathcal{V}$ (this term's presence does not depend on whether ρ is present or not). In the case of the point charge near the sphere, this is the charge $q_0 = 4\,\pi\,\epsilon_o\,V_0\,R$ that appears so the sphere sits at $V = V_0$. It has nothing to do with the point charge q. The condition $G_D(\vec{r},\vec{r}') = 0$ for $\vec{r}' \in \mathcal{S}$ is the sensible condition that this additional surface charge does not induce its own image charge. It is sort of amazing that this simple term does all that work — figures out the surface charge required to realize the Dirichlet boundary condition and calculates its potential in $\mathcal{V}$.

For a Dirichlet boundary condition, the symmetry of G_D in its arguments can be proven by applying Green's Theorem with $\phi = G_D(\vec{r}, \vec{x})$ and $\psi = G_D(\vec{r}', \vec{x})$, where $\vec{x}$ is the variable that is integrated over, and using the defining equation, Equation 3.45, and the defining boundary condition $G_D(\vec{r}, \vec{x}) = 0$ for $\vec{x}$ on the boundary and $\vec{r}$ in the volume and on the boundary (which also implies the same for $G_D(\vec{r}', \vec{x})$). Symmetry of G_D implies symmetry of F_D given that their difference is symmetric in $\vec{r}$ and $\vec{r}'$.

When this symmetry property is applied to Equation 3.45, and we also use the symmetry of the delta function, Equation 3.38 is recovered (after relabeling $\vec{r} \leftrightarrow \vec{r}'$). This resolves the apparent inconsistency between wanting the Green Function to satisfy Equation 3.38 but having to assume Equation 3.45 at the start to get Equation 3.46.

We can use the symmetry requirement to reinterpret the condition $G_D(\vec{r}, \vec{r}') = 0$ for $\vec{r}' \in \mathcal{S}$. We can now think of the unit charge as being at $\vec{r} \in \mathcal{V}, \mathcal{S}$ and the potential as being calculated at $\vec{r}' \in \mathcal{S}$. This condition requires that G_D yields zero contribution to the potential on the *boundary* from charges *in the volume*. The first half of this statement is the requirement that image charge appear such that the sum of the potentials of the real charge in the volume and its image charge do not modify the boundary condition. (We do not talk about real charge on the boundary because we are considering only the Dirichlet problem right now.)

We can also now provide an interpretation of $F_D(\vec{r}, \vec{r}')$ in the Dirichlet case. Because 1) $F_D(\vec{r}, \vec{r}')$ satisfies Laplace's Equation in the volume $\mathcal{V}$, and 2) when added to the potential of a unit point charge at $\vec{r}'$ (the first term in our expression relating G_D and F_D, Equation 3.44), the sum satisifies the specified boundary condition on $\mathcal{S}$, $F_D(\vec{r}, \vec{r}')$ can be interpreted as the potential function in the volume due to the image charge induced on the boundary by the real charges in the volume with the boundary grounded. This image charge depends on where the charges in the volume are, hence the integration over $\vec{r}' \in \mathcal{V}$ to calculate *this* effect of this term.

What remains a bit mysterious or magical is how the second term in Equation 3.48 works. Clearly, that term calculates the surface charge density on the boundary needed for the Dirichlet boundary condition to be satisfied and then calculates the potential in the volume due to that surface charge density. It requires both terms in G_D (*i.e.*, $|\vec{r} - \vec{r}'|^{-1}$ and F_D) to do that. It seems this part just falls out of the mathematics.

- *Neumann boundary condition*

  In this case, $\widehat{n} \cdot \vec{\nabla} V(\vec{r})$ is specified for $\vec{r} \in \mathcal{S}$, so we need to render irrelevant the term containing $V(\vec{r})$ because we should not have to simultaneously specify it. While we might be inclined to require $\widehat{n}(\vec{r}\,') \cdot \vec{\nabla}_{\vec{r}\,'} G_N(\vec{r}, \vec{r}\,') = 0$ for $\vec{r}\,' \in \mathcal{S}$ to make this happen, this requirement is not consistent with Equation 3.45 defining G: if one integrates this equation for G_N over $\vec{r}\,' \in \mathcal{V}(\mathcal{S})$, and turns it into a surface integral using the divergence theorem, one obtains the requirement

$$-\epsilon_o \oint_{\mathcal{S}(\mathcal{V})} da'\, \widehat{n}(\vec{r}\,') \cdot \vec{\nabla}_{\vec{r}\,'} G_N(\vec{r}, \vec{r}\,') = 1 \qquad \text{for} \quad \vec{r} \in \mathcal{V}, \mathcal{S}$$

  Thus, the simplest condition we can impose on G_N is

$$\widehat{n}(\vec{r}\,') \cdot \vec{\nabla}_{\vec{r}\,'} G_N(\vec{r}, \vec{r}\,') = -\left[\epsilon_o \oint_{\mathcal{S}(\mathcal{V})} da'\right]^{-1} \qquad \text{for} \quad \vec{r} \in \mathcal{V}, \mathcal{S}, \vec{r}\,' \in \mathcal{S} \quad (3.49)$$

Applying this condition, the solution for $V(\vec{r})$ reduces to

$$\boxed{V(\vec{r}) = \int_{\mathcal{V}} d\tau'\, \rho(\vec{r}')\, G_N(\vec{r},\vec{r}') + \epsilon_o \oint_{\mathcal{S}(\mathcal{V})} da'\, G_N(\vec{r},\vec{r}')\, \hat{n}(\vec{r}') \cdot \vec{\nabla}_{\vec{r}'} V(\vec{r}') + \langle V(\vec{r}) \rangle_{\mathcal{S}(\mathcal{V})}}$$

$$\text{with} \quad \langle V(\vec{r}) \rangle_{\mathcal{S}(\mathcal{V})} \equiv \frac{\oint_{\mathcal{S}(\mathcal{V})} da'\, V(\vec{r}')}{\oint_{\mathcal{S}(\mathcal{V})} da'} \tag{3.50}$$

While $V(\vec{r})$ on the boundary has not been completely eliminated, its only appearance is via its average value on the boundary. This makes sense, as the Neumann boundary condition does not specify the potential offset since it only specifies derivatives of the potential. The appearance of this term reflects the freedom we have to set the potential offset for problems with Neumann boundary conditions. Recall that the Uniqueness Theorem only showed uniqueness up to an overall offset.

What is the interpretation of a Neumann Green Function? Since $\hat{n}(\vec{r}') \cdot \vec{\nabla}_{\vec{r}'} V(\vec{r}')$ specifies the surface charge density on the boundary, $G_N(\vec{r},\vec{r}')$ simply calculates the potential at a point $\vec{r}$ in the volume due to this boundary surface charge density at $\vec{r}'$. Note that G_N is convolved with the volume charge density and the surface charge density in the same way, reinforcing this interpretation. A Neumann Green Function thus has a simpler interpretation than a Dirichlet Green Function. There is no interpretation of G_N or F_N as calculating contributions from image charge.

What is the interpretation of $F_N(\vec{r}, \vec{r}')$ for the Neumann case? One can show that it has no effect (one needs to make use of symmetry of F_N in its arguments, see below). Not that it is identically zero, but that all terms involving it vanish. This makes sense: if we specify the surface charge density everywhere in the volume and on the surface, we should be able to just use Coulomb's Law to calculate the potential everywhere, which just requires the Coulumb's Law part of G_N.

The triviality of the Neumann Green Function may seem to render pointless the extended discussion leading to this point. Recall, however, that Dirichlet boundary conditions are far more common: we tend to specify potentials on the boundary in real situations, not the charge density. We derived the Neumann Green Function for completeness, not because it is really needed.

For a Neumann boundary condition, the symmetry of G_N and F_N is not a result of the boundary condition, but it may be assumed without loss of generality; see K.-J. Kim and J. D. Jackson, *Am. J. Phys.* **61**:1144 (1993). As with the Dirichlet Green Function, this symmetry property allows Equation 3.38 to be obtained from the assumed defining equation, Equation 3.45, closing the loop on that apparent inconsistency.

To make further progress in obtaining a functional form for the Green Function, we must specify the boundary conditions in more detail. We will consider examples of this next.

Obtaining Green Functions from the Method of Images

We mentioned earlier that the component $F_D(\vec{r}, \vec{r}')$ of the full Dirichlet Green Function $G_D(\vec{r}, \vec{r}')$ can be determined by the method of images in some cases. Let's see how this works for the two cases we have considered:

- *Point charge near grounded conducting plane*

  The full potential at a point $\vec{r}$ for the point charge at $d\,\hat{z}$ is

$$V(\vec{r}) = \frac{1}{4\,\pi\,\epsilon_o}\left[\frac{q}{|\vec{r} - d\,\hat{z}|} - \frac{q}{|\vec{r} + d\,\hat{z}|}\right] \tag{3.51}$$

  We can see by inspection that the Dirichlet Green Function is given by taking $q = 1$ and by replacing $d\,\hat{z}$ in the first term with $\vec{r}'$ and $-d\,\hat{z}$ in the second term with $\vec{r}'$ mirrored through the $x'y'$ plane:

$$G_D(\vec{r}, \vec{r}') = \frac{1}{4\,\pi\,\epsilon_o}\left[\frac{1}{|\vec{r} - \vec{r}'|} - \frac{1}{|\vec{r} - (x'\hat{x} + y'\hat{y} - z'\hat{z})|}\right] \tag{3.52}$$

  One can test this by plugging into Equation 3.47 with $\rho(\vec{r}') = q\,\delta(\vec{r}' - d\,\hat{z})$.

  The second term accounts for the fact that induced charge appears on the grounded conducting plane and calculates the contribution to the potential due to it; it is the $F(\vec{r}, \vec{r}')$ term while the first term is the usual Coulomb's Law term. The first term solves Poisson's Equation while the second term solves Laplace's Equation. Both terms depend on the position of the point charge at $\vec{r}'$.

This G_D is not manifestly symmetric under exchange of $\vec{r}$ and $\vec{r}'$, but one can rewrite it so it is:

$$G_D(\vec{r},\vec{r}') = \frac{1}{4\,\pi\,\epsilon_o}\left[\frac{1}{[(x-x')^2+(y-y')^2+(z-z')^2]^{1/2}} - \frac{1}{[(x-x')^2+(y-y')^2+(z+z')^2]^{1/2}}\right]$$

One can now also see how $G(z=0,\vec{r}')=0$ always: the two terms become identical in this case.

It is also important to notice that, for our boundary condition $V(z=0)=0$, there is no term in $V(\vec{r})$ for the surface term because it vanishes in this case. That is, in the Dirichlet case, we expect a surface term from Equation 3.48

$$-\epsilon_o \oint_{\mathcal{S}(\mathcal{V})} da'\, V(\vec{r}')\,\hat{n}(\vec{r}')\cdot\vec{\nabla}_{\vec{r}'}\,G_D(\vec{r},\vec{r}') \tag{3.53}$$

Since the Dirichlet boundary condition is $V(z=0)=0$, this integral vanishes and we indeed only have the volume integral term from Equation 3.48 convolving the original charge distribution with G_D.

- *Point charge near conducting plane held at V_0*

  Suppose our boundary condition had instead been $V(z=0)=V_0$, a constant (and also $V(r\to\infty)=V_0$ for consistency; we will elaborate on this later). Is the above Green Function still valid? *Yes!* We have not changed the charge distribution in $\mathcal{V}$ or the *type* of boundary condition; all we have done is change the *value* of the boundary condition. We can check that the new *value* of the Dirichlet boundary condition is respected when we apply G_D derived on the basis of the $V_0=0$ case.

  This is an important point about the Dirichlet Green Function: while one may find it using a special case, it is, by construction, valid for any Dirichlet boundary condition for the same geometry. It does not care about the details of *either* the charge distribution or the boundary condition. Of course, the special case used must be general enough that one can find the entire Green Function. When we later do an example using Separation of Variables in Cartesian coordinates to solve Laplace's Equation, we will see how that example determines a portion of the Dirichlet Green Function but not all of it.

  Returning to the matter at hand: because $V(\vec{r}\,')=V_0$ for $\vec{r}\,'\in\mathcal{S}(\mathcal{V})$, we can pull it outside the integral, so we just have the surface integral of the normal gradient of G_D over the surface:

$$-\epsilon_o\oint_{\mathcal{S}(\mathcal{V})}da'\,V(\vec{r}\,')\,\hat{n}(\vec{r}\,')\cdot\vec{\nabla}_{\vec{r}\,'}G_D(\vec{r},\vec{r}\,')=-\epsilon_o V_0\oint_{\mathcal{S}(\mathcal{V})}da'\hat{n}(\vec{r}\,')\cdot\vec{\nabla}_{\vec{r}\,'}G_D(\vec{r},\vec{r}\,')$$

We recall that, by definition, $G_D(\vec{r},\vec{r}')$ is the potential at the point $\vec{r}$ due to a point charge of unit magnitude ($q=1$) at $\vec{r}'$. By the symmetry of its arguments, it is also the potential at the point $\vec{r}'$ due to a unit point charge at $\vec{r}$. Earlier, when we did the method of images solution for the grounded conducting plane, we calculated the surface charge density at $\vec{r}$ due to the point charge at $d\,\hat{z}$ from $-\epsilon_o\,\vec{\nabla}_{\vec{r}}V(\vec{r},d\,\hat{z})$. In this case, $-\epsilon_o\,\vec{\nabla}_{\vec{r}'}G_D(\vec{r},\vec{r}')$ is the surface charge density at $\vec{r}'$ due to a unit charge at $\vec{r}$. Since V_0 has come outside the integral, our surface integral is now just the integral of this surface charge density over the boundary, or the total induced charge on the boundary. We calculated this when we did the method of images and found it was $Q_{ind}=-q$, so, in this case, it will be -1. That is:

$$-\epsilon_o\oint_{\mathcal{S}(\mathcal{V})}da'\,V(\vec{r}')\,\hat{n}(\vec{r}')\cdot\vec{\nabla}_{\vec{r}'}G_D(\vec{r},\vec{r}')=-V_0\oint_{\mathcal{S}(\mathcal{V})}da'\,\sigma_{ind}(\vec{r}',q=1)$$
$$=-V_0\,Q_{ind}(q=1)=V_0 \qquad (3.54)$$

So, we see that the surface term serves to add the potential offset that the boundary condition $V(z=0)=V_0$ requires. Therefore, the solution is now

$$V(\vec{r})=\frac{1}{4\,\pi\,\epsilon_o}\left[\frac{q}{|\vec{r}-d\,\hat{z}|}-\frac{q}{|\vec{r}+d\,\hat{z}|}\right]+V_0 \qquad (3.55)$$

This solution has $V(z=0)=V_0$ and $V(r\to\infty)=V_0$.

This example serves to highlight the fact that one has to be careful about the self-consistency of boundary conditions, especially when they involve a condition at infinity. Consider two alternative, invalid BCs:

- One cannot set $V(z=0)=V_0$ and $V(r\to\infty)=0$ because that is not self-consistent for $z=0$, $(x,y)\to\infty$: should the BC be V_0 or 0 for this part of the boundary?

- One cannot even require $V(z=0)=V_0$ and $V(z\to\infty)=0$ because it leaves unspecified the boundary condition for $V(z,\sqrt{x^2+y^2}\to\infty)$. If one then thinks about what type of BC to specify there, one finds that it should be impossible to specify something that is consistent with $V(z\to\infty)=0$. Think about the case of the conductor held at V_0 and no point charge. We know the solution is a uniform sheet of surface charge on the conductor, and we know that the field is then a constant $\vec{E}(\vec{r})=(\sigma/\epsilon_o)\,\hat{z}$ and the potential is $V(\vec{r})=-(\sigma/\epsilon_o)\,z$. This potential does not vanish as $z\to\infty$. If one knows that a set of boundary conditions is not self-consistent for the case of no point charge, then linearity/superposition tells us there is no way to fix the inconsistency by adding charges to $\mathcal{V}$: one would have to add a potential that is also not self-consistent to cancel out the self-inconsistency of the $q=0$ potential!

- *Point charge near grounded conducting sphere*

  The full potential at a point $\vec{r}$ for the point charge at $a\,\hat{z}$ was (Equation 3.25):

$$V(\vec{r}) = \frac{1}{4\,\pi\,\epsilon_o}\left[\frac{q}{|\vec{r} - a\,\hat{z}|} - \frac{q\,\frac{R}{a}}{|\vec{r} - \frac{R^2}{a}\,\hat{z}|}\right] \tag{3.56}$$

  Thus, the Dirichlet Green Function is given by letting $\vec{r}\,' = a\,\hat{z}$ and taking $q = 1$:

$$G_D(\vec{r}, \vec{r}\,') = \frac{1}{4\,\pi\,\epsilon_o}\left[\frac{1}{|\vec{r} - \vec{r}\,'|} - \frac{R/r\,'}{\left|\vec{r} - \vec{r}\,'\frac{R^2}{(r\,')^2}\right|}\right] \tag{3.57}$$

  Again, the second term accounts for the potential due to the charge induced on the surface of the sphere and is the term that solves Laplace's Equation in this situation (the $F_D(\vec{r}, \vec{r}\,')$ term). And again, one can this test form for G_D by plugging into Equation 3.47 with $\rho(\vec{r}\,') = q\,\delta(\vec{r}\,' - a\,\hat{z})$.

It is perhaps not so obvious that the second term in this Green Function is symmetric in its arguments. Let's rewrite it:

$$\frac{R/r'}{|\vec{r} - \vec{r}'\frac{R^2}{(r')^2}|} = \frac{R}{|\hat{r}\, r\, r' - R^2\hat{r}'|} \quad = \frac{R}{\sqrt{(r\, r')^2 + R^4 - 2\, r\, r' R^2 \hat{r}\cdot\hat{r}'}} \tag{3.58}$$

Now the symmetry is manifest.

The same point about the surface integral term as for the conducting plane holds here: that term vanishes because $V(\vec{r}') = 0$ for $\vec{r}' \in \mathcal{S}$.

- *Point charge near conducting sphere held at fixed potential*

  In this case, we can see the effect of the surface integral term in Equation 3.48 because $V(\vec{r})$ on the boundary does not vanish. The integral term is, from Equation 3.48:

$$-\epsilon_o \oint_{\mathcal{S}(\mathcal{V})} da'\, V(\vec{r}\,')\, \hat{n}(\vec{r}\,') \cdot \vec{\nabla}_{\vec{r}\,'} G_D(\vec{r}, \vec{r}\,') \tag{3.59}$$

  When we encountered this nonvanishing surface term for the prior case of a point charge near a conducting plane, we recognized that $V(\vec{r}\,') = V_0$ could be pulled outside the integral and that the integral of the normal gradient of the Green Function gives the total charge induced on the boundary for a unit charge at $\vec{r}$. To calculate that total induced charge, we invoke the theorem (based on Gauss's Law) we discussed earlier. In this case, the surface encloses the image charge, so the total induced charge is equal to the image charge. That is:

$$-\epsilon_o \oint_{\mathcal{S}(\mathcal{V})} da'\, V(\vec{r}\,')\, \hat{n}(\vec{r}\,') \cdot \vec{\nabla}_{\vec{r}\,'} G_D(\vec{r}, \vec{r}\,') = -V_0 Q_{ind} = -V_0 q_{image} = V_0 \frac{R}{r} \tag{3.60}$$

  This is again just the potential due to a point charge at the origin whose magnitude is such that the potential at radius R is V_0.

With this integral evaluated, the full solution for $V(\vec{r})$ is given by summing the term that involves the integral with ρ, which we calculated already for the grounded sphere case, with the boundary term:

$$V(\vec{r}) = \frac{q}{4\,\pi\,\epsilon_o}\left[\frac{1}{|\vec{r} - a\widehat{z}|} - \frac{R/a}{|\vec{r} - \frac{R^2}{a}\,\widehat{z}|}\right] + V_0\frac{R}{r}$$

This is what we found earlier when we discussed the same problem using the method of images.

- *Point charge in the presence of a charged, insulated, conducting sphere*

  The prior situation is identical to this one: specifying the charge on a conductor is the same as specifying its potential. So the result for $V(\vec{r})$ is the same, where we must take $V_0 = (Q + (R/a)q)/(4\,\pi\,\epsilon_o\,R)$. Note that, even though we are talking about a boundary condition in which charge is specified, it is not a Neumann boundary condition because we do not specify $\sigma(\vec{r}\,' \in \mathcal{S})$, we are still effectively specifying $V(\vec{r}\,' \in \mathcal{S})$. This case is like the third special case of the Uniqueness Theorem we discussed earlier.

Introduction to Separation of Variables

General Points on Separation of Variables

Griffiths makes this seem harder than it is. In separation of variables, we assume that the solution of Laplace's Equation factors into functions of single coordinates. This allows us to reduce the partial differential equation to a set of ordinary differential equations, which can be solved by standard techniques. Constants of integration appear that help to define the solutions. We apply the boundary conditions as defined by the voltages and/or the charge densities (normal derivative of voltage) at the boundaries. Once we find a set of solutions, we know from Sturm-Liouville theory that they form a *complete* set, so we are assured that we can write any solution to Laplace's Equation for the given boundary conditions in terms of these solutions.

We will only develop separation of variables for Laplace's Equation and, in the near term, we will only apply it to solving problems with specific types of boundary conditions rather than trying to use it to find the F piece of the Green Function. (Recall, F satisfies Laplace's Equation while G satisfies Poisson's Equation.) We will see later, at the tail end of our discussion of separation of variables in spherical coordinates, that this technique will actually be sufficient to obtain the Green Function for an arbitrary geometry, which then provides us the solution to Poisson's Equation. (One will be able to see that it is not feasible to do separation of variables for Poisson's Equation in the same way we do it for Laplace's Equation: the process very much relies on the vanishing of one side of the equation!)

Digression on Orthonormal Functions

The general topic of the properties of solutions to second-order linear differential equations is beyond the scope of this course; it falls under the name *Sturm-Liouville theory*, and it is covered in ACM95/100. We will simply quote some results that are important for this course.

Sturm-Liouville theory consists of recognizing that the second-order linear ordinary differential equations we encounter in many places in this course are self-adjoint (Hermitian) operators on the Hilbert space of functions that satisfy the differential equation. You know from linear algebra that Hermitian operators are guaranteed to have a set of eigenvalues and eigenvectors (in this case, eigenfunctions), and that the eigenvectors form an orthonormal basis for the space under consideration (here, again, the space of functions that satisfy the differential equation). The same results apply here. What this means is that, for such equations, there are a set of solution functions $\{f_p(w)\}$ that are the eigenfunctions of the operator, and there are corresponding eigenvalues $\{\lambda_p\}$. These eigenfunctions form a *complete, orthonormal set*. (Note: w is intended to represent any coordinate, one- or multi-dimensional.) The original differential equation (with differential operator O_w and the eigenvalue-eigenvector equation are:

$$O_w\, f(w) = k\, f(w) \qquad\qquad O_w\, f_p(w) = \lambda_p(w) \tag{3.61}$$

where k is initially an undetermined constant; solving the the differential equation determines the allowed values of k, the $\{\lambda_p\}$, and the corresponding solutions $\{f_p(w)\}$.

Orthonormality is written mathematically as

$$\int_s^t dw\, f_p^*(w)\, f_q(w) = \delta_{pq} \tag{3.62}$$

where integration over the interval of interest $[s, t]$ is the Hilbert space inner product.

Completeness is defined to be

$$\sum_p f_p^*(w')\, f_p(w) = \delta(w' - w) \tag{3.63}$$

where the sum is over all eigenfunctions of the differential equation.

Completeness, as its name indicates, enables us to show that any function $g(w)$ on $[s,t]$ can be expanded in terms of the eigenfunctions $\{f_p\}$:

$$g(w) = \int_s^t dw'\, g(w')\, \delta(w'-w) = \int_s^t dw'\, g(w') \sum_p f_p^*(w') f_p(w)$$

$$= \sum_p f_p(w) \int_s^t dw'\, f_p^*(w') g(w')$$

That is, we have the expansion:

$$\boxed{g(w) = \sum_p A_p f_p(w)} \tag{3.64}$$

with coefficients given by

$$\boxed{A_p = \int_s^t dw'\, f_p^*(w')\, g(w')} \tag{3.65}$$

We could have derived Equation 3.65 also by applying orthornomality to the expansion Equation 3.64; this is the usual way we think of finding the $\{A_p\}$ as we will see below. They are of course equivalent derivations.

Separation of Variables in Cartesian Coordinates

We assume that the function $V(\vec{r})$ can be factorized as

$$V(\vec{r}) = X(x)\,Y(y)\,Z(z) \tag{3.66}$$

Plugging this into Laplace's Equation, we obtain

$$\begin{aligned} Y(y)\,Z(z)\,\frac{d^2X}{dx^2} + X(x)\,Z(z)\,\frac{d^2Y}{dY^2} + X(x)\,Y(y)\,\frac{d^2Z}{dz^2} &= 0 \\ \frac{1}{X(x)}\,\frac{d^2X}{dx^2} + \frac{1}{Y(y)}\,\frac{d^2Y}{dY^2} + \frac{1}{Z(z)}\,\frac{d^2Z}{dz^2} &= 0 \end{aligned} \tag{3.67}$$

We have three terms, the first a function of x, the second of y, and the third of z. Given these mismatched dependences, the only way the equation can hold is if each term is a constant. That is, it must hold that

$$\frac{1}{X(x)}\,\frac{d^2X}{dx^2} = K_1 \qquad \frac{1}{Y(y)}\,\frac{d^2Y}{dY^2} = K_2 \qquad \frac{1}{Z(z)}\,\frac{d^2Z}{dz^2} = K_3 \tag{3.68}$$

with $K_1 + K_2 + K_3 = 0$.

We know that the solution to these ordinary differential equations are exponentials,

$$X(x) = A\,\exp(x\sqrt{K_1}) + B\,\exp(-x\sqrt{K_1}) \tag{3.69}$$

$$Y(y) = C\,\exp(y\sqrt{K_2}) + C\,\exp(-y\sqrt{K_2}) \tag{3.70}$$

$$Z(z) = E\,\exp(z\sqrt{-(K_1+K_2)}) + F\,\exp(-z\sqrt{-(K_1+K_2)}) \tag{3.71}$$

We have not specified which of K_1, K_2, and K_3 are positive and which are negative (clearly, they cannot all be the same sign). That will be determined by the boundary conditions. Note that we are also neglecting linear solutions that also satisfy the individual ordinary differential equations; we will see they are not necessary in the examples we consider here (though they may be needed more generally).

At this point, we cannot make further generic progress; we need to apply a set of boundary conditions. These will place constraints on the allowed values of the exponents and coefficients and restrict the family of solutions. There are a number of examples in Griffiths. To avoid duplication, we use a different one here from Jackson §2.9.

Example 3.4: Empty box with five walls grounded and one held at a potential

Consider a box with side lengths a, b, and c in the x, y, and z dimensions and with one corner at the origin. The boundary conditions are

$$V(x=0)=0 \qquad V(y=0)=0 \qquad V(z=0)=0 \tag{3.72}$$

$$V(x=a)=0 \qquad V(y=b)=0 \qquad V(z=c)=\phi(x,y) \tag{3.73}$$

where $\phi(x,y)$ is a function that is given. *In SoV, we always apply the homogeneous (vanishing RHS) BCs first because, we will see, they restrict the functional form of the solutions.* The homogeneous BC in the ith dimension (*e.g.*, y) can only be satisfied if the ith function (*e.g.*, $Y(y)$) satisfies it alone because it must be satisfied for all values of the other coordinates. Let's do x, y first for convenience (with foreknowledge of solution):

$$X(0)=A+B=0 \qquad X(a)=A\exp(a\sqrt{K_1})+B\exp(-a\sqrt{K_1})=0 \tag{3.74}$$

$$Y(0)=C+D=0 \qquad Y(b)=C\exp(b\sqrt{K_2})+D\exp(-b\sqrt{K_2})=0 \tag{3.75}$$

Reducing,

$$A\left[\exp(a\sqrt{K_1})-\exp(-a\sqrt{K_1})\right]=0 \tag{3.76}$$

$$C\left[\exp(b\sqrt{K_2})-\exp(-b\sqrt{K_2})\right]=0 \tag{3.77}$$

There is no solution to these equations for $K_1 > 0$ and $K_2 > 0$: the unit-normalized decaying and rising exponentials are only equal when their arguments both vanish, and they do not. Therefore, let's take $K_1 = -\alpha^2$ and $K_2 = -\beta^2$ so these become oscillating exponentials. We thus obtain the conditions

$$\sin(\alpha\, a) = 0 \qquad \sin(\beta\, b) = 0 \tag{3.78}$$

This places conditions on the allowed values of α and β:

$$\alpha_n = \frac{n\,\pi}{a} \qquad \beta_m = \frac{m\,\pi}{b} \qquad n,\, m \text{ positive integers} \tag{3.79}$$

where n and m may only be positive integers because negative values are redundant with the positive ones and $n = 0$ and $m = 0$ yield vanishing functions. Thus, we have

$$X(x) = \sum_{n=1}^{\infty} A_n \sin \alpha_n x \qquad Y(y) = \sum_{m=1}^{\infty} C_m \sin \beta_m y \tag{3.80}$$

where the $\{A_n\}$ and $\{C_m\}$ are constants to be determined. These solutions clearly respect the $V = 0$ boundary conditions at $x = 0, a$ and $y = 0, b$ because they vanish at those points. There is no relationship between n and m or between α_n and β_m at this point.

Now, let's apply the remaining homogeneous BC to $Z(z)$. At $z = 0$, we have

$$Z(0) = E + F = 0 \qquad \Longrightarrow \qquad F = -E \tag{3.81}$$

Therefore, $Z(z)$ is of the form

$$Z(z) = E_{nm}\left[\exp(z\sqrt{\alpha_n^2 + \beta_m^2}) - \exp(-z\sqrt{\alpha_n^2 + \beta_m^2})\right] \tag{3.82}$$

$$= E'_{nm}\sinh(\gamma_{nm}z) \qquad \text{with} \qquad \gamma_{nm} = \sqrt{\alpha_n^2 + \beta_m^2} \tag{3.83}$$

(sinh not sin because we know $\alpha_n^2 + \beta_m^2 > 0$.) Note how the last BC only determined the form of $Z(z)$ while its eigenvalues were determined by the prior two BC.

Our full solution thus has the form

$$V(x, y, z) = \sum_{n,m=1}^{\infty} V_{nm}(x, y, z) \tag{3.84}$$

$$= \sum_{n,m=1}^{\infty} A_{nm}\sin(\alpha_n x)\sin(\beta_m y)\sinh(\gamma_{nm}z) \quad \text{with} \quad \gamma_{nm} = \sqrt{\alpha_n^2 + \beta_m^2}$$

where we have combined all the arbitrary coefficients A_m, C_n, and E'_{nm} into a single coefficient A_{nm}. Each $V_{nm}(\vec{r})$ satisfies all five homogeneous BCs, thus the solution so far (without the last BC applied) is the arbitrary sum over all such $V_{nm}(\vec{r})$.

Lecture 8:

Advanced Electrostatics IV:
Separation of Variables in Cartesian Coordinates (cont.)
Separation of Variables in Spherical Coordinates: General Theory
Separation of Variables in Spherical Coordinates
with Azimuthal Symmetry

Date Revised: 2024/02/01 07:15
Explicitly discuss which BC are
homogeneous and inhomogeneous in Example 3.6
Date Given: 2024/01/30

Now, we want to apply the last boundary condition, $V(x, y, z = c) = \phi(x, y)$. How? Not the same way as we applied the previous ones. The prior boundary conditions were *homogeneous*, meaning that they forced the solution to vanish on some boundary (not all boundaries, otherwise the solution would vanish by the "no extrema" property). The remaining one is *inhomogeneous* because it requires the solution to take on a particular functional form on a boundary. It must be treated differently, for two reasons.

- The first involves linearity and uniqueness. Because the right-hand side of a homogeneous BC is zero, the BC is satisfied by any linear combination of functions that satisfy the BC. The same is not true of inhomogeneous BC. If it were possible for two *different* functions to satisfy the inhomogeneous BC, then only a subset of linear combinations of them would satisfy the same BC: the linear combinations in which the coefficients sum to unity. This condition violates linearity. The only resolution is for there to be precisely one solution to the inhomogeneous BC. This requirement is consistent with uniqueness: the inhomogeneous BC is applied last, and it completes the application of the BC, so the solution *should be* unique once it is applied.
- From the purely calculational point of view, requiring the solution for a given n, m to satisfy the inhomogeneous boundary condition would imply

$$V_{nm}(x, y, z = c) = \phi(x, y) \tag{3.85}$$

$$A_{nm} \sin(\alpha_n x) \sin(\beta_m y) \sinh(\gamma_{nm} c) = \phi(x, y) \tag{3.86}$$

There simply is not enough freedom in the functional form on the left to satisfy the boundary condition for arbitrary $\phi(x, y)$.

The only way to have enough freedom to satisfy the inhomogeneous boundary condition is to consider a linear combination of the individual n, m:

$$V(\vec{r}) = \sum_{n,m=1}^{\infty} A_{nm} \sin(\alpha_n x) \sin(\beta_m y) \sinh(\gamma_{nm} z) \tag{3.87}$$

where A_{nm} are now constants to find based on requiring the above linear combination solution satisfies the inhomogeneous boundary condition at $z = c$, which now becomes

$$\phi(x, y) = V(x, y, z = c) = \sum_{n,m=1}^{\infty} A_{nm} \sin(\alpha_n x) \sin(\beta_m y) \sinh(\gamma_{nm} c) \tag{3.88}$$

This condition will let us determine the A_{nm}, but how, and why are we certain they exist? We make use of the theory of orthonormal functions we cited earlier.

We will use the fact (not proven here) that the functions $\{\sqrt{2/a}\sin(\alpha_n x)\}$ for $n \geq 1$ form a complete, orthonormal set on the $x \in [0, a]$ interval (*with the given boundary conditions at* $x = 0$, a), as do $\{\sqrt{2/b}\sin(\beta_n y)\}$ for $m \geq 1$ on $y \in [0, b]$ (again, with BC). Therefore, we may recover the A_{nm} by multiplying by them and integrating:

$$\int_0^a dx \int_0^b dy\, \phi(x,y)\sqrt{\frac{2}{a}}\sin(\alpha_p x)\sqrt{\frac{2}{b}}\sin(\beta_q y)$$

$$= \int_0^a dx \int_0^b dy \sum_{n,m=1}^{\infty} A_{nm}\sinh(\gamma_{nm}c)\,\sin(\alpha_p x)\sqrt{\frac{2}{a}}\sin(\alpha_n x)\sin(\beta_m y)\sqrt{\frac{2}{b}}\sin(\beta_q y)$$

$$= \sum_{n,m=1}^{\infty} A_{nm}\sinh(\gamma_{nm}c)\sqrt{\frac{a}{2}}\delta_{pn}\sqrt{\frac{b}{2}}\delta_{qm} = \frac{\sqrt{a\,b}}{2}A_{pq}\,\sinh(\gamma_{pq}c) \qquad (3.89)$$

Now, be aware that we did more work than necessary above. Once we are told that the $\{\sqrt{2/a}\,\sin(\alpha_n x)\,\sqrt{2/b}\,\sin(\beta_m y)\}$ form an orthonormal set, we do not need to do the integrals on the right-hand side! We only need write the right-hand side of the original equation in terms of the orthonormal functions, then use orthonormality (Equation 3.65) to obtain the equations for the individual coefficients; *i.e.*:

$$\phi(x,y) = \sqrt{\frac{a\,b}{4}}\sum_{n,m=1}^{\infty} A_{nm}\sqrt{\frac{2}{a}}\sin(\alpha_n x)\sqrt{\frac{2}{b}}\sin(\beta_m y)\sinh(\gamma_{nm}c) \qquad (3.90)$$

$$\Longrightarrow \quad \int_0^a dx \int_0^b dy\sqrt{\frac{2}{a}}\sin(\alpha_p x)\sqrt{\frac{2}{b}}\sin(\beta_q y)\,\phi(x,y) = \sqrt{\frac{a\,b}{4}}A_{pq}\sinh(\gamma_{pq}c) \qquad (3.91)$$

Next, we move the coefficients to one side to obtain (replacing pq with mn):

$$A_{nm} = \frac{1}{\sinh(\gamma_{nm} c)} \int_0^a dx \int_0^b dy \, \frac{2}{a} \sin(\alpha_n x) \, \frac{2}{b} \sin(\beta_m y) \, \phi(x, y) \tag{3.92}$$

Our full solution for the applied set of boundary conditions is

$$V(\vec{r}) = \frac{4}{a\,b} \sum_{n,m=1}^{\infty} \sin(\alpha_n x) \sin(\beta_m y) \, \frac{\sinh(\gamma_{nm} z)}{\sinh(\gamma_{nm} c)} \int_0^a dx' \int_0^b dy' \phi(x', y') \sin(\alpha_n x') \sin(\beta_m y') \tag{3.93}$$

Summary: The homogeneous boundary conditions restricted the solutions to a specific orthonormal set, and the single inhomogeneous boundary condition sets the coefficients of the appropriate linear combination of that orthonormal set.

A good exercise is to write down the solutions for the five other inhomogeneous boundary condition cases (especially the ones with the inhomogeneous condition on the x, y, or $z = 0$ planes) "by inspection" — *i.e.*, by simply changing the solution we already have by replacing z with x, y, or $a - x$, $b - y$, or $c - z$ — rather than by rederiving. Clearly, these other problems are not different in any conceptual way, they are only different calculationally, and only barely. There is no reason to redo all that calculation from scratch!

If we had used a more general boundary condition, specifying V to be nonzero on all six sides of the box, then we could solve the similar problem for each of the six faces independently (*i.e.*, let V be nonzero and arbitrary on that face and zero on all the other faces) and then sum the solutions since each individual solution does not affect the capability of the other solutions to satisfy their boundary conditions. (Of course, the boundary conditions themselves must be consistent with each other at the edges and corners where they meet.) *In fact, we would have to do this; the separation of variables technique provides no way to satisfy two generic, independent inhomogeneous boundary conditions simultaneously.* Rather, to solve problems involving multiple inhomogeneous boundary conditions, one must use the property that an inhomogeneous boundary condition solution can always be summed with an arbitrary number of homogeneous boundary condition solutions and still satisfy the inhomogeneous boundary condition.

It is interesting to consider the intermediate case, consisting of the same geometry with constant potentials ϕ_0 at the $z = c$ face and $-\phi_0$ at the $z = 0$ face. As stated above, one can solve the two cases of ϕ_0 and $-\phi_0$ separately and add them. One can also solve the problem directly by simultaneously applying the two boundary conditions, and one can show that the two solutions are the same (using some hyperbolic trigonometry identities). This is possible because the double-inhomogeneous boundary condition in this case is very simple, having only one free parameter, ϕ_0. A generic double-inhomogeneous boundary condition problem cannot be solved in this way.

Referring back to our discussion of Green Functions, the above solution is the surface term in Equation 3.48 for the particular boundary condition we have applied. By comparison of the two expressions, we infer (not derive!)

$$
\begin{aligned}
-\epsilon_o \hat{n}(\vec{r}\,') \cdot \vec{\nabla}_{\vec{r}\,'} G_D(\vec{r}, \vec{r}\,' = x'\hat{x} + y'\hat{y} + c\hat{z}) \qquad (3.94)\\
= \frac{4}{a\,b} \sum_{n,m=1}^{\infty} \sin(\alpha_n x) \sin(\beta_m y) \frac{\sinh(\gamma_{nm} z)}{\sinh(\gamma_{nm} c)} \sin(\alpha_n x') \sin(\beta_m y')
\end{aligned}
$$

Note that this expression does not fully specify G_D (or F_D)! The above information is sufficient for the particular physical situation we have set up, which consists of no physical charge in the volume and the above boundary condition, because:

- The term consisting of the integral of the charge density in the volume convolved with G_D is zero in this case because the charge density vanishes *in the volume*. Therefore, we do not need to know G_D (or F_D) completely.
- The above surface term is the only one needed because $V = 0$ on the other boundaries.

For the more general problem of an arbitrary charge distribution in the volume and arbitrary Dirichlet boundary conditions on the surfaces, we would need to find the full G_D. It may seem like one could do as suggested earlier, finding the solution for each option for which face is held at nonzero potential, then using the results analogous to the above as six Neumann boundary conditions on G_D, and applying separation of variables to find G_D. But one would have to require that G_D solve *Poisson's Equation* for a unit point charge, not Laplace's Equation. This, as we noted earlier, is not feasible with separation of variables because of the nonzero right side of the equation. There is a way to deal with this, which we will show a bit later when we develop the spherical harmonic expansion for the Green Function in spherical coordinates.

Another approach that does work would be the method of images with the condition $V = 0$ on all the surfaces. It is left as an exercise for the reader to think about what set of image charges is appropriate; the situation gets complicated for a charge at an arbitrary position in the box, but it is solvable. Certainly, from the resulting G_D, we could compute the normal gradient of G_D on any surface and thus obtain the general solution for V in the volume for any Dirichlet boundary condition. We should find that the normal gradient of G_D on the $z = c$ surface is what is given above.

It may seem like separation of variables is unsatisfactory for this reason — the procedure does not give you the full Green Function, while the method of images does. But, as we have seen, the method of images is not a systematic procedure — one has to guess the correct image charge distribution. By contrast, separation of variables is an entirely algorithmic procedure to give you a solution if a separable one exists for the particular boundary condition you are applying. It is less general but more reliable. *More importantly, we will show later how, by applying separation of variables in a more sophisticated way, we can in fact find the full Green Function.*

There is, nevertheless, no guarantee that there will be a separable solution; this depends on the geometry of the boundary conditions. The boundary conditions need to respect the separability assumed. For example, a boundary condition on a spherical boundary would not likely yield a solution via separation of variables in Cartesian coordinates!

Note also that the method of images technique is not appropriate for a Neumann boundary condition because the method of images solution generally solves the $V = 0$ Dirichlet BC problem. One needs a technique like separation of variables for such cases.

Separation of Variables in Spherical Coordinates: General Theory

Doing the Separation in Spherical Coordinates

We do this in a slightly more general manner than Griffiths, dropping the assumption of azimuthal symmetry until it is time to solve the separated differential equations.

Laplace's Equation in spherical coordinates is:

$$\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial V}{\partial r}\right)+\frac{1}{r^2\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial V}{\partial\theta}\right)+\frac{1}{r^2\sin^2\theta}\frac{\partial^2 V}{\partial\phi^2}=0 \tag{3.95}$$

If we assume a separable form

$$V(r,\theta,\phi)=R(r)\,\Theta(\theta)\,\Phi(\phi) \tag{3.96}$$

then, after dividing through by $V(r,\theta,\phi)$ and multiplying by $r^2\sin^2\theta$, we have

$$\sin^2\theta\left[\frac{1}{R(r)}\frac{d}{dr}\left(r^2\frac{dR}{dr}\right)+\frac{1}{\Theta(\theta)}\frac{1}{\sin\theta}\frac{d}{d\theta}\left(\sin\theta\frac{d\Theta}{d\theta}\right)\right]+\frac{1}{\Phi(\phi)}\frac{d^2\Phi}{d\phi^2}=0 \tag{3.97}$$

We see that the first term depends only on r and θ while the second term depends only on ϕ, so we can immediately assume they are each equal to a constant:

$$\frac{1}{\Phi(\phi)}\frac{d^2\Phi}{d\phi^2} = -m^2 \tag{3.98}$$

The choice of the form of the constant is motivated by what will come next, but we can see why it needs to be of this form. As we saw in Cartesian coordinates, the above differential equation is solved either by growing/decaying exponentials (right side positive) or oscillating exponentials (right side negative). Since ϕ is a coordinate that repeats on itself ($\phi = 2\,n\,\pi$ are the same physical coordinate) the solutions $\Phi(\phi)$ must also be periodic, forcing the choice of the oscillating exponential. (For the same reason, the linear solutions we ignored in the Cartesian case are disallowed here.) We saw before that it is convenient to define the constant to incorporate a squaring.

The solutions of this equation are straightforward:

$$\Phi(\phi) = A\,\exp(i\,m\,\phi) + B\,\exp(-i\,m\,\phi) \tag{3.99}$$

Periodicity in ϕ with period $2\,\pi$ requires m be an integer. One can either require $m \geq 0$ and keep the $\{A_m\}$ and $\{B_m\}$ or allow m to be any integer and drop the $\{B_m\}$ (which would be redundant with the $\{A_m\}$ for $m < 0$). In either case, only one of A_0 or B_0 is required.

Returning to the other term, we now have

$$\sin^2\theta\left[\frac{1}{R(r)}\frac{d}{dr}\left(r^2\frac{dR}{dr}\right)+\frac{1}{\Theta(\theta)}\frac{1}{\sin\theta}\frac{d}{d\theta}\left(\sin\theta\frac{d\Theta}{d\theta}\right)\right]=m^2 \tag{3.100}$$

$$\frac{1}{R(r)}\frac{d}{dr}\left(r^2\frac{dR}{dr}\right)+\left[\frac{1}{\Theta(\theta)}\frac{1}{\sin\theta}\frac{d}{d\theta}\left(\sin\theta\frac{d\Theta}{d\theta}\right)-\frac{m^2}{\sin^2\theta}\right]=0 \tag{3.101}$$

Now, we see that the first term depends only on r and the second only on θ, so we can separate again by setting the two terms equal to constants that sum to zero. Here, we rely on prior knowledge of the result to choose the constant to be $\ell(\ell+1)$ so that

$$\frac{1}{R(r)}\frac{d}{dr}\left(r^2\frac{dR}{dr}\right)=\ell(\ell+1) \tag{3.102}$$

$$\frac{1}{\Theta(\theta)}\frac{1}{\sin\theta}\frac{d}{d\theta}\left(\sin\theta\frac{d\Theta}{d\theta}\right)-\frac{m^2}{\sin^2\theta}=-\ell(\ell+1) \tag{3.103}$$

Note that the radial equation does not depend on m. This implies that the $R(r)$ functions will not depend on the azimuthal properties of the problem, in particular whether it has azimuthal symmetry. But $R(r)$ depends on ℓ, so it will depend on the polar properties of the problem. $\Theta(\theta)$ depends on ℓ and m, so its behavior depends on both the polar and azimuthal properties of the problem. $\Phi(\phi)$ looks like it may only depend on the azimuthal properties because it depends only on m, but m is tied to ℓ through the polar equation, so there will be some relationship.

Solving the Radial Equation

Here, we add another item to our "bag of tricks" and define $U(r)$ by $R(r) = U(r)/r$ and plug in. (This is motivated by the r^2 that the second d/dr must act on: assuming this dependence gets rid of the extra terms arising because of that factor.) We find

$$\frac{d^2U}{dr^2} - \frac{\ell(\ell+1)}{r^2}\,U(r) = 0 \tag{3.104}$$

Since the two derivatives would reduce the exponent of a power-law solution by 2, and the second term does the same by dividing by r^2, the above equation suggests $U(r)$ is a power law in r. (Or, try making it work with a transcendental function: you can't.) If we plug in such a form $U(r) = r^a$, we find

$$a(a-1)r^{a-2} - \ell(\ell+1)r^{a-2} = 0 \qquad \Longrightarrow \qquad a_1 = \ell+1 \quad \text{or} \quad a_2 = -\ell \tag{3.105}$$

$$\Longrightarrow \qquad \boxed{R(r) = \frac{U(r)}{r} = A\,r^{a_1-1} + B\,r^{a_2-1} = A\,r^{\ell} + \frac{B}{r^{\ell+1}}} \tag{3.106}$$

There is no constraint on ℓ yet.

The Polar Equation and the Generalized Legendre Equation

We may rewrite the polar angle equation as

$$\frac{1}{\sin\theta}\frac{d}{d\theta}\left(\sin\theta\frac{d\Theta}{d\theta}\right)+\left[\ell(\ell+1)-\frac{m^2}{\sin^2\theta}\right]\Theta(\theta)=0 \tag{3.107}$$

Motivated by the fact that $\sin\theta\, d\theta = -d(\cos\theta)$, we add another trick to our bag of tricks by writing

$$x=\cos\theta \qquad \Theta(\theta)=P(\cos\theta)=P(x) \qquad 1-x^2=\sin^2\theta \tag{3.108}$$

Then we may rewrite the polar differential equation as

$$\boxed{\frac{d}{dx}\left[(1-x^2)\frac{dP}{dx}\right]+\left[\ell(\ell+1)-\frac{m^2}{1-x^2}\right]P(x)=0} \tag{3.109}$$

This is called the *generalized Legendre equation*.

As you have seen in ACM95/100, differential equations of this type can be solved by assuming the solution is a polynomial in x and requiring termination after a finite number of terms. That is, one assumes

$$P_\ell^m(x) = \sum_{k=1}^{\infty} a_k x^k \tag{3.110}$$

and then, plugging the above form into the differential equation, one requires the series to terminate ($a_k = 0$ for some k). This condition forces ℓ to be a nonnegative integer and $-\ell \leq m \leq \ell$. (We already know m is an integer to ensure $\Phi(\phi)$ is single-valued.) These polynomials are the *associated Legendre polynomials.*

Mathematically, there should be a second solution for each ℓ, m because the equation is second order. These are the solutions one finds by not requiring termination but simply convergence for $-1 < x < 1$ (corresponding to $0 < \theta < \pi$). If one has a geometry that excludes the z-axis (where these solutions diverge), these solutions must be considered. If the z-axis is in the space, then these solutions are unphysical and can be discarded.

Separation of Variables in Spherical Coordinates with Azimuthal Symmetry

The Polar Equation Solution with Azimuthal Symmetry: the Legendre Equation and Legendre Polynomials

Consider the special case of azimuthal symmetry, for which $m = 0$ and $\Phi(\phi) =$ constant. The generalized Legendre Equation reduces to the *Legendre Equation*:

$$\boxed{\frac{d}{dx}\left[(1-x^2)\frac{dP}{dx}\right] + \ell(\ell+1)\,P(x) = 0} \tag{3.111}$$

The same series solution applies here with $m = 0$, so ℓ must still be a nonnegative integer. These solutions are the *Legendre Polynomials*. One can show they obey *Rodrigues' Formula*:

$$\boxed{P_\ell(x) = \frac{1}{2^\ell \ell!}\left(\frac{d}{dx}\right)^\ell (x^2-1)^\ell} \tag{3.112}$$

Properties of the Legendre Polynomials

One can see by inspection or prove the following properties:

- $P_\ell(x)$ is a ℓth-order polynomial in x.
- $P_\ell(x)$ has only even powers of x if ℓ is even and only odd powers if ℓ is odd. $\Longrightarrow P_\ell(x)$ is an even function of x for ℓ even and an odd function for ℓ odd.
- The Legendre polynomials are a complete, orthonormal set: any function that remains finite on the interval $[-1, 1]$ can be written in terms of them. Their orthonormality relation is

$$\boxed{\int_{-1}^{1} dx \sqrt{\frac{2\ell+1}{2}} P_\ell(x) \sqrt{\frac{2\ell'+1}{2}} P_{\ell'}(x) = \delta_{\ell\,\ell'}} \tag{3.113}$$

and their completeness relation is

$$\boxed{\sum_{\ell=0}^{\infty} \frac{2\ell+1}{2} P_\ell(x) P_\ell(x') = \delta(x-x')} \tag{3.114}$$

- $P_\ell(1) = 1$ and $P_\ell(-1) = (-1)^\ell$.
- $P_\ell(0) = [(-1)^n\,(2n-1)!!]/2^n\,n!$ for even $\ell = 2\,n$. $P_\ell(0) = 0$ for odd ℓ.

Full Solution to Laplace's Equation with Azimuthal Symmetry

Combining our radial and polar equation solutions, we have that, for any problem with azimuthal symmetry and in which the z-axis is included, the potential must have the form

$$\boxed{V(r,\theta) = \sum_{\ell=0}^{\infty} \left(A_\ell \, r^\ell + \frac{B_\ell}{r^{\ell+1}} \right) P_\ell(\cos\theta)} \tag{3.115}$$

The coefficients $\{A_\ell\}$ and $\{B_\ell\}$ are set by the boundary conditions. If the volume includes the origin and the boundary conditions imply the potential must be finite there, the $\{B_\ell\}$ may be eliminated, and, if the volume includes infinity and the boundary conditions require the potential be finite (usually zero) there, the $\{A_\ell\}$ may be eliminated. In other cases, some or all of the $\{A_\ell\}$ and $\{B_\ell\}$ can be nonzero. Usually, application of the boundary conditions on V will require use of the orthonormality relations for the Legendre polynomials.

We note that, in the process of doing separation of variables, we have proven that the angular solution satisfies the *eigenvalue-eigenfunction equation*

$$\boxed{\nabla^2 P_\ell(\cos\theta) = -\frac{\ell(\ell+1)}{r^2} P_\ell(\cos\theta)} \tag{3.116}$$

For the angular equation, r acts as a constant and so appears in the eigenvalue.

Examples of Separation of Variables with Azimuthal Symmetry

We will start first with a case in which the boundary condition is quite obviously Dirichlet and the application is very much like what we did in Cartesian coordinates. Generally speaking, however, boundary conditions are not always so obvious. One has to use whatever information one is given and turn it into boundary conditions of the type that we know provides uniqueness.

Example 3.5: Dirichlet Boundary Condition on a Spherical Boundary with Azimuthal Symmetry

Suppose $V(R,\theta)$, the potential as a function of θ on a sphere of radius R, is specified, where the sphere is either the outer boundary or the inner boundary of the space. What is the explicit form for the resulting potential?

Let's consider the two cases together. If the space is $r < R$, then we require the $\{B_\ell\}$ to vanish to ensure a finite potential at the origin. (There is no charge in the volume, so we are assured that the potential cannot be infinite there.) If the space is $r > R$, then we require the $\{A_\ell\}$ to vanish so the potential goes to zero at infinity. That is:

$$V(r,\theta) = \sum_{\ell=0}^{\infty} A_\ell \, r^\ell \, P_\ell(\cos\theta) \qquad \text{or} \qquad V(r,\theta) = \sum_{\ell=0}^{\infty} \frac{B_\ell}{r^{\ell+1}} \, P_\ell(\cos\theta) \tag{3.117}$$

To apply the boundary condition at R, we evaluate the above equations at that value:

$$V(R,\theta) = \sum_{\ell=0}^{\infty} A_\ell \, R^\ell \, P_\ell(\cos\theta) \qquad \text{or} \qquad V(R,\theta) = \sum_{\ell=0}^{\infty} \frac{B_\ell}{R^{\ell+1}} \, P_\ell(\cos\theta) \tag{3.118}$$

Then, to find the coefficients, we apply orthonomality to both sides, as we did for separation of variables in Cartesian coordinates. For the case of $r < R$, we have:

$$\frac{2\,\ell+1}{2}\int_0^\pi \sin\theta\, d\theta\, V(R,\theta)\, P_\ell(\cos\theta) \tag{3.119}$$

$$= \sum_{\ell'=0}^{\infty} A_{\ell'}\, R^{\ell'} \int_0^\pi \sin\theta\, d\theta\, \frac{2\,\ell+1}{2}\, P_\ell(\cos\theta)\, P_{\ell'}(\cos\theta) \tag{3.120}$$

$$= \sum_{\ell'=0}^{\infty} A_{\ell'}\, R^{\ell'}\, \delta_{\ell\ell'} = A_\ell\, R^\ell \tag{3.121}$$

which we can solve for A_ℓ. Or, based on the orthonormality relation Equation 3.113, we can just state by inspection (yielding the same result as the above calculation):

$$A_\ell = \frac{2\,\ell+1}{2}\,\frac{1}{R^\ell}\int_0^\pi \sin\theta\, d\theta\, V(R,\theta)\, P_\ell(\cos\theta) \tag{3.122}$$

Notice how R^ℓ appears in the formula for A_ℓ. This is analogous to the same way that $\sinh(\gamma_{nm}\, c)$ appeared in the solution for the coefficients A_{nm} in the Cartesian case (Equation 3.92).

Similarly, for the case $r > R$,

$$B_\ell = \frac{2\,\ell + 1}{2}\, R^{\ell+1} \int_0^\pi \sin\theta\, d\theta\, V(R,\theta)\, P_\ell(\cos\theta) \tag{3.123}$$

Therefore, the solutions are

$$V(r < R, \theta) = \sum_{\ell=0}^{\infty} \frac{2\,\ell + 1}{2}\, \frac{r^\ell}{R^\ell}\, P_\ell(\cos\theta) \int_0^\pi \sin\theta'\, d\theta'\, V(R,\theta')\, P_\ell(\cos\theta') \tag{3.124}$$

$$V(r > R, \theta) = \sum_{\ell=0}^{\infty} \frac{2\,\ell + 1}{2}\, \frac{R^{\ell+1}}{r^{\ell+1}}\, P_\ell(\cos\theta) \int_0^\pi \sin\theta'\, d\theta'\, V(R,\theta')\, P_\ell(\cos\theta') \tag{3.125}$$

Notice how the units of the coefficients cancel the powers of r in the solution so our result has the same units of electrostatic potential as the boundary condition.

Let's make some other observations, connecting to separation of variables in Cartesian coordinates.

- In our Cartesian example, we had five homogeneous boundary conditions and one inhomogeneous one. The five homogeneous ones determined the form of the individual terms in the solution: they created relationships between the coefficients, and also imposed quantization requirements, that reduced the form from being a product of three sums of two exponentials with six arbitrary argument coefficients and four arbitrary normalization coefficients to being a product of two sines and a hyperbolic sine with quantized argument coefficients with one overall arbitrary normalization coefficient. The same happened here: the homogeneous boundary condition at $r = 0$ or $r \to \infty$ eliminated one of the two coefficients in each term. (Why five homogeneous boundary conditions in the Cartesian case and only one here? Requiring single-valued behavior in ϕ and at the poles imposes another three boundary conditions, and azimuthal symmetry is a fourth. So we effectively already applied four in the form for the solution we assumed.) In the Cartesian case, those conditions had the effect of both "quantizing" the argument coefficients (restricting the freedom in the arguments of the exponentials) and restricting the normalization coefficients (showing we had only sines and hyperbolic sines, eliminating cosines and hyperbolic cosines). In this case, the "quantization" is imposed by the geometry and azimuthal symmetry from the start, yielding the "already-quantized" form we started with.

- In our Cartesian example, we applied the homogeneous boundary conditions term-by-term and then finally we were forced to consider a sum of them to match the inhomogeneous boundary condition. In this case, we started off with the sum and applied the homogeneous boundary conditions to the sum. But one can see that, by use of orthonormality, this process really was applied term-by-term. In the Cartesian case, we could not write down such a sum so early because we had not yet obtained the quantization conditions on the argument coefficients: in Cartesian coordinates, those conditions come from the specific geometry of the problem and its homogeneous boundary conditions rather than from the coordinate system. At the end of the general derivation, we did not even know whether the argument coefficients were purely real or purely imaginary numbers! Any sum would have had to be written down as an integral over an unspecified domain. So, we had to apply the homogeneous boundary conditions first to even be able to write down a sum.

- In both cases, the application of the inhomogeneous boundary condition is done to the entire sum, and the result even looks quite similar, involving an integration of the inhomogeneous boundary condition over the surface with the orthonormal functions of which the solution is composed.

Example 3.6: Dirichlet Boundary Conditions at $r = 0$ and ∞, Neumann Boundary Condition at $r = R$

Griffiths does an example in which a surface charge density is specified at $r = R$ and the potential has to be found over all of space. This is almost a Neumann boundary condition, but not quite, since the surface charge density specifies the change in the normal derivative of V at r, not the normal derivative of V itself. By solving for V over all of space, one effectively turns it into a Neumann boundary condition by using the solution in one region to specify the condition on the normal derivative as one approaches the surface from the other side. One writes down different solutions for the two regions: the $\{B_\ell\}$ vanish for the $r < R$ solution to avoid a divergence at the origin, and the $\{A_\ell\}$ vanish for the $r > R$ solution to ensure the potential vanishes at infinity (as we saw above). Then, one applies the conditions that the potential must be continuous at R (which is a homogeneous boundary condition — the difference of the potentials on the two sides of the boundary equals zero) and that the normal derivative must change by the surface charge density (divided by $-\epsilon_o$; this is the inhomogeneous boundary condition because the charge density appears on one side of the equation). The first condition is effectively the specification of $\langle V \rangle_R$, which we recall from our generic discussion of Green Functions for Neumann boundary conditions. The second condition is the actual Neumann boundary condition. The first condition relates the $\{A_\ell\}$ and $\{B_\ell\}$ at each ℓ. With now just a single set of coefficients to determine, the Neumann boundary condition can be used with the orthonormality relation to find a formula for the coefficient for each ℓ.

Note the use of two different solutions in the two regions: this is a generally useful technique.

Example 3.7: Uncharged Metal Sphere in a Uniform Field: Unusual Dirichlet Boundary Conditions

Griffiths does the example of an uncharged metal sphere in a uniform electric field in the z direction, $\vec{E} = E_0\hat{z}$. The boundary condition is a bit mixed again. Because the sphere is metal, it is an equipotential. But that doesn't specify the value of V on the sphere. Since the field is uniform, we cannot set V to vanish at infinity. Instead, $V(z = 0) = 0$ is chosen. From that choice and the fact that the equipotential sphere is in contact with $z = 0$, we can conclude that the sphere satisfies $V = 0$. But now V at infinity is not specified, so we don't yet have a Dirichlet boundary condition. The sensible thing to do is to require the potential approach $V(\vec{r}) = -E_0 z$ at infinity: whatever induced charge the sphere picks up, its contribution to the potential and field must fall off at infinity, leaving only the uniform field. Now we have a Dirichlet boundary condition. Because the potential is allowed to diverge at infinity, we cannot eliminate the $\{A_\ell\}$ in this case. But it is easy to see that only A_1 is nonzero: for $\ell > 0$, the behavior goes like r^ℓ, and since the potential must go like $z = r\cos\theta$ at large r, all the $\ell > 1$ terms must vanish. This large r behavior sets $A_1 = -E_0$. $A_0 = 0$ because the potential has no offset. That leaves the $\{B_\ell\}$ to be determined.

Applying the boundary condition $V = 0$ at $r = R$ gives:

$$0 = A_1 R \cos\theta + \sum_{\ell=0}^{\infty} \frac{B_\ell}{R^{\ell+1}} P_\ell(\cos\theta) \tag{3.126}$$

$$-A_1 R \cos\theta = \sum_{\ell=0}^{\infty} \frac{B_\ell}{R^{\ell+1}} P_\ell(\cos\theta) \tag{3.127}$$

Since the left side has a $\ell = 1$ term, and the Legendre polynomials are orthonormal, there can also be only a $\ell = 1$ term on the right side, implying $B_\ell = 0$ for $\ell \neq 1$ and $B_1/R^2 = -A_1 R$ or $B_1 = E_0 R^3$. Thus, the solution is

$$V(\vec{r}) = -E_0 \left(r - \frac{R^3}{r^2} \right) \cos\theta \tag{3.128}$$

Note the use of a nontrivial boundary condition at infinity and the need to realize that the sphere has the same potential as the $z = 0$ plane; without these boundary conditions, it would have been impossible to start the problem.

Lecture 9:

Advanced Electrostatics V:
Separation of Variables in Spherical Coordinates with Azimuthal Symmetry (cont.)
Separation of Variables in Spherical Coordinates without Azimuthal Symmetry
Spherical Harmonic Expansion of Green Functions

Date Revised: 2024/02/06 10:00
Revised lecture break
Date Given: 2024/02/01

Example 3.8: Separation of Variables for a Point Charge near a Grounded Conducting Sphere

Let's reconsider the situation we looked at before via method of images, the point charge near the conducting sphere. The setup is as before, with the point charge at $a\,\widehat{z}$ and the sphere centered on the origin with radius R and $V = 0$ on its surface. One difficulty is that the presence of the point charge implies Laplace's equation is not satisfied in the full volume! It is, however, satisfied separately in the regions $R < r < a$ and $a < r < \infty$, and we have the charge density at $r = a$, so we should somehow solve separately in the two regions and then join the solutions together (*as we did before for the spherical shell of charge, which we recognized was a Neumann boundary condition (Example 3.6)*).

Since we have seen how the method of images can provide the Green Function for a system, the aforementioned equivalence suggests that we may be able to use separation of variables to find the full Green Function for a system in the "sum over orthonormal functions" form rather than in the "system of point charges form." This is indeed true and we will do this in general fashion for spherical coordinates later in §3.9.4 using a technique similar to the one we use for this example.

We may guess that the appropriate way to write the charge density at $r = a$ is

$$\sigma(\theta, \phi) = \frac{q}{2\,\pi\,a^2 \sin\theta}\,\delta(\theta) \tag{3.129}$$

The rationale for this guess is that $a^2 \sin\theta$ cancels the $r^2 \sin\theta$ portion of the volume element and $2\,\pi$ cancels the ϕ integral. It has the right units, too, surface charge density, charge/length^2; remember, $\delta(\theta)/\sin\theta$ is unitless because θ is unitless. One can see the form is correct because integration returns q:

$$\int_0^{\pi}\int_0^{2\,\pi} da\,\sigma(\theta,\phi) = \int_0^{\pi}\sin\theta\,d\theta\int_0^{2\,\pi} d\phi\,a^2\,\frac{q}{2\,\pi\,a^2\sin\theta}\,\delta(\theta) \tag{3.130}$$

$$= \frac{1}{2\,\pi}\int_0^{\pi} d\theta\int_0^{2\,\pi} d\phi\,q\,\delta(\theta) = q \tag{3.131}$$

Notice that no $\delta(\phi)$ is required.

We can largely apply what we did in the case of the Example 3.6 except that we cannot eliminate the $\{B_\ell\}$ for $r < a$ because the inner boundary is at $r = R$, not $r = 0$. Let's apply the homogeneous boundary condition $V(r = R) = 0$ first:

$$0 = \sum_{\ell=0}^{\infty} \left(A_\ell^{in} R^\ell + \frac{B_\ell^{in}}{R^{\ell+1}} \right) P_\ell(\cos\theta) \tag{3.132}$$

where we use the ^{in} superscript to indicate these are the coefficients for the solution in the region *inside* of the charge at $r = a$; *i.e.*, the $R < r < a$ region. Since this a homogeneous boundary condition, we know from prior discussion we can apply it term-by-term. Perhaps easier to remember/justify is to apply orthonormality to the sum, which forces the coefficent of P_ℓ at each ℓ to vanish independently:

$$A_\ell^{in} R^\ell = -\frac{B_\ell^{in}}{R^{\ell+1}} \quad \Longrightarrow \quad V(r < a, \theta) = \sum_{\ell=0}^{\infty} A_\ell^{in} \left(r^\ell - \frac{R^{2\ell+1}}{r^{\ell+1}} \right) P_\ell(\cos\theta) \tag{3.133}$$

For $r > a$, we start with the same form for the solution, but of course now with different coefficients $\{A_\ell^{out}\}$ and $\{B_\ell^{out}\}$. *Do not confuse these coefficients with the $\{A_\ell^{in}\}$ and $\{B_\ell^{in}\}$ determined above: these are solutions in different regions, so they are different functions and there is no reason to expect the coefficients are the same!* The $\{A_\ell^{out}\}$ must all vanish so the potential vanishes at infinity (homogeneous boundary condition). So we have

$$V(r > a, \theta) = \sum_{\ell=0}^{\infty} \frac{B_\ell^{out}}{r^{\ell+1}} P_\ell(\cos\theta) \tag{3.134}$$

Next, we join the solutions at the boundary between them by applying the Neumann boundary condition there, which requires that V be continuous at $r = a$ and that $\partial V/\partial r$ be continuous there except at $\theta = 0$, where it has a discontinuity specified by $\sigma(0)$. We apply the first (homogeneous) condition, term-by-term like any homogeneous boundary condition or via the orthonormality of the P_ℓ:

$$A_\ell^{in}\left(a^\ell - \frac{R^{2\ell+1}}{a^{\ell+1}}\right) = \frac{B_\ell^{out}}{a^{\ell+1}} \qquad \Longrightarrow \qquad B_\ell^{out} = A_\ell^{in}\left(a^{2\ell+1} - R^{2\ell+1}\right) \tag{3.135}$$

Let's put everything we have so far together in a suggestive form:

$$V^{in}(r,\theta) \equiv V(r<a,\theta) = \sum_{\ell=0}^{\infty} A_\ell^{in}\, a^{\ell+1}\left(\frac{r^\ell}{a^{\ell+1}} - \frac{\frac{R}{a}\left(\frac{R^2}{a}\right)^\ell}{r^{\ell+1}}\right) P_\ell(\cos\theta) \tag{3.136}$$

$$V^{out}(r,\theta) \equiv V(r>a,\theta) = \sum_{\ell=0}^{\infty} A_\ell^{in}\, a^{\ell+1}\left(\frac{a^\ell}{r^{\ell+1}} - \frac{\frac{R}{a}\left(\frac{R^2}{a}\right)^\ell}{r^{\ell+1}}\right) P_\ell(\cos\theta) \tag{3.137}$$

Notice the length^{-1} units of the portion in parentheses, implying that A_ℓ^{in} will have units of ϵ_o^{-1}(length)$^{-(\ell+1)}$. Next, we apply the derivative matching (Neumann) condition:

$$\left(\frac{\partial V^{out}}{\partial r} - \frac{\partial V^{in}}{\partial r}\right)\Bigg|_{r=a} = -\frac{\sigma(\theta)}{\epsilon_o} \tag{3.138}$$

The derivatives are

$$\frac{\partial V^{in}}{\partial r} = \sum_{\ell=0}^{\infty} A_\ell^{in} a^{\ell+1} \left(\frac{\ell\, r^{\ell-1}}{a^{\ell+1}} + (\ell+1) \frac{\frac{R}{a}\left(\frac{R^2}{a}\right)^\ell}{r^{\ell+2}} \right) P_\ell(\cos\theta) \tag{3.139}$$

$$\frac{\partial V^{out}}{\partial r} = \sum_{\ell=0}^{\infty} A_\ell^{in} a^{\ell+1} (\ell+1) \left(-\frac{a^\ell}{r^{\ell+2}} + \frac{\frac{R}{a}\left(\frac{R^2}{a}\right)^\ell}{r^{\ell+2}} \right) P_\ell(\cos\theta) \tag{3.140}$$

Evaluating at $r = a$ gives

$$\left.\frac{\partial V^{in}}{\partial r}\right|_{r=a} = \sum_{\ell=0}^{\infty} A_\ell^{in} a^{\ell+1} \left(\frac{\ell}{a^2} + (\ell+1) \frac{\frac{R}{a}\left(\frac{R^2}{a}\right)^\ell}{a^{\ell+2}} \right) P_\ell(\cos\theta) \tag{3.141}$$

$$\left.\frac{\partial V^{out}}{\partial r}\right|_{r=a} = \sum_{\ell=0}^{\infty} A_\ell^{in} a^{\ell+1} (\ell+1) \left(-\frac{1}{a^2} + \frac{\frac{R}{a}\left(\frac{R^2}{a}\right)^\ell}{a^{\ell+2}} \right) P_\ell(\cos\theta) \tag{3.142}$$

When we difference the two, the second terms in the expressions cancel, leaving

$$-\sum_{\ell=0}^{\infty}(2\,\ell+1)A_{\ell}^{in}\,a^{\ell-1}\,P_{\ell}(\cos\theta)=-\frac{q\,\delta(\theta)}{2\,\pi\,a^{2}\epsilon_{o}\,\sin\theta} \tag{3.143}$$

This is our inhomogeneous boundary condition so, as usual, we must use orthonormality to obtain a formula for the coefficients in terms of an integral of the boundary condition with the orthonormal functions. We can multiply by $P_{\ell'}(\cos\theta)\sin\theta$ and integrate over θ, or we can just apply orthonormality. (Recall the orthonormality relation: $[2/(2\ell+1)]\int_0^{\pi}\sin\theta\,d\theta\,P_{\ell}(\cos\theta)\,P_{\ell'}(\cos\theta)=\delta_{\ell\ell'}$). This extracts the $A_{\ell'}^{in}$ term we want, and it also simplifies the right-hand side:

$$-2\,A_{\ell'}^{in}\,a^{\ell'-1}=-\frac{q}{2\,\pi\,a^{2}\epsilon_{o}}\int_{0}^{\pi}\sin\theta\,d\theta\,\frac{\delta(\theta)\,P_{\ell'}(\cos\theta)}{\sin\theta} \tag{3.144}$$

$$=-\frac{q}{2\,\pi\,a^{2}\epsilon_{o}}\,P_{\ell'}(\cos(\theta=0))=-\frac{q}{2\,\pi\,a^{2}\epsilon_{o}} \tag{3.145}$$

$$A_{\ell'}^{in}=\frac{1}{a^{\ell'+1}}\,\frac{q}{4\,\pi\,\epsilon_{o}} \tag{3.146}$$

Writing the full solution, we have

$$V(r<a,\theta)=\frac{q}{4\,\pi\,\epsilon_o}\sum_{\ell=0}^{\infty}\left(\frac{r^\ell}{a^{\ell+1}}-\frac{\frac{R}{a}\left(\frac{R^2}{a}\right)^\ell}{r^{\ell+1}}\right)P_\ell(\cos\theta) \tag{3.147}$$

$$V(r>a,\theta)=\frac{q}{4\,\pi\,\epsilon_o}\sum_{\ell=0}^{\infty}\left(\frac{a^\ell}{r^{\ell+1}}-\frac{\frac{R}{a}\left(\frac{R^2}{a}\right)^\ell}{r^{\ell+1}}\right)P_\ell(\cos\theta) \tag{3.148}$$

The form is hardly one we would have guessed! Separation of variables is more algorithmic than method of images, but it is also less intuitive. We will connect the two next.

Recognize that the integral over the inhomogeneous boundary condition that we expect from past experience with separation of variables has already been done on the prior page, so it is no longer explicit here; it yielded the $a^{-(\ell+1)}$ factor that is present in all the terms (in the first term for the $r>a$ solution, it is multiplied by $a^{2\ell+1}$, a factor that came from the homogeneous boundary conditions, so it is no longer obvious). Also, that integral did not include an integral over ϕ as we might have expected. We could have integrated over ϕ on both sides if we wanted, yielding a closer analogy to Equation 3.92, but it would have just yielded a common factor of $2\,\pi$ on the two sides since neither side has ϕ dependence. We did not need to do this because the problem is azimuthally symmetric and thus we know the solution must include only the $m=0$ term. We will see such an integral when we consider cases without azimuthal symmetry.

Connecting Method of Images to Separation of Variables via a Useful Expansion in Legendre Polynomials

We have two techniques — method of images and separation of variables — that we can evidently use for the same problem. By the Uniqueness Theorem, the solutions must be the same. Comparing Equations 3.147 and 3.148 that we just obtained via separation of variables to Equation 3.25 obtained via method of images, the connection is hardly obvious! To see it, we must first prove a theorem.

We will show

$$\boxed{\frac{1}{|\vec{r}-\vec{r}'|} = \sum_{\ell=0}^{\infty} \frac{r_<^\ell}{r_>^{\ell+1}} P_\ell(\cos\gamma)} \tag{3.149}$$

$$\text{with} \qquad r_< = \min(|\vec{r}|, |\vec{r}'|) \qquad r_> = \max(|\vec{r}|, |\vec{r}'|) \qquad \cos\gamma = \hat{r}\cdot\hat{r}'$$

This will let us go back and forth between separation-of-variables solutions and functions that look like the Coulomb potential (*e.g.*, point charge near the grounded sphere!). Griffiths *sort of* derives this, using a far less interesting and powerful technique. He also does it in §3.4.1, after the discussion of separation of variables, so he is unable to use this theorem to connect the method of images and separation of variables solutions for the point charge near the grounded, conducting sphere.

To prove this, orient the coordinate system so $\vec{r}\,' = r'\,\hat{z}$. The function on the left-hand side of Equation 3.149 is the potential at $\vec{r}$ of a point charge $q = 4\pi\epsilon_o$ in magnitude (not units!) at r' along the z-axis. It satisfies azimuthal symmetry and thus is expandable in terms of the above solutions of Laplace's Equation in spherical coordinates with azimuthal symmetry (because these solutions form a complete, orthonormal set!):

$$\frac{1}{|\vec{r} - \vec{r}\,'|} = \sum_{\ell=0}^{\infty} \left(A_\ell\, r^\ell + \frac{B_\ell}{r^{\ell+1}} \right) P_\ell(\cos\theta) \tag{3.150}$$

Consider two cases separately:

- $r < r'$
  We must eliminate the B_ℓ coefficients to keep the function finite as $r \to 0$. To find the A_ℓ, let's consider the point $\vec{r} = r\,\hat{z}$ (*i.e.*, $\cos\gamma = 1$), which implies

$$\frac{1}{r' - r} = \frac{1}{|\vec{r} - \vec{r}\,'|} = \sum_{\ell=0}^{\infty} A_\ell\, r^\ell \tag{3.151}$$

  (Recall, $P_\ell(1) = 1$.) Thus, the A_ℓ are just the coefficients of the power series expansion of the left side, which we know (recall: $(1-x)^{-1} = 1 + x + x^2 + \cdots$ for $0 < x < 1$) is

$$\frac{1}{r' - r} = \frac{1}{r'}\,\frac{1}{1 - \frac{r}{r'}} = \frac{1}{r'} \sum_{\ell=0}^{\infty} \left(\frac{r}{r'} \right)^\ell \tag{3.152}$$

  The series converges because $x = r/r' < 1$. Thus, $A_\ell = 1/(r')^{\ell+1}$. This now sets the $\{A_\ell\}$ for arbitrary $\vec{r}$ (*i.e.*, arbitrary $\cos\gamma$ rather than the special case $\cos\gamma = 1$ we have considered).

- $r > r'$
  We must eliminate the A_ℓ coefficients to keep the function finite as $r \to \infty$. Again, consider $\vec{r} = r\,\hat{z}$, which implies

$$\frac{1}{r - r'} = \frac{1}{|\vec{r} - \vec{r}'|} = \sum_{\ell=0}^{\infty} \frac{B_\ell}{r^{\ell+1}} \tag{3.153}$$

  For this case, we consider an expansion in r'/r rather than r/r' because now $0 < r'/r < 1$ while, above, $0 < r/r' < 1$. Again, the B_ℓ are just the coefficients of the power series expansion of the left side, which we know is

$$\frac{1}{r - r'} = \frac{1}{r}\,\frac{1}{1 - \frac{r'}{r}} = \frac{1}{r}\sum_{\ell=0}^{\infty}\left(\frac{r'}{r}\right)^{\ell} \tag{3.154}$$

  Thus, $B_\ell = (r')^\ell$.

Combining the above two cases, and generalizing back from $\cos\theta$ to $\cos\gamma$, yields Equation 3.149.

A few notes on the above derivation:

- Note some elements of technique: without loss of generality, we: a) set $\vec{r}' = r'\,\hat{z}$ so $\cos\gamma = \cos\theta$; and b) evaluated the expression at $\cos\theta = 1$, similar to the manner in which we applied the boundary conditions for the point charge near the grounded sphere. These are useful techniques to keep in mind for the future.

- Note that the process was effectively separation of variables, separately in the $r < r'$ and $r > r'$ spaces (like our separate consideration of $r < R$ and $r > R$ in the previous example) but with an unusual boundary condition: Rather than specifying a condition on the function (the "potential") on the boundary $r = r'$, we used the fact that we knew the solution along the line $\vec{r} = r\,\hat{r}'$ (which we took to be $\vec{r} = r\,\hat{z}$ in this case). That is, we specified the potential for a locus of points in the volume $\mathcal{V}$ rather than on the surface $\mathcal{S}(\mathcal{V})$. We do not have a general theorem about such boundary conditions because the derivation of the Uniqueness Theorem used Green's Theorem, which involves $\mathcal{S}$. Evidently, though, appropriate specification of the potential on some locus of points in $\mathcal{V}$ is also sufficient to yield a unique solution!

- We could have derived the result instead by treating this as a separation of variables problem with a Neumann boundary condition at $r = r'$ due to the singularity there, which is the same singularity one gets for the potential of a point charge of value $q = 4\,\pi\,\epsilon_o$ as we explained above. Explicitly, we would combine the solution to Example 3.6 (a spherical shell of charge $\sigma(R, \theta)$ in free space) with the charge density we developed for the point charge near the conducting sphere. In fact, we can read off the result from Equations 3.147 and 3.148: let $R \to 0$, replace a by r', and replace q by $4\,\pi\,\epsilon_o$! That approach is a bit more cumbersome but benefits from the Uniqueness Theorem and the full separation-of-variables machinery. We did not do that here because we knew ahead of time the solution on the $\vec{r} = r\hat{r}'$ locus, and this approach introduced you to a new technique.

With that theorem proven, we can make the advertised connection. If we compare Equations 3.147 and 3.148 from separation of variables for the point charge near the conducting sphere to Equation 3.149, we see that all four terms in the former are of the form used in the latter. The first term of the first equation has $r_< = r$ and $r_> = a$ as appropriate for $r < a$, while the first term of the second equation has $r_< = a$ and $r_> = r$ as needed for $r > a$. The second terms of both equations are of the same form with $r_< = R^2/a$, $r_> = r$ and the charge multiplied by $-R/a$. Thus, we recover

$$V(\vec{r}) = \frac{q}{4\,\pi\,\epsilon_o}\left[\frac{1}{|\vec{r} - a\,\widehat{z}|} - \frac{R/a}{\left|\vec{r} - \frac{R^2}{a}\,\widehat{z}\right|}\right] \tag{3.155}$$

which matches Equation 3.25. Remarkable! This is a case where we were able to use separation of variables to recover the full potential and thus the full method of images solution, which we know then gives us the Green Function: it is possible!

Could we have done a similar thing if we had a point charge in the five-sides-grounded box problem? There is no reason to think it would not work.

In fact, we will later show how to use a similar technique to find the Green Function in spherical coordinates for systems without azimuthal symmetry.

Separation of Variables in Spherical Coordinates without Azimuthal Symmetry

The Full Polar Equation Solution: the Associated Legendre Polynomials

There is a relation yielding the associated Legendre polynomials for $m \geq 0$ from the Legendre polynomials:

$$P_\ell^m(x) = (-1)^m (1-x^2)^{m/2} \frac{d^m}{dx^m} P_\ell(x) \tag{3.156}$$

which, using Rodrigues' Formula (Equation 3.112), implies

$$P_\ell^m(x) = \frac{(-1)^m}{2^\ell \, \ell!} (1-x^2)^{m/2} \frac{d^{\ell+m}}{dx^{\ell+m}} (x^2-1)^\ell \tag{3.157}$$

which is now valid for all m.

Some properties of the P_ℓ^m:

- $P_\ell^0 = P_\ell$ ($m = 0$ in Equation 3.156).
- The parity in x (evenness/oddness) of the associated Legendre functions is given by $(-1)^{\ell+m}$ (where -1 implies oddness): the parity of P_ℓ is given by $(-1)^\ell$, and each derivative changes the parity by a factor of -1 (note that the powers of $(1-x^2)$ have no effect on the parity because it is an even function).
- The $\{P_\ell^m\}$ for a given m form an orthogonal set. We can see this by going back to the polar equation (Equation 3.103 or 3.109) and recognizing that it is literally a *different* differential equation for every value of m. The solutions at different m are related to each other by Equation 3.156 because the differential equations are related to each other, but they are different equations and thus their solutions form different orthogonal sets. We will address their normalization imminently.

There are a number of other properties of these functions, but it is more useful to consider them together with the ϕ solutions.

The Full Solution to the Angular Piece of Laplace's Equation: the Spherical Harmonics

When one combines the $P_\ell^m(\cos\theta)$ and the $e^{im\phi}$ solutions of the polar and azimuthal equations, one obtains the *Spherical Harmonics*

$$\boxed{Y_{\ell m}(\theta,\phi) = \sqrt{\frac{2\,\ell+1}{4\,\pi}\,\frac{(\ell-m)!}{(\ell+m)!}}\,P_\ell^m(\cos\theta)\,e^{im\phi}} \tag{3.158}$$

They are an orthonormal, complete basis for functions on the sphere (θ,ϕ) (assuming the z-axis is part of the sphere; recall our comment about a second set of solutions to the Legendre equation if it is not). The $(\ell-m)!/(\ell+m)!$ factors come from normalizing the P_ℓ^m, while the $2\,\pi$ comes from normalizing the $e^{im\phi}$. They satisfy numerous important and useful conditions:

- Conjugation:

$$Y_{\ell(-m)}(\theta,\phi) = (-1)^m\,Y_{\ell m}^*(\theta,\phi) \tag{3.159}$$

- Orthonormality:

$$\int_0^{2\,\pi} d\phi \int_0^{\pi} \sin\theta\,d\theta\;Y_{\ell' m'}^*(\theta,\phi)\;Y_{\ell m}(\theta,\phi) = \delta_{\ell\ell'}\,\delta_{mm'} \tag{3.160}$$

- Completeness ($\cos\theta$ is the argument because the differential is $\sin\theta\, d\theta = -d(\cos\theta)$):

$$\sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} Y^*_{\ell m}(\theta', \phi') Y_{\ell m}(\theta, \phi) = \delta(\phi - \phi')\, \delta(\cos\theta - \cos\theta') \tag{3.161}$$

- $m = 0$ devolves to Legendre polynomials:

$$Y_{\ell\, 0}(\theta, \phi) = \sqrt{\frac{2\,\ell + 1}{4\,\pi}}\, P_\ell(\cos\theta) \tag{3.162}$$

This should be obvious from Equation 3.156, the relation between the Legendre and the associated Legendre polynomials.

- The $\theta = 0$ behavior is simple given Equation 3.156 (the $(1 - x^2)$ factor):

$$P_\ell^{m\neq 0}(\pm 1) = 0 \quad \Longrightarrow \quad Y_{\ell\, m\neq 0}(\theta = 0, \phi) = Y_{\ell\, m\neq 0}(\theta = \pi, \phi) = 0 \tag{3.163}$$

This condition ensures the $Y_{\ell\, m\neq 0}$ are single-valued at the poles. (Single-valuedness is automatic for $m = 0$ because $e^{i(0)\phi} = 1$.) Recall that we also stated $P_\ell(1) = 1$, $P_\ell(-1) = (-1)^\ell$, which implies

$$Y_{\ell\, 0}(\theta = 0, \phi) = \sqrt{\frac{2\,\ell + 1}{4\,\pi}} \qquad Y_{\ell\, 0}(\theta = \pi, \phi) = (-1)^\ell \sqrt{\frac{2\,\ell + 1}{4\,\pi}} \tag{3.164}$$

- The above implies that any expansion in terms of $Y_{\ell m}$ simplifies at $\theta = 0,\ \pi$:

$$\text{given} \quad g(\theta,\phi) = \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} A_{\ell m}\, Y_{\ell m}(\theta,\phi) \tag{3.165}$$

$$\text{then} \quad g(\theta=0,\phi) = \sum_{\ell=0}^{\infty} \sqrt{\frac{2\,\ell+1}{4\,\pi}}\, A_{\ell\,0} \tag{3.166}$$

$$\text{and} \quad g(\theta=\pi,\phi) = \sum_{\ell=0}^{\infty} (-1)^{\ell} \sqrt{\frac{2\,\ell+1}{4\,\pi}}\, A_{\ell\,0} \tag{3.167}$$

- The *Addition Theorem for Spherical Harmonics*: Given $\widehat{r}$ and $\widehat{r}\,'$ pointing in the directions (θ,ϕ) and (θ',ϕ'), respectively, then

$$\boxed{P_{\ell}(\widehat{r}\cdot\widehat{r}\,') = \frac{4\,\pi}{2\,\ell+1} \sum_{m=-\ell}^{\ell} Y^{*}_{\ell m}(\theta',\phi')\, Y_{\ell m}(\theta,\phi)} \tag{3.168}$$

where $\widehat{r}\cdot\widehat{r}\,' = \cos\gamma = \cos\theta\,\cos\theta' + \sin\theta\,\sin\theta'\cos(\phi-\phi')$. The proof of this can be found in Jackson §3.6.

- An important corollary of the Addition Theorem can be obtained by combining the above with Equation 3.149, the formula for the inverse of the relative distance between two points in terms of the Legendre polynomials:

$$\frac{1}{|\vec{r}-\vec{r}\,'|} = \sum_{\ell=0}^{\infty} \frac{r_<^\ell}{r_>^{\ell+1}} P_\ell(\cos\gamma)$$

Plugging in the Addition Theorem gives us

$$\boxed{\frac{1}{|\vec{r}-\vec{r}\,'|} = 4\pi \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} \frac{1}{2\ell+1} \frac{r_<^\ell}{r_>^{\ell+1}} Y_{\ell m}^*(\theta',\phi') Y_{\ell m}(\theta,\phi)} \qquad (3.169)$$

The utility of this relation is even more obvious than that of Equation 3.149, especially for doing integrals over charge distributions with the relative distance function (*i.e.*, calculating the potential due to Coulomb's Law): decompose the charge distribution in terms of spherical harmonics and integrate the charge distribution in a particular spherical harmonic $Y_{\ell m}$ over r' with weighting by $(r')^\ell$ to obtain the component of the potential at a distance r from the origin with spatial dependence $Y_{\ell m}(\theta,\phi)/r^{\ell+1}$.

The Full Solution of Laplace's Equation in Spherical Coordinates

Putting it all together, we see that the most general solution to Laplace's Equation in spherical coordinates is r

$$\boxed{V(r,\theta,\phi) = \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} \left(A_{\ell m}\, r^{\ell} + \frac{B_{\ell m}}{r^{\ell+1}} \right) Y_{\ell m}(\theta,\phi)} \tag{3.170}$$

Again, the coefficients $\{A_{\ell m}\}$ and $\{B_{\ell m}\}$ are set by the volume under consideration and one or the other entire set may vanish. As well, application of the boundary conditions will require the orthonormality relations for the spherical harmonics.

As with the case of azimuthal symmetry, we note that, in the process of doing separation of variables, we have proven that the angular solution satisfies the *eigenvalue-eigenfunction equation*

$$\boxed{\nabla^2 Y_{\ell m}(\theta,\phi) = -\frac{\ell(\ell+1)}{r^2} Y_{\ell m}(\theta,\phi)} \tag{3.171}$$

As before, the appearance of r^2 on the right side is not surprising. Note also that m does not appear in the angular equation. This is because Laplace's Equation itself is spherically (and therefore azimuthally) symmetric. The charge distribution and boundary conditions are what may break the spherical symmetry.

Expansion of the Green Function in Spherical Coordinates in Terms of the Spherical Harmonics

The fact that the spherical harmonics combined with the usual power laws in radius solve Laplace's Equation for problems that are separable in spherical coordinates can be used to show that the Green Function for such problems will have a convenient expansion in terms of the radial solutions and spherical harmonics, like Equation 3.170. It is convenient to recall at this point that a Green Function is specified (is unique) once one specifies the geometry and the type of boundary condition; the value of the boundary condition *does not* affect the Green Function. So, once we have specified a geometry and *type* of boundary condition, the expansion can be determined and is unique. Alternatively, one can think of this expansion as a generalization of the corollary of the Addition Theorem, Equation 3.169. It is shown by using the completeness property of the spherical harmonics and the eigenvalue-eigenfunction equation for the angular solution. But let's see that this is true explicitly for a couple example geometries first:

- *Free space*
  The corollary of the Addition Theorem above is the desired expansion of the Green Function for charge in free space with no finite radius boundaries and with the condition $V \to 0$ as $r \to \infty$.

- *Point charge near a grounded, conducting sphere*
  For this geometry, we saw that the Green Function can be written as sum of the Coulomb potential of two point charges, the original one at $r'\,\hat{z}$ and the image charge $q' = -q\,R/r'$ at $\hat{z}\,R^2/r'$:

$$G(\vec{r},\vec{r}') = \frac{1}{4\,\pi\,\epsilon_o}\left[\frac{1}{|\vec{r}-\vec{r}'|} - \frac{R/r'}{\left|\vec{r}-\vec{r}'\left(\frac{R}{r'}\right)^2\right|}\right] \tag{3.172}$$

Using the same corollary of the Addition Theorem, we can immediately write (using the fact $r'\,(R/r')^2 < r'$ always because the the image charge is always at radius $< R$ while the true charge is at $r' > R$):

$$G(\vec{r},\vec{r}') = \frac{1}{\epsilon_o}\sum_{\ell=0}^{\infty}\sum_{m=-\ell}^{\ell}\left[\frac{r_<^\ell}{r_>^{\ell+1}} - \frac{R}{r'}\frac{\left[r'\left(\frac{R}{r'}\right)^2\right]^\ell}{r^{\ell+1}}\right]\frac{Y^*_{\ell m}(\theta',\phi')\,Y_{\ell m}(\theta,\phi)}{2\,\ell+1} \tag{3.173}$$

$$= \frac{1}{\epsilon_o}\sum_{\ell=0}^{\infty}\sum_{m=-\ell}^{\ell}\left[\frac{r_<^\ell}{r_>^{\ell+1}} - \frac{1}{R}\left(\frac{R^2}{r\,r'}\right)^{\ell+1}\right]\frac{Y^*_{\ell m}(\theta',\phi')\,Y_{\ell m}(\theta,\phi)}{2\,\ell+1} \tag{3.174}$$

Note also the symmetry in $\vec{r}$ and $\vec{r}'$ is manifest.

In both cases, we finally have forms for the Green Function that could plausibly come from separation of variables. Note, however, that we did not use separation of variables to obtain it; we used the method of images combined with the corollary of the Addition Theorem.

Earlier, we solved for the potential of the latter configuration using separation of variables with azimuthal symmetry, Equations 3.147 and 3.148 reproduced here but rewritten using the $r_<$, $r_>$ notation:

$$V(r,\theta) = \frac{q}{4\,\pi\,\epsilon_o}\sum_{\ell=0}^{\infty}\left(\frac{r_<^{\ell}}{r_>^{\ell+1}} - \frac{\frac{R}{a}\left(\frac{R^2}{a}\right)^{\ell}}{r^{\ell+1}}\right)P_\ell(\cos\theta)$$

$$\text{with} \quad r_< = \min(r,a) \qquad r_> = \max(r,a) \tag{3.175}$$

Why was this not enough to give us the full Green Function? Because this solution for the potential in terms of Legendre polynomials assumed the point charge was along the z-axis.

What we can do is generalize this solution by replacing $\cos\theta$ with $\cos\gamma = \widehat{r}\cdot\widehat{r}'$ and a with r' followed by application of the Addition Theorem. Then the solution would be in a form where we could read off the Green Function expansion in spherical coordinates not assuming azimuthal symmetry. But this approach is not the same as obtaining the solution directly, and clearly the above approach does not generalize to a system that does not have azimuthal symmetry.

The general approach to the problem of finding the Green Function for an arbitrary (spherical) geometry is to go back to the definition of the Green Function:

$$-\epsilon_o \nabla^2_{\vec{r}} G(\vec{r}, \vec{r}\,') = \delta(\vec{r} - \vec{r}\,') \tag{3.176}$$

and decompose both sides in terms of spherical harmonics. We do not know the Green Function yet, so its expansion is the arbitrary general form, which here we write

$$G(\vec{r}, \vec{r}\,') = \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} A_{\ell m}(r|\vec{r}\,')\, Y_{\ell m}(\theta, \phi) \tag{3.177}$$

where the coefficients in the expansion $A_{\ell m}$ depend on r, as usual, and they also depend parametrically on $\vec{r}\,'$ because it is a parameter in the differential equation. (We do not know the solutions for the radial dependence of the $A_{\ell m}$ yet for the general case we are trying to solve (which is not Laplace's Equation!), so we cannot assume they are the power laws we saw for solutions to Laplace's Equation.)

The right side can be rewritten using the breakdown of the delta function into delta functions in each spherical coordinate followed by completeness of the spherical harmonics. The breakdown of the delta function is:

$$\delta(\vec{r}-\vec{r}\,') = \frac{\delta(r-r')}{r^2}\,\delta(\phi-\phi')\,\delta(\cos\theta-\cos\theta') \tag{3.178}$$

The $1/r^2$ on the radial component is required to cancel the r^2 in the volume element in spherical coordinates. The fact that the delta function in θ is a function of $\cos\theta$ and $\cos\theta'$ is because the volume element contains $\sin\theta\,d\theta = d(\cos\theta)$. One could have instead written $\delta(\theta-\theta')/\sin\theta$ as we did when rewriting the point charge near the grounded, conducting sphere as a surface charge density $\sigma(\theta)$, Equation 3.129. Using completeness of the spherical harmonics, we have

$$\delta(\vec{r}-\vec{r}\,') = \frac{\delta(r-r')}{r^2}\sum_{\ell=0}^{\infty}\sum_{m=-\ell}^{\ell} Y^*_{\ell m}(\theta',\phi')\,Y_{\ell m}(\theta,\phi) \tag{3.179}$$

Thus, our differential equation for the Green Function becomes

$$-\epsilon_o \nabla^2_{\vec{r}} \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} A_{\ell m}(r|\vec{r}\,') \, Y_{\ell m}(\theta,\phi) = \frac{\delta(r-r')}{r^2} \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} Y^*_{\ell m}(\theta',\phi') \, Y_{\ell m}(\theta,\phi) \tag{3.180}$$

Note that the Laplacian acts on the unprimed coordinates only. When we evaluate the action of the Laplacian, a cross term $\vec{\nabla}_{\vec{r}} A_{\ell m}(r|\vec{r}\,') \cdot \vec{\nabla}_{\vec{r}} Y_{\ell m}(\theta,\phi)$ appears, but it vanishes because the first term points along $\widehat{r}$ while the second is along $\widehat{\theta}$ and $\widehat{\phi}$, leaving only $\nabla^2_{\vec{r}}$ acting on each factor in the product individually. We wrote down earlier Equation 3.171, the eigenvalue-eigenfunction equation satisfied by the angular solutions of Laplace's Equation, which we use here to evaluate $\nabla^2_{\vec{r}} Y_{\ell m}(\theta,\phi)$:

$$\begin{aligned} -\epsilon_o \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} \left[\left(\nabla^2_{\vec{r}} - \frac{\ell(\ell+1)}{r^2} \right) A_{\ell m}(r|\vec{r}\,') \right] Y_{\ell m}(\theta,\phi) \\ = \frac{\delta(r-r')}{r^2} \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} Y^*_{\ell m}(\theta',\phi') \, Y_{\ell m}(\theta,\phi) \end{aligned} \tag{3.181}$$

Note that the Laplacian on the left side is now acting with its radial derivatives only on $A_{\ell m}$; its action on the spherical harmonics has yielded the $\ell(\ell+1)/r^2$ term.

Lecture 10:

Advanced Electrostatics VI:

Spherical Harmonic Expansion of Green Functions (cont.)
Examples of Using the Spherical Harmonic Expansion
for Green Functions

Date Revised: 2024/02/08 06:15
Revised selection of skipped material
Date Given: 2024/02/06

The coefficients of the individual $Y_{\ell m}(\theta, \phi)$ on the two sides must be equal because of the orthonormality relation for the spherical harmonics, implying

$$-\epsilon_o \left[\left(\nabla^2_{\vec{r}} - \frac{\ell(\ell+1)}{r^2} \right) A_{\ell m}(r|\vec{r}\,') \right] = \frac{\delta(r-r')}{r^2} \, Y^*_{\ell m}(\theta', \phi') \tag{3.182}$$

Now, given that we have $Y^*_{\ell m}(\theta', \phi')$ on the right side (from applying completeness), and again the spherical harmonics are orthonormal functions, the dependence of $A_{\ell m}(r|\vec{r}\,')$ on its $\vec{r}\,'$ angular coordinates must be proportional to $Y^*_{\ell m}(\theta', \phi')$. Therefore, we may write (with $g_\ell(r, r')$ still to be determined)

$$A_{\ell m}(r|r', \theta', \phi') = g_\ell(r, r') \, Y^*_{\ell m}(\theta', \phi') \tag{3.183}$$

Plugging in this form to the above reduced version of Laplace's Equation and canceling $Y^*_{\ell m}(\theta', \phi')$, we get:

$$-\epsilon_o \left(\nabla^2_{\vec{r}} - \frac{\ell(\ell+1)}{r^2} \right) g_\ell(r, r') = \frac{\delta(r-r')}{r^2} \tag{3.184}$$

Only the Laplacian's radial derivatives yield a nonzero contribution here, so we have (also multiplying both sides by $-r^2/\epsilon_o$):

$$\frac{d}{dr} \left[r^2 \frac{d}{dr} \, g_\ell(r, r') \right] - \ell(\ell+1) \, g_\ell(r, r') = -\frac{\delta(r-r')}{\epsilon_o} \tag{3.185}$$

We see that, when $r \neq r'$ (r' is a parameter, not a variable, here), $g_\ell(r, r')$ satisfies the radial ODE in r from separation of variables in spherical coordinates, Equation 3.102. Therefore, in the two separate regions $r < r'$ and $r > r'$, the solutions to that ODE are also our solutions here:

$$g_\ell(r, r') = \begin{cases} A_\ell^{in}(r')\, r^\ell + B_\ell^{in}(r')\, r^{-(\ell+1)} & r < r' \\ A_\ell^{out}(r')\, r^\ell + B_\ell^{out}(r')\, r^{-(\ell+1)} & r > r' \end{cases} \tag{3.186}$$

Because r' is a parameter of the differential equation, the coefficients and therefore the solutions depend on it parametrically. Therefore, the general form for the expansion of the Green Function in spherical harmonics is

$$\boxed{r < r' : G(\vec{r}, \vec{r}') = \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} \left[A_\ell^{in}(r')\, r^\ell + \frac{B_\ell^{in}(r')}{r^{\ell+1}} \right] Y_{\ell m}(\theta, \phi)\, Y_{\ell m}^*(\theta', \phi')} \tag{3.187}$$

$$\boxed{r > r' : G(\vec{r}, \vec{r}') = \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} \left[A_\ell^{out}(r')\, r^\ell + \frac{B_\ell^{out}(r')}{r^{\ell+1}} \right] Y_{\ell m}(\theta, \phi)\, Y_{\ell m}^*(\theta', \phi')} \tag{3.188}$$

To determine the coefficients, we need to apply boundary conditions. Since we have not yet specified the geometry and boundary conditions, the only generic boundary condition we can write down is the one at $r = r'$, which we obtain by integrating Equation 3.185 from $r = r' - \epsilon$ to $r = r' + \epsilon$ and letting $\epsilon \to 0$:

$$\int_{r'-\epsilon}^{r'+\epsilon} dr \left\{ \frac{d}{dr}\left[r^2 \frac{d}{dr} g_\ell(r, r') \right] - \ell(\ell+1)\, g_\ell(r, r') \right\} = -\int_{r'-\epsilon}^{r'+\epsilon} dr \frac{\delta(r - r')}{\epsilon_o} \tag{3.189}$$

The first term is the integral of a total differential, so it is trivially integrated. For the second term, the form of $g_\ell(r, r')$, where it is sum of two terms, each of which includes a power law in r and some function of r' not dependent on r, ensures it cannot diverge at $r = r'$. Therefore, the second term is an integral of a function with no singularity at $\epsilon = 0$ (*i.e.*, at $r = r'$) and thus, as $\epsilon \to 0$, that integral vanishes. The right side gives $-1/\epsilon_0$ when integrated. Therefore, we have

$$\begin{aligned} \lim_{\epsilon \to 0} \left[r^2 \frac{d}{dr} g_\ell(r, r') \right] \Big|_{r=r'-\epsilon}^{r=r'+\epsilon} &= -\frac{1}{\epsilon_o} \\ \frac{d}{dr} g_\ell^{out}(r, r') \Big|_{r=r'} - \frac{d}{dr} g_\ell^{in}(r, r') \Big|_{r=r'} &= -\frac{1}{\epsilon_0 (r')^2} \end{aligned} \tag{3.190}$$

where $g_\ell^{out}(r, r')$ is the $r > r'$ solution and $g_\ell^{in}(r, r')$ is the $r < r'$ solution. This is a Neumann-type boundary condition as we had for the examples of the arbitrary charge density on a sphere $\sigma(R, \theta)$ and for the point charge near the conducting sphere $\sigma(a, \theta) = \delta(\cos\theta)/2\,\pi\, a^2$, but this form is equivalent to what we obtained in those cases *after* we used orthonormality in θ to isolate the expansion coefficients.

We note, as an aside, that we derived the above matching condition Equation 3.190 in a somewhat different way here than when we considered the above examples. In the examples, we used the fact that we knew the boundary condition on the normal derivative of the potential from Gauss's Law. Here, we effectively rederived that boundary condition for the special case of a radial boundary because we have only to determine the radial function $g(r, r')$. We could have gone back a step and written down the boundary condition on the normal derivative of the potential and derived the same condition above, but it would have required going back to the full potential and applying orthonormality and completeness again. We circumvented that step by rederiving the boundary condition considering only the radial function.

Evaluating the above condition explicitly using the $r < r'$ and $r > r'$ pieces of the solution, and multiplying both sides by $(r')^2$, we obtain

$$\ell \left[A_\ell^{in}(r') - A_\ell^{out}(r') \right] (r')^{\ell+1} + (\ell+1) \left[B_\ell^{out}(r') - B_\ell^{in}(r') \right] (r')^{-\ell} = \frac{1}{\epsilon_o} \quad (3.191)$$

Since A_ℓ^{in}, B_ℓ^{in}, A_ℓ^{out}, and B_ℓ^{out} all depend on r', all the powers of r' match up. The above is an inhomogeneous boundary condition on $g_\ell(r, r')$.

The finite discontinuity in the radial derivative of $g_\ell(r, r')$ implies that $g_\ell(r, r')$ itself must be continuous at $r = r'$: the derivative would need to have a singularity in order for there to be a discontinuity in $g_\ell(r, r')$. Therefore, we also have the condition

$$g_\ell^{out}(r = r', r') - g_\ell^{in}(r = r', r') = 0 \tag{3.192}$$

Explicitly evaluating this condition, again using the two portions of the solution, yields

$$\left[A_\ell^{in}(r') - A_\ell^{out}(r')\right](r')^{2\,\ell+1} + \left[B_\ell^{in}(r') - B_\ell^{out}(r')\right] = 0 \tag{3.193}$$

The above is a homogeneous boundary condition on $g_\ell(r, r')$.

The above two matching conditions, along with application of the boundary conditions that define Dirichlet or Neumann Green Functions (Equations 3.47 and 3.49), provide four conditions for the four unknowns $A_\ell^{in}(r')$, $B_\ell^{in}(r')$, $A_\ell^{out}(r')$, and $B_\ell^{out}(r')$, which should fully specify them. We finally have a completely algorithmic way to obtain the full Green Function! What a powerful technique! This general approach can be applied for any coordinate system in which Laplace's Equation and the boundary conditions are separable.

We also note that the above two equations imply the solutions $A_\ell^{in}(r')$, $B_\ell^{in}(r')$, $A_\ell^{out}(r')$, and $B_\ell^{out}(r')$ will be power laws in r'. This is sensible: because we expect the Green Function to be symmetric in $\vec{r}$ and $\vec{r}'$, the functional dependences on $\vec{r}$ and $\vec{r}'$, and thus on r and r', must be the same, and, so, because G has power-law dependence on r, it must also for r'.

We note that the above procedure is almost identical to what we would have done had we used separation of variables more explicitly. We certainly assumed a SoV-style expansion for the Green Function, Equation 3.177. As noted above, we could have applied a Neumann-style boundary condition at $r = r'$ by arguing that $G(\vec{r}, \vec{r}')$ is the potential at $\vec{r}$ due to a point charge at $\vec{r}'$. That would have yielded the same conditions on $g_\ell(r, r')$ that integrating the ODE, Equation 3.185, yielded. We chose to do it the way we did so we would have to rely less on such analogies. But one can see that the approaches are equivalent.

Example 3.9: Expansion in spherical harmonics for the Green Function for $R < r < \infty$ with Dirichlet boundary conditions at $r = R$ and $r \to \infty$

These boundary conditions impose the requirement $G_D(\vec{r}, \vec{r}') = 0$ for $\vec{r} \in \mathcal{S}, \mathcal{V}$, $\vec{r}' \in \mathcal{S}$. We use the symmetry of the Dirichlet Green Function to convert this to the requirement $G_D(\vec{r}, \vec{r}') = 0$ for $\vec{r} \in \mathcal{S}$, $\vec{r}' \in \mathcal{S}, \mathcal{V}$ because we do not know the dependence of the coefficients on r' and we want to obtain relations between the coefficients of the expansion that are valid at all r', not just values on the boundary, because those full dependences are needed to use the matching conditions at $r = r'$ that we just derived. One can check that applying these conditions at $\vec{r}' \in \mathcal{S}$ does not result in useful information.

Our condition implies

$$0 = G_D(\vec{r} \in \mathcal{S}, \vec{r}' \in \mathcal{S}, \mathcal{V}) = \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} g_\ell(r \in \mathcal{S}, r' \in \mathcal{S}, \mathcal{V})\, Y^*_{\ell m}(\theta', \phi')\, Y_{\ell m}(\theta, \phi) \tag{3.194}$$

Applying orthonormality of the $Y_{\ell m}(\theta', \phi')$, we obtain

$$0 = g_\ell(r \in \mathcal{S}, r' \in \mathcal{S}, \mathcal{V})\, Y_{\ell m}(\theta, \phi) \tag{3.195}$$

By applying orthonormality of the $Y_{\ell m}(\theta, \phi)$, or simply noting $Y_{\ell m}(\theta, \phi)$ is in general nonzero, we obtain

$$g_\ell(r \in \mathcal{S}, r' \in \mathcal{S}, \mathcal{V}) = 0 \tag{3.196}$$

We will apply the above homogeneous boundary condition first, then the matching conditions (one homogeneous and one inhomogeneous boundary condition) at $r = r'$, because the Dirichlet BC are simpler algebraically (this is the same order of steps we used when we solved this problem using separation of variables).

First, consider the boundary at $r = R$. Since $r' > r = R$ for all $\vec{r}' \in \mathcal{V}$, this implies that we should require $g_\ell(r = R, r') = 0$ for the $r < r'$ solution, yielding:

$$A_\ell^{in}(r')\,R^\ell + B_\ell^{in}(r')\,R^{-(\ell+1)} = 0 \qquad \Longrightarrow \qquad B_\ell^{in}(r') = -R^{2\,\ell+1}\,A_\ell^{in}(r') \quad (3.197)$$

Next, consider the boundary at $r \to \infty$: $g_\ell(r \to \infty, r' \in \mathcal{V}) = 0$. Here, it is the $r > r'$ solution that applies, which implies $A_\ell^{out}(r') = 0$ for all r'.

Next, we apply the matching conditions at $r = r'$. First, the homogeneous one, continuity of $g_\ell(r, r')$ at $r = r'$ (Equation 3.193), which implies

$$A_\ell^{in}(r')(r')^{2\ell+1} + \left\{\left[-A_\ell^{in}(r')\,R^{2\ell+1}\right] - B_\ell^{out}(r')\right\} = 0$$

$$\Longrightarrow \qquad B_\ell^{out}(r') = A_\ell^{in}(r')\left[(r')^{2\ell+1} - R^{2\ell+1}\right] \qquad (3.198)$$

The condition on the change in the radial derivative at $r = r'$, the inhomogeneous boundary condition, yielded Equation 3.191, which we plug into to obtain

$$\ell\,A_\ell^{in}(r')\,(r')^{\ell+1} + (\ell+1)A_\ell^{in}(r')\left[(r')^{2\ell+1} - R^{2\ell+1} + R^{2\ell+1}\right](r')^{-\ell} = \frac{1}{\epsilon_o}$$

$$\Longrightarrow \qquad A_\ell^{in}(r') = \frac{1}{2\,\ell+1}\,\frac{1}{\epsilon_o}\,\frac{1}{(r')^{\ell+1}} \qquad (3.199)$$

Putting it all together, we have that the Green Function for this Dirichlet boundary condition, expanded in terms of spherical harmonics, is

$$G(\vec{r},\vec{r}\,') = \begin{cases} \frac{1}{\epsilon_o}\sum_{\ell=0}^{\infty}\sum_{m=-\ell}^{\ell}\left[\frac{r^\ell}{(r\,')^{\ell+1}} - \frac{1}{R}\left(\frac{R^2}{r\,r\,'}\right)^{\ell+1}\right]\frac{Y_{\ell m}(\theta,\phi)\,Y^*_{\ell m}(\theta\,',\phi\,')}{2\,\ell+1} & r < r\,' \\ \frac{1}{\epsilon_o}\sum_{\ell=0}^{\infty}\sum_{m=-\ell}^{\ell}\left[\frac{(r\,')^\ell}{r^{\ell+1}} - \frac{1}{R}\left(\frac{R^2}{r\,r\,'}\right)^{\ell+1}\right]\frac{Y_{\ell m}(\theta,\phi)\,Y^*_{\ell m}(\theta\,',\phi\,')}{2\,\ell+1} & r > r\,' \end{cases} \tag{3.200}$$

$$= \frac{1}{\epsilon_o}\sum_{\ell=0}^{\infty}\sum_{m=-\ell}^{\ell}\left[\frac{r_<^\ell}{r_>^{\ell+1}} - \frac{1}{R}\left(\frac{R^2}{r\,r\,'}\right)^{\ell+1}\right]\frac{Y_{\ell m}(\theta,\phi)\,Y^*_{\ell m}(\theta\,',\phi\,')}{2\,\ell+1} \tag{3.201}$$

This solution is of course consistent with Equation 3.174, where we used the Addition Theorem for Spherical Harmonics to rewrite the Green Function for this geometry and type of boundary conditions in terms of the spherical harmonics, except now that we used separation of variables from the start rather than relying on the method of images and the Addition Theorem.

Note that, as predicted, the solution consists of sums of power laws in $r\,'$ as well as r and is of course symmetric under exchange of $\vec{r}$ and $\vec{r}\,'$.

Interesting exercises would be to see that the above expression approaches $|\vec{r}-\vec{r}\,'|^{-1}$ as $\vec{r}\to\vec{r}\,'$ (use the Addition Theorem to recover the method of images solution) and also to recover the defining differential equation, Equation 3.176.

Examples of Using the Expansion of the Green Function in Terms of the Spherical Harmonics

We did a lot of gymnastics to get the expansion of the Green Function in terms of spherical harmonics. Let's see how it can be used. For each of the examples we will consider, it would be possible to solve for the potential without explicitly using our expansion by splitting the volume into regions on two sides of $\vec{r}'$ and using separation of variables with application of boundary conditions (including matching conditions at the chosen internal boundary). The advantage of using the Green Function is that it obviates re-solving the same kind of problem many times by simply providing integrals that need to be done.

Another point: in separation of variables, we always end up using orthonormality of the specific set of solutions to Laplace's Equation for the geometry to obtain the solution coefficients from the inhomogeneous boundary condition(s), Dirichlet or Neumann, and the matching conditions, if any. That general approach will become codified here in the way the Green Function is integrated with the charge distribution and boundary conditions in Equations 3.47 and 3.49. In particular, the Green Function connects particular spherical harmonic modes of the charge distribution and/or the voltage (Dirichlet) and/or charge (Neumann) boundary conditions to the corresponding spherical harmonic modes of the potential. This correspondence makes the structure of the solution much easier to understand. The effect of a spherical harmonic mode in charge distribution and/or the boundary conditions at one radius r' on the potential at another radius r is just a function of the two radii, the $g(r, r')$ function (charge distribution in volume or Neumann boundary condition) or its radial derivative (Dirichlet boundary condition).

The application of the Green Function is like a propagator in QM, propagating from the initial condition to later times. We have to do less work to obtain the QM propagator because the solution to the time piece of Schrödinger's Equation is trivial once one has the eigenvalues of the space piece.

For our examples, we will consider charge distributions inside a conducting sphere. We quote the general result from Jackson for the Green Function expansion in spherical harmonics for a geometry consisting of the volume between two spheres at $r = a$ and $r = b$ with Dirichlet BC on the two surfaces:

$$G_D(\vec{r}, \vec{r}\,') \tag{3.202}$$
$$= \frac{1}{\epsilon_o} \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} \left[r_<^\ell - \frac{1}{a} \left(\frac{a^2}{r_<} \right)^{\ell+1} \right] \left[\frac{1}{r_>^{\ell+1}} - \frac{1}{b} \left(\frac{r_>}{b^2} \right)^\ell \right] \frac{Y^*_{\ell m}(\theta', \phi')\, Y_{\ell m}(\theta, \phi)}{\left[1 - \left(\frac{a}{b} \right)^{2\,\ell+1} \right] (2\,\ell + 1)}$$

where, as usual, $r_< = \min\{r, r'\}$ and $r_> = \max\{r, r'\}$. Obtaining this more general result is a matter of doing the same thing as we did to obtain the result for a spherical conducting boundary at $r = R$ except that the A_ℓ^{out} term cannot be assumed to vanish.

Next, taking the limit $a \to 0$, we get the result we will need for our work below where we want to solve for the potential inside a sphere at $r = b$ with Dirichlet BC:

$$G_D(\vec{r}, \vec{r}\,') = \frac{1}{\epsilon_o} \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} r_<^\ell \left[\frac{1}{r_>^{\ell+1}} - \frac{1}{b} \left(\frac{r_>}{b^2} \right)^\ell \right] \frac{Y^*_{\ell m}(\theta', \phi')\, Y_{\ell m}(\theta, \phi)}{2\,\ell + 1} \tag{3.203}$$

You will also be able to read off this simpler result from a method of images problem you will do in homework. On to our examples!

Example 3.10: Potential inside a conducting sphere of radius b due to an arbitrary Dirichlet boundary condition potential at b but no charge in the volume

With no charge in the volume, we just need to calculate the surface term in Equation 3.48, for which we need the normal gradient of G_D at the surface (remember, $\widehat{n}$ points out of $\mathcal{V}$):

$$\begin{aligned}
&\widehat{n}(\vec{r}\,') \cdot \vec{\nabla}_{\vec{r}\,'} G_D(\vec{r}, \vec{r}\,') \Big|_{\vec{r}\,' \in \mathcal{S}} \\
&= \frac{1}{\epsilon_o} \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} \frac{Y^*_{\ell m}(\theta', \phi')\, Y_{\ell m}(\theta, \phi)}{2\,\ell + 1}\, r^{\ell} \frac{d}{dr'} \left[\frac{1}{(r')^{\ell+1}} - \frac{1}{b} \left(\frac{r'}{b^2} \right)^{\ell} \right] \Bigg|_{r'=b} \\
&= -\frac{1}{\epsilon_o} \frac{1}{b^2} \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} \left(\frac{r}{b} \right)^{\ell} Y^*_{\ell m}(\theta', \phi')\, Y_{\ell m}(\theta, \phi) \qquad (3.204)
\end{aligned}$$

Therefore, the potential in the volume for the Dirichlet B.C. $V(b, \theta, \phi)$ is

$$V(\vec{r}) = \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} \left(\frac{r}{b} \right)^{\ell} Y_{\ell m}(\theta, \phi) \int d\Omega'\, Y^*_{\ell m}(\theta', \phi')\, V(b, \theta', \phi') \qquad (3.205)$$

We see that the spherical harmonic component ℓm of the potential at r is determined by the spherical harmonic component ℓm of the potential on the boundary: very simple and consistent with the QM propagator picture.

Example 3.11: Potential inside a grounded spherical conductor with a ring of charge of radius a in the xy plane

This time, we do the volume integral but there is no integral over the surface. The charge density due to the ring is

$$\rho(\vec{r}\,') = \frac{Q}{2\,\pi\,a^2}\,\delta(r'-a)\,\delta(\cos\theta') \tag{3.206}$$

Again, one can check that the charge density is correct by integrating it: the a^{-2} cancels the $(r')^2$ factor in the volume element and the argument of the θ' delta function is $\cos\theta'$ because the volume element contains $d(\cos\theta')$.

We use Equation 3.48 as usual, in this case with no surface term because the boundary has $V = 0$. The potential is then

$$\begin{aligned} V(\vec{r}) &= \int_{\mathcal{V}} d\tau'\rho(\vec{r}\,')\,G_D(\vec{r},\vec{r}\,') \\ &= \frac{Q}{2\,\pi\,\epsilon_o\,a^2}\sum_{\ell=0}^{\infty}\sum_{m=-\ell}^{\ell} Y_{\ell m}(\theta,\phi) \\ &\qquad \times \int_{\mathcal{V}} d\tau'\delta(r'-a)\,\delta(\cos\theta')\,r_<^{\ell}\left[\frac{1}{r_>^{\ell+1}} - \frac{1}{b}\left(\frac{r_>}{b^2}\right)^{\ell}\right]\frac{Y^*_{\ell m}(\theta',\phi')}{2\,\ell+1} \end{aligned} \tag{3.207}$$

Because the charge density has no azimuthal dependence, the ϕ' integral picks out the $m = 0$ term. Recall that $Y_{\ell 0} = \sqrt{(2\ell+1)/4\,\pi} P_\ell$, so we may rewrite as

$$V(\vec{r}) = \frac{Q}{4\,\pi\,\epsilon_o\,a^2} \sum_{\ell=0}^{\infty} P_\ell(\cos\theta) \int_{-1}^{1} d(\cos\theta')\,\delta(\cos\theta')\,P_\ell(\cos\theta') \tag{3.208}$$

$$\times \int_0^b (r')^2 dr'\,\delta(r'-a)\,r_<^\ell \left[\frac{1}{r_>^{\ell+1}} - \frac{1}{b}\left(\frac{r_>}{b^2}\right)^\ell\right]$$

$$= \frac{Q}{4\,\pi\,\epsilon_o} \sum_{\ell=0}^{\infty} P_\ell(\cos\theta)\,P_\ell(0)\,r_<^\ell \left[\frac{1}{r_>^{\ell+1}} - \frac{1}{b}\left(\frac{r_>}{b^2}\right)^\ell\right] \tag{3.209}$$

where now $r_< = \min\{r, a\}$ and $r_> = \max\{r, a\}$ because the δ function does the r' integral for us, effectively replacing r' with a. (The next example will show the case of a more complex charge distribution for which the radial integral is not done so easily.) Now, recall $P_\ell(0) = 0$ for odd ℓ and $P_\ell(0) = [(-1)^n\,(2n-1)!!]/2^n\,n!$ for even $\ell = 2\,n$, so we may reduce the above further to (replacing ℓ with $2\,n$ so n runs over all nonnegative integers rather than ℓ running over all nonnegative even integers):

$$V(\vec{r}) = \frac{Q}{4\,\pi\,\epsilon_o} \sum_{n=0}^{\infty} \frac{(-1)^n\,(2n-1)!!}{2^n\,n!} r_<^{2n} \left[\frac{1}{r_>^{2n+1}} - \frac{1}{b}\left(\frac{r_>}{b^2}\right)^{2n}\right] P_{2n}(\cos\theta) \tag{3.210}$$

where $r_< = \min(r, a)$ and $r_> = \max(r, a)$ again: *i.e.*, not surprisingly, the solution has a different form depending on whether one wants to know the potential inside the ring ($r < a$) or outside the ring ($r > a$). This is now the complete solution.

To get some intuition for the solution, let's calculate the induced surface charge density at $r = b$. We obtain it from the normal gradient of V, which, recall, is just $\vec{E}$, and the change in its normal component at a boundary gives the surface charge density. Since the normal gradient is just d/dr for this particular geometry, it does not act at all on P_{2n}. In calculating this gradient, $r_< = a$ and $r_> = r$ since we will in the end evaluate at $r = b > a$. Therefore:

$$\begin{aligned}
\sigma(\vec{r}) &= \epsilon_o \left.\frac{dV}{dr}\right|_{r_>=r=b, r_<=a} \\
&= -\frac{Q}{4\,\pi\,b^2} \sum_{n=0}^{\infty} \frac{(4n+1)\,(-1)^n\,(2n-1)!!}{2^n\,n!} \left(\frac{a}{b}\right)^{2n} P_{2n}(\cos\theta) \\
&= -\frac{Q}{4\,\pi\,b^2} \left[1 + \sum_{n=1}^{\infty} \frac{(4n+1)\,(-1)^n\,(2n-1)!!}{2^n\,n!} \left(\frac{a}{b}\right)^{2n} P_{2n}(\cos\theta)\right] \qquad (3.211)
\end{aligned}$$

The expression is written in the above suggestive form on the last line so that it is easy to obtain the total induced surface charge. Since $P_0(\cos\theta) = 1$, the integral of the $n > 0$ terms over $\cos\theta$ can be viewed as integrating P_{2n} with P_0; by orthonormality of the Legendre polynomials, these terms all yield zero. The first term yields $-Q$ when integrated over the sphere. This is what we would expect from Gauss's Law applied just inside the $r = b$ boundary. All the other terms yield θ dependence in the charge density but their averages vanish (easily seen by using orthonormality with P_0, which is a constant).

We have seen in this example how the integration of the charge density with the Green Function breaks the charge density down into its spherical harmonic components, calculates the potential due to each component individually (and fairly trivially, just multiplying by a function of the radius of the source charge and the radius at which the potential is desired) and then sums up those components. The same kind of correspondence clearly holds for the induced surface charge density.

Note that the additional $4n + 1$ factor implies the θ dependence of the induced surface charge density is different from that of the original ring charge; *i.e.*, the induced surface charge is not just a ring.

To get some intuition about the surface charge distribution, let's go back to the potential and rewrite it into a method of images solution. Using Equation 3.149 (the expansion of $|\vec{r}-\vec{r}'|^{-1}$ in Legendre polynomials), we can imagine that the first term arises from the convolution of $|\vec{r}-\vec{r}'|^{-1}$ with the ring charge distribution (though we won't prove it explicitly). What about the second term? Let's manipulate it a bit:

$$-r_<^{2n}\,\frac{1}{b}\left(\frac{r_>}{b^2}\right)^{2n}\Bigg|_{r_<=a,r_>=r} = -a^{2n}\,\frac{r^{2n}}{b^{4n+1}} = -\frac{b}{a}\,\frac{r^{2n}a^{2n+1}}{b^{4n+2}} = -\frac{b}{a}\,\frac{r^{2n}}{(b^2/a)^{2n+1}} \tag{3.212}$$

$$= -\frac{b}{a}\,\frac{r_<^{2n}}{r_>^{2n+1}}\Bigg|_{r_<=r,r_>=b^2/a} \tag{3.213}$$

note: meaning of $r_<$ and $r_>$ changed!

Thus, we see the second term has the right form for the potential at $r_< = r$ due to an image charge at radius $r_> = b^2/a$ and normalization $-b/a$ relative to the true charge. (Note that the meaning of $r_<$ and $r_>$ change between the initial and final expression above.) The ring shape comes from the weighted sum over Legendre polynomials, which is the same as the corresponding sum for the potential of the true charge, the first term.

Seeing that the image charge is a ring at radius b^2/a explains the induced surface charge density distribution via its proportionality to the field lines from the true charge to the image charge at the $r = b$ surface. Drawing a picture using the image charge configuration should make this clear.

Example 3.12: Potential inside a grounded spherical conductor with a line charge density along the z axis

This is done in Jackson Section 3.10. We reproduce it here because it has some calculational twists.

The first twist is figuring out how to write down the charge density in spherical coordinates. One could probably rigorously derive the form by writing down the charge density trivially in Cartesian or cylindrical coordinates and then applying Jacobian transformation to convert it to spherical coordinates, but there is an easier, more intuitive way.

It is all present at $\cos\theta = 1$ and $\cos\theta = -1$, so clearly delta functions for these positions need to be included. It has azimuthal symmetry, so there will be no ϕ dependence, only a factor of $1/2\,\pi$. The charge is distributed in radius, so there is some to-be-determined radial dependence $f(r)$. To figure out $f(r)$, let's write down the requirement that the integral be the total charge Q:

$$\rho(\vec{r}) = \frac{Q}{2\,\pi}\, f(r)\, [\delta(\cos\theta - 1) + \delta(\cos\theta + 1)] \tag{3.214}$$

$$\begin{aligned} Q &= \int_{\mathcal{V}} d\tau\, \rho(\vec{r}) \\ &= \frac{Q}{2\,\pi} \int_0^b dr\, r^2\, f(r) \int_{-1}^1 d(\cos\theta)\, [\delta(\cos\theta - 1) + \delta(\cos\theta + 1)] \int_0^{2\,\pi} d\phi \\ &= 2\,Q \int_0^b dr\, r^2\, f(r) \end{aligned}$$

If we choose $f(r) = c/r^2$ where c is a constant to be determined, then the remaining integral becomes trivial and yields b, which we can use to find c:

$$Q = 2\,Q\,c\,b \qquad \Longrightarrow \qquad c = \frac{1}{2b} \tag{3.215}$$

$$\Longrightarrow \qquad \rho(\vec{r}) = \frac{Q}{4\,\pi\,b\,r^2}\,\left[\delta(\cos\theta - 1) + \delta(\cos\theta + 1)\right] \tag{3.216}$$

Now, since the sphere is grounded, we just need to do the integral of the charge density with the Dirichlet Green Function:

$$V(\vec{r}) = \frac{1}{\epsilon_o}\int_{\mathcal{V}} d\tau'\,\rho(\vec{r}\,')\,G_D(\vec{r},\vec{r}\,') \tag{3.217}$$

$$= \frac{Q}{4\,\pi\,\epsilon_o\,b}\sum_{\ell=0}^{\infty}\sum_{m=-\ell}^{\ell}\int_{\mathcal{V}} d\tau'\,\frac{\delta(\cos\theta' - 1) + \delta(\cos\theta' + 1)}{(r')^2}$$

$$\times\; r_<^{\ell}\left[\frac{1}{r_>^{\ell+1}} - \frac{1}{b}\left(\frac{r_>}{b^2}\right)^{\ell}\right]\frac{Y^*_{\ell m}(\theta',\phi')\,Y_{\ell m}(\theta,\phi)}{2\ell+1}$$

We apply azimuthal symmetry as we did in the previous example, selecting the $m = 0$ terms that we can write as Legendre polynomials. The normalization of the spherical harmonics cancels the factor of $2\ell + 1$ in the denominator but adds a factor of $4\,\pi$ in the denominator. The ϕ integral cancels a factor of $2\,\pi$ in the denominator. The θ' integrals can be done trivially, selecting $P_\ell(1)$ and $P_\ell(-1)$. Note also that the $(r')^2$ from the $d\tau'$ cancels the $(r')^2$ in the denominator from the charge density. Thus, we have

$$V(\vec{r}) = \frac{Q}{8\pi\,\epsilon_0\,b}\sum_{\ell=0}^{\infty} P_\ell(\cos\theta)\int_0^b dr'\,r_<^\ell\left[\frac{1}{r_>^{\ell+1}} - \frac{1}{b}\left(\frac{r_>}{b^2}\right)^\ell\right][P_\ell(1) + P_\ell(-1)]$$

We know $P_\ell(1) = 1$ and $P_\ell(-1) = (-1)^\ell$, so the term containing these two factors yields 2 for even ℓ and 0 for odd ℓ. Thus, the above reduces to

$$V(\vec{r}) = \frac{Q}{4\,\pi\,\epsilon_0\,b}\sum_{\ell\ \text{even}} P_\ell(\cos\theta)\int_0^b dr'\,r_<^\ell\left[\frac{1}{r_>^{\ell+1}} - \frac{1}{b}\left(\frac{r_>}{b^2}\right)^\ell\right]$$

The integral over radius must be broken into two pieces, one for $r' < r$ and one for $r' > r$, because the $r_<$ and $r_>$ variables take on different values for these two regions (by definition!). Doing so, and doing the integrals (they are straightforward) yields

$$\int_0^b dr'\, r_<^\ell \left[\frac{1}{r_>^{\ell+1}} - \frac{1}{b}\left(\frac{r_>}{b^2}\right)^\ell\right] = \frac{2\ell+1}{\ell+1}\,\frac{1}{\ell}\left[1-\left(\frac{r}{b}\right)^\ell\right] \tag{3.218}$$

The second portion of the above quantity is well-defined for $\ell \neq 0$, but not for $\ell = 0$. We need to use L'Hôpital's rule to evaluate it for $\ell = 0$:

$$\lim_{\ell\to 0}\frac{1}{\ell}\left[1-\left(\frac{r}{b}\right)^\ell\right] = \lim_{\ell\to 0}\frac{\frac{d}{d\ell}\left[1-\left(\frac{r}{b}\right)^\ell\right]}{\frac{d}{d\ell}\,\ell} = -\lim_{\ell\to 0}\frac{\left(\frac{r}{b}\right)^\ell\left(\ln\frac{r}{b}\right)\frac{d}{d\ell}\,\ell}{\frac{d}{d\ell}\,\ell} = \ln\frac{b}{r} \tag{3.219}$$

Therefore, we may write the full solution as, separating out the $\ell = 0$ term and rewriting in terms of $\ell = 2\,n$,

$$V(\vec{r}) = \frac{Q}{4\,\pi\,\epsilon_o\,b}\left[\ln\frac{b}{r} + \sum_{n=1}^{\infty}\frac{4n+1}{2n\,(2n+1)}\left[1-\left(\frac{r}{b}\right)^{2n}\right]P_{2n}(\cos\theta)\right] \tag{3.220}$$

Let's calculate the induced surface charge density and the total induced charge again:

$$\sigma(\theta) = \epsilon_o \left.\frac{\partial V}{\partial r}\right|_{r=b} = -\frac{Q}{4\,\pi\,b^2}\left[1 + \sum_{n=1}^{\infty} \frac{4n+1}{2n+1} P_{2n}(\cos\theta)\right] \tag{3.221}$$

Note again how the surface charge density has a different n-dependent weighting than the potential. Finally, integrating over the sphere to get the total induced charge, all $n \geq 1$ terms vanish, yielding

$$Q_{ind} = \int_{r=b} b^2\, d\phi\, d\cos\theta\, \sigma(\theta) = -Q \tag{3.222}$$

as we expect from Gauss's Law. It would again be interesting to rewrite the solution in the form of a method of images solution, which you have the tools to do. It clearly should look like a line charge at the north and south poles. Its density will presumably fall off as $1/z^2$ because the true charge density is uniform (in linear units, z) and the image charge magnitude and position both scale as $1/z$.

Lecture 11:

Advanced Electrostatics VII:
Multipoles

Date Revised: 2024/02/13 05:30
Revised lecture break
Date Given: 2024/02/08

Multipole Expansions

Dipoles: Quick Review

Recall from Ph1b the idea of an *electric dipole*: two charges of equal and opposite magnitude $\pm q$ spaced very close together at $\vec{r}_+$ and $\vec{r}_-$. The net charge cancels almost perfectly, so, rather than the potential falling off like $1/r$ at large radius, it falls off as $1/r^2$ with functional form

$$V(\vec{r}) = \frac{1}{4\,\pi\,\epsilon_o}\frac{\vec{p}\cdot\widehat{r}}{r^2} \qquad \text{as} \qquad \frac{r}{|\vec{r}_+|}, \frac{r}{|\vec{r}_-|}, \frac{r}{|\vec{r}_+ - \vec{r}_-|} \to \infty \tag{3.223}$$

where $\vec{p} = q(\vec{r}_+ - \vec{r}_-)$ is the dipole moment.

This idea generalizes. When one has a charge distribution with vanishing net charge, but inside of which there is a variation in the charge density, that variation is still noticeable at large distance as a set of potentials that fall off more quickly than $1/r$. The first additional term is the *dipole*, falling as $1/r^2$, the second is the *quadrupole*, falling as $1/r^3$, the third is the *octupole*, falling as $1/r^4$, and so on. The nomenclature comes from the minimum number of different source charges one must have to obtain that moment: one for monopole, two for dipole, four for quadrupole, etc.

Multipoles: Full Derivation

We derive the full form by considering the potential due to *a charge distribution near the origin as viewed at a point $\vec{r}$ such that r is much larger than the extent of the charge distribution*. This the key assumption! We begin with

$$V(\vec{r}) = \frac{1}{4\,\pi\,\epsilon_o}\int_{\mathcal{V}} d\tau'\,\frac{\rho(\vec{r}\,')}{|\vec{r}-\vec{r}\,'|} \tag{3.224}$$

We now use Equation 3.149, taking $r_< = r'$ and $r_> = r$ because r is outside the charge distribution. Thus,

$$V(\vec{r}) = \frac{1}{4\,\pi\,\epsilon_o}\int_{\mathcal{V}} d\tau'\,\rho(\vec{r}\,')\sum_{\ell=0}^{\infty}\frac{(r')^{\ell}}{r^{\ell+1}}\,P_{\ell}(\cos\gamma) \tag{3.225}$$

where $\cos\gamma = \widehat{r}\cdot\widehat{r}\,'$ is the angle between the two vectors. There is a common $1/r$ we can factor out, leaving

$$\boxed{V(\vec{r}) = \frac{1}{4\,\pi\,\epsilon_o}\frac{1}{r}\sum_{\ell=0}^{\infty}\frac{1}{r^{\ell}}\int_{\mathcal{V}} d\tau'\rho(\vec{r}\,')\,(r')^{\ell}\,P_{\ell}(\cos\gamma)} \tag{3.226}$$

This is the *multipole expansion* of the potential of the charge distribution. One can see that the successive terms fall off as successively higher powers of $1/r$. The angular dependence is given by the Legendre polynomials in $\cos\gamma = \widehat{r}\cdot\widehat{r}\,'$.

The Monopole, Dipole, and Quadrupole Terms

Let's write out the first three terms more explicitly to get some physical intuition:

- *Monopole term*
  The first term is

$$V_1(\vec{r}) = \frac{1}{4\,\pi\,\epsilon_o}\,\frac{1}{r}\int_{\mathcal{V}} d\tau'\rho(\vec{r}') = \frac{1}{4\,\pi\,\epsilon_o}\,\frac{Q}{r} \tag{3.227}$$

  This is the standard Coulomb's Law term due to the total charge. Far enough away, all charge distributions look pointlike. But, if $Q = 0$, this term vanishes identically and the higher-order terms must be considered. Even if $Q \neq 0$, if one is close enough to the charge distribution to see its non-pointlike nature, the higher-order terms will be important corrections to the monopole term.

- *Dipole term*
  The second term is

$$V_2(\vec{r}) = \frac{1}{4\,\pi\,\epsilon_o}\,\frac{1}{r^2}\int_{\mathcal{V}} d\tau'\,\rho(\vec{r}\,')\,r'\cos\gamma = \frac{1}{4\,\pi\,\epsilon_o}\,\frac{1}{r^2}\int_{\mathcal{V}} d\tau'\,\rho(\vec{r}\,')\,r'\,\hat{r}'\cdot\hat{r}$$
$$= \frac{1}{4\,\pi\,\epsilon_o}\,\frac{1}{r^2}\,\hat{r}\cdot\int_{\mathcal{V}} d\tau'\,\rho(\vec{r}\,')\,\vec{r}\,' \qquad (3.228)$$

or
$$\boxed{V_2(\vec{r}) = \frac{1}{4\,\pi\,\epsilon_o}\,\frac{1}{r^2}\,\hat{r}\cdot\vec{p} \qquad \text{where} \qquad \vec{p} = \int_{\mathcal{V}} d\tau'\,\rho(\vec{r}\,')\,\vec{r}\,'} \qquad (3.229)$$

is the dipole moment vector. It is the generalization of $\vec{p} = q\,(\vec{r}_+ - \vec{r}_-)$. It can be written in component form (which is how you would actually calculate it — recall our discussion during the first lecture of how to break vector integrals into a set of scalar integrals!) as

$$\boxed{p_j = \int_{\mathcal{V}} d\tau'\,\rho(\vec{r}\,')\,r_j' = \hat{r}_j\cdot\vec{p}} \qquad (3.230)$$

- *Quadrupole term*
  The third term is

$$\begin{aligned} V_3(\vec{r}) &= \frac{1}{4\,\pi\,\epsilon_o}\,\frac{1}{r^3}\int_{\mathcal{V}} d\tau'\rho(\vec{r}')\,(r')^2\,\frac{1}{2}\left(3\,\cos^2\gamma - 1\right) \\ &= \frac{1}{4\,\pi\,\epsilon_o}\,\frac{1}{r^3}\int_{\mathcal{V}} d\tau'\rho(\vec{r}')\,(r')^2\,\frac{1}{2}\left(3\,\left(\widehat{r}\cdot\widehat{r}'\right)\left(\widehat{r}'\cdot\widehat{r}\right) - 1\right) \\ &= \frac{1}{4\,\pi\,\epsilon_o}\,\frac{1}{r^3}\,\widehat{r}\cdot\left[\int_{\mathcal{V}} d\tau'\rho(\vec{r}')\,(r')^2\frac{1}{2}\left(3\,\widehat{r}'\widehat{r}' - 1\right)\right]\cdot\widehat{r} \end{aligned} \tag{3.231}$$

or
$$\boxed{V_3(\vec{r}) = \frac{1}{4\,\pi\,\epsilon_o}\,\frac{1}{r^3}\,\frac{1}{2}\,\widehat{r}\cdot\underline{\underline{Q}}\cdot\widehat{r} \quad \text{where} \quad \underline{\underline{Q}} = \int_{\mathcal{V}} d\tau'\rho(\vec{r}')\left[3\,\vec{r}'\,\vec{r}' - (r')^2\underline{\underline{1}}\right]} \tag{3.232}$$

is the quadrupole moment and $\underline{\underline{1}} = \text{diag}(1,1,1)$ is the identity tensor with ones along the diagonal. Because it is composed of $\vec{r}'\,\vec{r}'$ and $\underline{\underline{1}}$, $\underline{\underline{Q}}$ is a tensor, implying that one can take a dot product with a vector on each side. Written out in component form (which is again how you would calculate it):

$$\boxed{Q_{jk} = \int_{\mathcal{V}} d\tau'\,\rho(\vec{r}')\,\left[3\,r_j'\,r_k' - (r')^2\delta_{jk}\right] = \widehat{r}_j\cdot\underline{\underline{Q}}\cdot\widehat{r}_k} \tag{3.233}$$

It is now obvious that Q_{jk} is symmetric in its indices.

Origin Dependence of the Dipole Moment

Suppose we take a charge distribution and shift the origin by a vector $\vec{a}$ such that the charge distribution is now centered around $\vec{a}$. Then the new dipole moment is

$$\vec{p}' = \int d\tau' \rho'(\vec{r}')\, \vec{r}' = \int d\tau\, \rho(\vec{r})\, (\vec{a} + \vec{r}) = \vec{a}\, Q + \vec{p} \tag{3.234}$$

where we define the charge distribution in the new coordinate system $\rho'(\vec{r}')$ in terms of the original charge distribution $\rho(\vec{r})$ to be such that $\rho'(\vec{r}') = \rho(\vec{r} = \vec{r}' - \vec{a})$ when $\vec{r}' = \vec{r} + \vec{a}$. Thus, an origin shift can induce an artificial dipole moment for a charge distribution that has a monopole moment. This part of the dipole moment is not real: it is a reflection of the fact that the multipole potentials are written in terms of distance from the origin under the assumption that the charge distribution is centered around the origin. When it is not, this is an unnatural coordinate system to use, requiring corrections of order $1/r^2$ to the standard monopole term ($\propto Q/r$) to handle the fact that the charge distribution is displaced. The above tells us the correction term has the same form as a dipole term. Obviously, one should choose the origin wisely to avoid such complications.

Note also the somewhat counterintuitive implication that, if $Q = 0$, then the dipole moment is independent of origin! Our assumption that the charge distribution, including its displacement from the origin, is small compared to the to the observation point implies that a must also be small so that any corrections, which are of order a^3/r^3, are small compared to the leading $1/r^2$ dependence for the dipole term.

Field of an Electric Dipole

This is simply a matter of taking the gradient. If we let $\vec{p} = p\,\hat{z}$, then this is easy:

$$V_2(\vec{r}) = \frac{p\,\cos\theta}{4\,\pi\,\epsilon_o\,r^2} \tag{3.235}$$

$$\Longrightarrow \quad E_r(\vec{r}) = -\frac{\partial V_2}{\partial r} = -\frac{2\,p\,\cos\theta}{4\,\pi\,\epsilon_o\,r^3} \tag{3.236}$$

$$E_\theta(\vec{r}) = -\frac{1}{r}\,\frac{\partial V_2}{\partial \theta} = \frac{p\,\sin\theta}{4\,\pi\,\epsilon_o\,r^3} \tag{3.237}$$

$$E_\phi(\vec{r}) = -\frac{1}{r\,\sin\theta}\,\frac{\partial V_2}{\partial \phi} = 0 \tag{3.238}$$

$$\text{or} \quad \vec{E}(\vec{r}) = \frac{p}{4\,\pi\,\epsilon_o\,r^3}\left(2\,\hat{r}\,\cos\theta + \hat{\theta}\,\sin\theta\right) \tag{3.239}$$

To generalize this result for an arbitrary orientation of $\vec{p}$ requires some vector algebra. We have Equation 3.229 for the dipole potential in generic form, which we write out as

$$V_2(\vec{r}) = \frac{1}{4\,\pi\,\epsilon_o}\,\frac{1}{r^3}\,\vec{r}\cdot\vec{p} = \frac{1}{4\,\pi\,\epsilon_o}\,\frac{1}{r^3}\,\sum_i r_i\,p_i \tag{3.240}$$

Now, we take the gradient, first noting

$$\frac{\partial}{\partial r_j}\,\frac{r_i}{r^3} = \frac{\partial}{\partial r_j}\,\frac{r_i}{(r^2)^{3/2}} = -\frac{3}{2}\,\frac{r_i}{(r^2)^{5/2}}\,\frac{\partial r^2}{\partial r_j} + \frac{\delta_{ij}}{r^3} = -\frac{3}{2}\,\frac{r_i}{r^5}\,(2\,r_j) + \frac{\delta_{ij}}{r^3} \tag{3.241}$$

Where we used $r^3 = \left(r^2\right)^{3/2}$ and $r^2 = \sum_k r_k^2$ to more easily calculate the partial derivative. Therefore, with r_j and $\widehat{r}_j$ being the jth Cartesian coordinate and unit vector,

$$\begin{aligned}
\vec{E}_2(\vec{r}) = -\vec{\nabla}V_2(\vec{r}) &= -\sum_j \widehat{r}_j\,\frac{\partial V_2}{\partial r_j} = \frac{1}{4\,\pi\,\epsilon_o\,r^5}\sum_{ij}\widehat{r}_j\,p_i\left[3\,r_i\,r_j - \delta_{ij}\,r^2\right] \\
&= \frac{1}{4\,\pi\,\epsilon_o\,r^5}\sum_j\left[r_j\,\widehat{r}_j\left(3\sum_i p_i\,r_i\right) - p_j\,\widehat{r}_j\,r^2\right] = \frac{1}{4\,\pi\,\epsilon_o\,r^5}\left[3\,(\vec{p}\cdot\vec{r})\,\vec{r} - r^2\vec{p}\right] \\
&\Longrightarrow \qquad \boxed{\vec{E}_2(\vec{r}) = \frac{1}{4\,\pi\,\epsilon_o\,r^3}\left[3\,(\vec{p}\cdot\widehat{r})\,\widehat{r} - \vec{p}\right]} \qquad (3.242)
\end{aligned}$$

Electrostatic Potential Energy of a Multipole Distribution in an External Potential

The general expression for the potential energy of a charge distribution in an external potential is

$$U = \int_{\mathcal{V}} \rho(\vec{r}\,') \, V(\vec{r}\,') \tag{3.243}$$

Under the assumption that $V(\vec{r}\,')$ varies slowly (but need not be constant!) over the spatial extent of the charge distribution, we can rewrite this in terms of moments of the charge distribution and derivatives of the potential. To do so, we need to expand $V(\vec{r})$ about some point in the distribution. To make the calculation easier, assume the charge distribution is centered on the origin, around which we will expand. (We will generalize later.) We use the multidimensional Taylor expansion of $V(\vec{r})$:

$$V(\vec{r}\,') = V(\vec{r}\,' = \vec{0}) + \sum_{j=1}^{3} r_j' \left.\frac{\partial V}{\partial r_j}\right|_{\vec{r}\,'=\vec{0}} + \frac{1}{2}\sum_{j,k=1} r_j' r_k' \left.\frac{\partial^2 V}{\partial r_j \, \partial r_k}\right|_{\vec{r}\,'=\vec{0}} + \cdots \tag{3.244}$$

We can already foresee how integrating the above form for $V(\vec{r}\,')$ with $\rho(\vec{r}\,')$ is going to result in a dipole moment in the first term and quadrupole moment in the second.

Using $E_j = -\frac{\partial V}{\partial r_j}$, we may simplify

$$V(\vec{r}\,') = V(\vec{0}) - \vec{r}\,' \cdot \vec{E}(\vec{0}) + \frac{1}{6}\sum_{j,k=1}^{3} 3\, r_j{}' r_k' \left.\frac{\partial^2 V}{\partial r_j\, \partial r_k}\right|_{\vec{0}} + \cdots \tag{3.245}$$

$$= V(\vec{0}) - \vec{r}\,' \cdot \vec{E}(\vec{0}) + \frac{1}{6}\sum_{j,k=1}^{3} \left(3\, r_j{}' r_k' - (r\,')^2 \delta_{jk}\right) \left.\frac{\partial^2 V}{\partial r_j\, \partial r_k}\right|_{\vec{0}} + \cdots \tag{3.246}$$

where we were able to add the $(r\,')^2\delta_{jk}$ term because

$$\sum_{j,k} (r\,')^2 \delta_{jk} \left.\frac{\partial^2 V}{\partial r_j\, \partial r_k}\right|_{\vec{0}} = (r\,')^2 \nabla^2 V(\vec{r}\,' = 0) = 0 \tag{3.247}$$

because the charge distribution *sourcing* V is not present near the origin. Remember, $\rho(\vec{r})$ is not the distribution sourcing V; V is provided to us and is due to some charge distribution far away from the origin.

With the above expansion, the electrostatic potential energy is now (note that $\vec{E}(\vec{0})$ and $\partial^2 V/\partial r_j \partial r_k\big|_{\vec{0}}$ are constant with respect to r', so they come outside of the r' integral)

$$U = V(\vec{0}) \int_{\mathcal{V}} d\tau' \rho(\vec{r}') - \vec{E}(\vec{0}) \cdot \int_{\mathcal{V}} d\tau' \, \rho(\vec{r}') \, \vec{r}' \tag{3.248}$$

$$+ \frac{1}{6} \sum_{j,k=1}^{3} \frac{\partial^2 V}{\partial r_j \, \partial r_k}\bigg|_{\vec{0}} \int_{\mathcal{V}} d\tau' \rho(\vec{r}') \left[3 \, r_j' \, r_k' - \delta_{jk} (r')^2 \right] + \cdots$$

$$= Q \, V(\vec{0}) - \vec{p} \cdot \vec{E}(\vec{0}) + \frac{1}{6} \sum_{j,k=1}^{3} Q_{jk} \, \frac{\partial^2 V}{\partial r_j \, \partial r_k}\bigg|_{\vec{0}} + \cdots \tag{3.249}$$

We generalize for a charge distribution centered around $\vec{r}$ instead of the origin:

$$U(\vec{r}) = Q \, V(\vec{r}) - \vec{p} \cdot \vec{E}(\vec{r}) + \frac{1}{6} \sum_{j,k=1}^{3} Q_{jk} \, \frac{\partial^2 V}{\partial r_j \, \partial r_k}\bigg|_{\vec{r}} + \cdots \tag{3.250}$$

$$= Q \, V(\vec{r}) - \vec{p} \cdot \vec{E}(\vec{r}) + \frac{1}{6} \, \vec{\nabla}_{\vec{r}} \cdot \underline{\underline{Q}} \cdot \vec{\nabla}_{\vec{r}} \, V(\vec{r}) + \cdots \tag{3.251}$$

where we have written the last term in tensor dot product form. There are now contributions to the potential energy from the relative alignment of $\vec{p}$ and $\vec{E}$ and from the orientation of $\underline{\underline{Q}}$'s principal axes relative to the principal axes of the potential's curvature matrix. Note that $\vec{\nabla}_{\vec{r}}$ acts on the spatial dependence of $V(\vec{r})$; $\vec{r}'$ has already been integrated over to obtain $\underline{\underline{Q}}$.

Force on a Multipole Distribution in an External Field

We can calculate the force on the charge distribution by taking the derivative of U with respect to the *charge distribution's* nominal position $\vec{r}$, now replacing one derivative of V with the electric field $\vec{E}$ in the quadrupole term:

$$\begin{aligned}\vec{F}(\vec{r}) &= -\vec{\nabla}U(\vec{r}) = Q\left(-\vec{\nabla}V(\vec{r})\right) + \vec{\nabla}\left(\vec{p}\cdot\vec{E}(\vec{r})\right) + \frac{1}{6}\sum_{j,k,m=1}^{3}\widehat{r}_m\, Q_{jk}\,\frac{\partial^2 E_j}{\partial r_m\,\partial r_k} + \cdots \\ &= Q\,\vec{E}(\vec{r}) + \left(\vec{p}\cdot\vec{\nabla}\right)\vec{E}(\vec{r}) + \frac{1}{6}\sum_{j,k,m=1}^{3}\widehat{r}_m\, Q_{jk}\,\frac{\partial^2 E_j}{\partial r_m\,\partial r_k} + \cdots \\ &= Q\,\vec{E}(\vec{r}) + \left(\vec{p}\cdot\vec{\nabla}\right)\vec{E}(\vec{r}) + \frac{1}{6}\,\vec{\nabla}\left[\vec{\nabla}\cdot\left(\underline{\underline{Q}}\cdot\vec{E}(\vec{r})\right)\right] + \cdots \end{aligned} \tag{3.252}$$

In going from the first to the second row, we used the vector identity $\vec{\nabla}\left(\vec{a}\cdot\vec{f}(\vec{r})\right) = \left(\vec{a}\cdot\vec{\nabla}\right)\vec{f}(\vec{r})$ when $\vec{a}$ is a constant vector and $\vec{f}(\vec{r})$ has no curl. Note that all $\vec{\nabla}$ are with respect to $\vec{r}$ (since $\vec{r}\,'$ has been integrated over already).

We see that the total force is a sum of contributions from the interaction of the monopole with the electric field, the dipole with gradients in the electric field and, the quadrupole with the local curvature (second derivatives) of the electric field.

Lecture 12:

Advanced Electrostatics VIII:
Multipoles (cont.)

Electrostatics in Matter I:
Polarizability and Polarization
Bound Charges and their Potential
Electric Displacement Field
Linear Dielectrics

Date Revised: 2024/02/15 06:15
Revised lecture break
Date Given: 2024/02/13

Torque on a Multipole in an External Field

Let's also calculate the torque. As you know from Ph106a, to calculate a torque, we need to take the gradient of the potential energy in spherical coordinates with respect to the orientation of the charge distribution relative to the electric field.

The monopole term yields no torque because there is no orientation angle involved: Q and $V(\vec{r})$ are scalars.

Considering the dipole term, we understand that there are only two vectors involved, $\vec{p}$ and $\vec{E}$, and the potential energy only depends on the angle between them. So the torque will be given by the derivative with respect to this angle, which we call θ_p to differentiate it from the θ coordinate of the system in which we consider $\vec{E}$. This angle will be measured from $\vec{E}$ to $\vec{p}$. Then,

$$\vec{N}_{elec} = -\frac{\partial}{\partial \theta_p}\left(-\vec{p}\cdot\vec{E}(\vec{r})\right) \tag{3.253}$$

$$= \frac{\partial}{\partial \theta_p} p\left|\vec{E}(\vec{r})\right| \cos\theta_p = -p\left|\vec{E}(\vec{r})\right| \sin\theta_p$$

$$= \vec{p}\times\vec{E}(\vec{r}) \tag{3.254}$$

This is a result you are familiar with from Ph1b, indicating the torque acts in a direction to align the dipole moment with the field direction.

Moving on to the quadrupole term, we recognize from Ph106a that any symmetric tensor can be diagonalized via a rotation. Let's write

$$\underline{\underline{Q}} = \mathcal{R}(\phi_Q, \theta_Q, \psi_Q)\, \underline{\underline{\mathcal{Q}}}\, [\mathcal{R}(\phi_Q, \theta_Q, \psi_Q)]^T \qquad \text{with} \quad \underline{\underline{\mathcal{Q}}} = \text{diag}(Q_1, Q_2, Q_3) \quad (3.255)$$

where the Q_i are quadrupole moments along the principal axes of the quadrupole tensor and $\mathcal{R}(\phi_Q, \theta_Q, \psi_Q)$ is the rotation matrix that rotates from the frame in which the coordinate axes align with the quadrupole tensor's principal axes to the arbitrary frame we started in, with the three Euler angles $(\phi_Q, \theta_Q, \psi_Q)$ defining the orientation of the principal axes of $\underline{\underline{\mathcal{Q}}}$ relative to the this arbitrary frame. This kind of diagonalization should be familiar to you from Ph106a, with $\mathcal{R}$ rotating from the "body" frame (the one fixed to the charge distribution's quadrupole principal axes) to the "space" frame. The quadrupole potential energy term is then

$$U_3 = -\frac{1}{6}\, \vec{\nabla}_{\vec{r}} \cdot \left\{ \mathcal{R}(\phi_Q, \theta_Q, \psi_Q)\, \underline{\underline{\mathcal{Q}}}\, [\mathcal{R}(\phi_Q, \theta_Q, \psi_Q)]^T \right\} \cdot \vec{E}(\vec{r}) \qquad (3.256)$$

To calculate the torque, we need to take the gradient of U_3 with respect to the orientation of the quadrupole. This amounts to taking gradients of $\mathcal{R}$ and $\mathcal{R}^T$ with respect to this orientation. As you know from the case of the symmetric top, the Euler angles are particularly useful angles with respect to which these derivatives can be taken. $\partial/\partial\phi_Q$ gives the torque about the z-axis of the space frame, which causes precession around that axis. $\partial/\partial\theta_Q$ gives the torque that causes motion in the polar angle direction with respect to the same space-frame z-axis, which is like nutation in the case of a top. And $\partial/\partial\psi_Q$ calculates the torque about one particular principal axis of the quadrupole, chosen at will, which accelerates or decelerates the rotation about that axis. You are familiar with symmetric tops, with $I_1 = I_2$. Here, we can have symmetric quadrupoles, with $Q_1 = Q_2$. In this case, the ψ_Q angle is the angle about the 3 axis of the quadrupole (the principal axis that aligns with the z-axis in the body frame). We do not take this further because, as you know from the study of tops in Ph106a, the phenomenology can be quite rich.

Section 4
Electrostatics in Matter

Polarizability and Polarization

Review of Polarizability of Materials

Griffiths §4.1 does a good job of providing physical motivation for the study of the polarizability of materials, and also reviews material you saw in Ph1b, so we only summarize the basics here.

- Atoms and molecules are *polarizable*, meaning that they can acquire a dipole moment when an external electric field is applied because of the separation of the positive and negative charge in response to the applied field. The charge distribution that results is such that its field is in the opposite direction as the applied field at the location of the atom or molecule.
- We assume that this polarizability is *a linear process*, so that the induced dipole moment is linear in the applied electric field, though the response may be anisotropic. The *polarizability tensor* $\underline{\underline{\alpha}}$ relates the induced dipole moment to the applied field:

$$\vec{p} = \underline{\underline{\alpha}} \cdot \vec{E} \tag{4.1}$$

- As we showed in our discussion of multipoles, dipoles can experience torques and forces in an electric field. If a dipole is placed in an electric field, it feels a torque (Equation 3.254)

$$\vec{N} = \vec{p} \times \vec{E} \tag{4.2}$$

If the electric field is nonuniform, the dipole feels a force (Equation 3.252)

$$\vec{F} = \left(\vec{p} \cdot \vec{\nabla}\right) \vec{E} \tag{4.3}$$

- If a medium consists of polarizable atoms or molecules, then that medium can become polarized under the application of an electric field. The *polarization* (or *polarization density*) of the medium is

$$\vec{P} = n\,\vec{p} \tag{4.4}$$

where n is the density of polarizable atoms or molecules and $\vec{p}$ is the induced dipole per atom or molecule.

Bound Charges and the Potential of a Polarizable Material

When a medium is polarized and acquires a polarization vector $\vec{P}$, then it can generate its own electric field. This comes from the superposition of the dipole fields of the individual polarized atoms or molecules. In Ph1b, you saw how the polarization could be interpreted as yielding *bound charge* densities: when the medium polarizes, the positive components of some dipoles are cancelled by the negative components of nearby dipoles, but there can appear a net effective charge: on the boundaries, where the cancellation fails, and in the bulk if the dipole density is not uniform, also causing the cancellation to fail. This argument was made in Purcell in Ph1b to derive the bound charge densities, and Griffiths makes it in §4.2.2. Here we derive the relationship between the polarization vector and the bound charge density in rigorous fashion.

The total electric potential generated by a polarizable medium is found by summing up the dipole potentials of the individual dipoles:

$$V(\vec{r}) = \frac{1}{4\,\pi\,\epsilon_o}\int_{\mathcal{V}} d\tau'\,\frac{\vec{P}(\vec{r}\,')\cdot(\vec{r}-\vec{r}\,')}{|\vec{r}-\vec{r}\,'|^3} \tag{4.5}$$

We use the identity $(\vec{r}-\vec{r}\,')/|\vec{r}-\vec{r}\,'|^3 = \vec{\nabla}_{\vec{r}\,'}(1/|\vec{r}-\vec{r}\,'|)$ (note: no minus sign because this is $\vec{\nabla}_{\vec{r}\,'}$, not $\vec{\nabla}_{\vec{r}}$, and we have $\vec{r}-\vec{r}\,'$ in the numerator, not $\vec{r}\,'-\vec{r}$) to rewrite this as

$$V(\vec{r}) = \frac{1}{4\,\pi\,\epsilon_o}\int_{\mathcal{V}} d\tau'\,\vec{P}(\vec{r}\,')\cdot\vec{\nabla}_{\vec{r}\,'}\left(\frac{1}{|\vec{r}-\vec{r}\,'|}\right) \tag{4.6}$$

We can integrate by parts to obtain

$$V(\vec{r}) = \frac{1}{4\,\pi\,\epsilon_o}\left[\int_{\mathcal{V}} d\tau' \vec{\nabla}_{\vec{r}'} \cdot \left(\frac{\vec{P}(\vec{r}')}{|\vec{r}-\vec{r}'|}\right) - \int_{\mathcal{V}} d\tau' \, \frac{1}{|\vec{r}-\vec{r}'|}\left(\vec{\nabla}_{\vec{r}'} \cdot \vec{P}(\vec{r}')\right)\right] \tag{4.7}$$

The first term can be converted to a surface integral via the divergence theorem:

$$V(\vec{r}) = \frac{1}{4\,\pi\,\epsilon_o}\left[\int_{\mathcal{S}(\mathcal{V})} da' \, \frac{\widehat{n}(\vec{r}') \cdot \vec{P}(\vec{r}')}{|\vec{r}-\vec{r}'|} - \int_{\mathcal{V}} d\tau' \, \frac{1}{|\vec{r}-\vec{r}'|}\left(\vec{\nabla}_{\vec{r}'} \cdot \vec{P}(\vec{r}')\right)\right] \tag{4.8}$$

We thus see that the potential appears to be that of a surface charge density $\sigma_b(\vec{r}')$ on $\mathcal{S}(\mathcal{V})$ and a volume charge density $\rho_b(\vec{r}')$ in $\mathcal{V}$ with ($\widehat{n}$ is the outward normal from the polarizable material):

$$\sigma_b(\vec{r}') = \widehat{n}(\vec{r}') \cdot \vec{P}(\vec{r}') \qquad \rho_b(\vec{r}') = -\vec{\nabla}_{\vec{r}'} \cdot \vec{P}(\vec{r}') \tag{4.9}$$

$$V(\vec{r}) = \frac{1}{4\,\pi\,\epsilon_o}\left[\int_{\mathcal{S}(\mathcal{V})} da' \, \frac{\sigma_b(\vec{r}')}{|\vec{r}-\vec{r}'|} + \int_{\mathcal{V}} d\tau' \, \frac{\rho_b(\vec{r}')}{|\vec{r}-\vec{r}'|}\right] \tag{4.10}$$

These charges are called "bound charges" because they are bound to the polarizable medium.

Example 4.1: Potential and Field of a Uniformly Polarized Sphere

This problem from Ph1b is much easier to solve with our knowledge of solutions to Laplace's Equation than it was without such techniques. The polarization density is a constant $\vec{P} = P\,\hat{z}$. The bound volume charge density vanishes because $\vec{P}$ is constant. The bound surface charge density on the surface at radius R is

$$\sigma_b = \hat{n}(\vec{r}) \cdot \vec{P} = \hat{r} \cdot P\,\hat{z} = P\,\cos\theta \tag{4.11}$$

Therefore, Laplace's Equation holds everywhere except at $r = R$, so we apply the same techniques as we developed previously. Recall that we discussed this same problem in Example 3.6 (Griffiths Example 3.9) for a generic $\sigma(\theta)$. The generic solution was

$$V(r < R, \theta) = \sum_{\ell=0}^{\infty} A_\ell\, r^\ell\, P_\ell(\cos\theta) \qquad V(r > R, \theta) = \sum_{\ell=0}^{\infty} \frac{B_\ell}{r^{\ell+1}}\, P_\ell(\cos\theta) \tag{4.12}$$

$$\text{with} \quad A_\ell = \frac{1}{2\,\epsilon_o\, R^{\ell-1}} \int_0^\pi d\theta'\, \sin\theta'\, \sigma(\theta')\, P_\ell(\cos\theta') \qquad B_\ell = A_\ell\, R^{2\,\ell+1} \tag{4.13}$$

Since $\sigma(\theta) = P\cos\theta = P\,P_1(\cos\theta)$, the orthonormal functions do their job and we get (making sure to include the normalization factor $2/(2\,\ell+1) = 2/3$):

$$V(r < R, \theta) = \frac{P\,r\,\cos\theta}{3\,\epsilon_o} \qquad V(r > R, \theta) = \frac{P\,R^3\cos\theta}{3\,\epsilon_o\,r^2} \tag{4.14}$$

We can write these more simply. We recognize $z = r\cos\theta$ and that the total dipole moment of the sphere is $\vec{p} = 4\pi R^3 P\,\hat{z}/3$, yielding

$$V(r<R,\theta) = \frac{P\,z}{3\,\epsilon_o} \qquad V(r>R,\theta) = \frac{\vec{p}\cdot\hat{r}}{4\pi\,\epsilon_o\,r^2} \tag{4.15}$$

Thus, the field inside the sphere is *uniform*, $\vec{E} = -\vec{P}/3\,\epsilon_o$, and the field outside the sphere is that of a dipole $\vec{p}$. Note that the field outside the sphere is a perfect dipole field all the way to $r = R$; this is not an approximation (until you get so close to the surface that you can see the discretization of the dipoles).

We remind the reader of the Ph1b technique, where we obtained this same result by treating the sphere as two spheres of uniform charge density $\rho = q/(4\pi R^3/3)$ with their centers displaced by $\vec{d} = \vec{p}/q$. The field inside a uniform sphere of charge is proportional to the radial vector outward from its center, so the two vectors $\vec{r} - \vec{d}/2$ and $\vec{r} + \vec{d}/2$ end up differencing (because the two spheres have opposite charge) to yield $\vec{d}$, yielding the uniform internal field. Outside the spheres, they look like point charges, so the system looks like a point dipole $\vec{p}$.

One could also use this argument to figure out that the charge density on the surface is $\sigma = P\cos\theta$ and evaluate the potential and field of that charge distribution.

The Electric Displacement Field

Definition of the Electric Displacement Field

We proved earlier that the potential due to a polarization density $\vec{P}(\vec{r})$ is

$$V(\vec{r}) = \frac{1}{4\,\pi\,\epsilon_o}\left[\int_{\mathcal{S}(\mathcal{V})} da'\,\frac{\widehat{n}(\vec{r}\,')\cdot\vec{P}(\vec{r}\,')}{|\vec{r}-\vec{r}\,'|} + \int_{\mathcal{V}} d\tau'\,\frac{-\vec{\nabla}_{\vec{r}\,'}\cdot\vec{P}(\vec{r}\,')}{|\vec{r}-\vec{r}\,'|}\right] \tag{4.16}$$

These are analogues of Coulomb's law for ρ_b, so the potential and field due to the polarization density satisfy

$$\nabla^2 V_b = -\frac{1}{\epsilon_o}\,\rho_b \qquad \vec{\nabla}\cdot\vec{E}_b = \frac{1}{\epsilon_o}\,\rho_b = -\frac{1}{\epsilon_o}\,\vec{\nabla}\cdot\vec{P} \tag{4.17}$$

If there is a *free* charge density ρ_f (which will contribute to V and $\vec{E}$!), then we see that the total potential and field satisfy

$$\nabla^2 V = -\frac{1}{\epsilon_o}\,(\rho_f + \rho_b) \qquad \vec{\nabla}\cdot\vec{E} = \frac{1}{\epsilon_o}\left(\rho_f - \vec{\nabla}\cdot\vec{P}\right) \tag{4.18}$$

We will see later that it will be convenient to have a field that depends primarily on the free charge density. Thus, we define the *electric displacement field* by

$$\boxed{\vec{D} = \epsilon_o \vec{E} + \vec{P}} \tag{4.19}$$

We immediately see that Gauss's Law can be written as

$$\boxed{\vec{\nabla} \cdot \vec{D} = \rho_f \qquad \Longleftrightarrow \qquad \oint_{\mathcal{S}} da\, \hat{n} \cdot \vec{D} = Q_{free,encl}} \tag{4.20}$$

The Helmholtz Theorem tells us that any vector field can be written as the sum of a curl-free component (sourced by the divergence of the field) and a divergence-free component (sourced by the curl of the field). Thus, to fully understand $\vec{D}$, we also need to determine its curl:

$$\boxed{\vec{\nabla} \times \vec{D} = \epsilon_o \vec{\nabla} \times \vec{E} + \vec{\nabla} \times \vec{P} = \vec{\nabla} \times \vec{P}} \tag{4.21}$$

Because the right side may not vanish, the left side may not vanish. This possibly nonzero curl is an important distinction between $\vec{D}$ and $\vec{E}$.

While Gauss's Law does indeed hold for $\vec{D}$, the possibility that $\vec{\nabla} \times \vec{D} \neq 0$ implies that the standard symmetry assumptions we make to apply Gauss's Law to find the field may not apply.

However, if one knows that, due to symmetry or some other consideration, $\vec{\nabla} \times \vec{P} = 0$, then *one can apply the standard techniques for using Gauss's Law combined with symmetry to calculate the displacement field.* ($\vec{\nabla} \times \vec{P} = 0$ should be interpreted as also requiring that any boundaries be normal to $\vec{P}$ because we will see below that, unlike for $\vec{E}$, the tangential component of $\vec{D}$ is not continuous if $\vec{P}$ has a tangential component.)

When the above is true, $\vec{D}$ provides a calculational convenience: if a free charge density ρ_f and a polarization field $\vec{P}$ are specified, then we should calculate $\vec{D}$ from the free charge density using Gauss's Law and then obtain the electric field from $\vec{E} = (\vec{D} - \vec{P})/\epsilon_o$. This simplification is possible only because of the particular form of the bound charge density, $\rho_b = -\vec{\nabla} \cdot \vec{P}$, which parallels the mathematical form of Gauss's Law, along with the condition $\vec{\nabla} \times \vec{P} = 0$.

Note the extra condition $\vec{\nabla} \times \vec{P} = 0$; this reflects the fact that $\vec{P}$ has more degrees of freedom than a scalar field ρ_b, so those extra degrees of freedom need to be specified via the curl-free condition for $\vec{D}$ to be derivable from ρ_f alone.

The situation will simplify somewhat when we consider linear, uniform dielectrics where $\vec{P} \propto \vec{E}$; then $\vec{\nabla} \times \vec{P} = 0$ is guaranteed, though the requirement that $\vec{P}$ be normal to any boundaries may still create complications.

Boundary Conditions on the Displacement Field

We derived boundary conditions on $\vec{E}$ earlier, Equations 2.59 and 2.61:

$$\hat{n} \cdot \left(\vec{E}_2 - \vec{E}_1 \right) = \frac{1}{\epsilon_o} \sigma \qquad \hat{s} \cdot \left(\vec{E}_2 - \vec{E}_1 \right) = 0 \tag{4.22}$$

where $\hat{n}$ is the normal vector pointing from region 1 into region 2 and $\hat{s}$ is any tangential vector (*i.e.*, $\hat{s} \cdot \hat{n} = 0$). We derived the equation for the normal component using the divergence of $\vec{E}$. So, here, we can use the fact that $\vec{\nabla} \cdot \vec{D} = \rho_f$, which yields

$$\boxed{\hat{n} \cdot \left(\vec{D}_2 - \vec{D}_1 \right) = \sigma_f} \tag{4.23}$$

Note that, by definition, we have $\sigma_b = \hat{n} \cdot \vec{P}$ where $\hat{n}$ is the outward normal going from a region with a polarization density to vacuum. Therefore, by superposition,

$$\boxed{\hat{n} \cdot \left(\vec{P}_2 - \vec{P}_1 \right) = -\sigma_b} \tag{4.24}$$

We could also have used $\rho_b = -\vec{\nabla} \cdot \vec{P}$ and followed the same type of derivation as used for $\vec{E}$ and $\vec{D}$. The sign on the right side of the boundary condition enters because of the sign in $\vec{\nabla} \cdot \vec{P} = -\rho_b$.

In general, we know nothing about $\vec{\nabla} \times \vec{P}$, so the boundary condition on the tangential component of $\vec{D}$ just reflects the fact that its curl is the curl of the polarization field. We obtain this condition by inserting the relation between $\vec{E}$, $\vec{D}$, and $\vec{P}$ into the above tangential condition:

$$\boxed{\widehat{s} \cdot \left(\vec{D}_2 - \vec{D}_1 \right) = \widehat{s} \cdot \left(\vec{P}_2 - \vec{P}_1 \right)} \tag{4.25}$$

Note that, even in the case of linear dielectrics, the right side can be nonzero, as we will see below.

What Sources $\vec{D}$? When Does It Vanish?

Consider the uniformly polarized sphere we just discussed in Example 4.1. In that example, the displacement field is

$$\vec{D}(r < R) = \epsilon_o \vec{E} + \vec{P} = -\frac{\vec{P}}{3} + \vec{P} = \frac{2}{3}\vec{P} \tag{4.26}$$

$$\vec{D}(r > R) = \epsilon_o \vec{E} = \epsilon_o \left(\text{field of an electric dipole } \vec{p} = \frac{4\pi}{3} R^3 \vec{P} \right) \tag{4.27}$$

There is no free charge in the problem, yet we have a nonzero $\vec{D}$! The nonzero nature of $\vec{\nabla} \times \vec{D}$ implies that $\vec{D}$ has another sourcing term that is not captured by Gauss's Law for $\vec{D}$. In this case, this sourcing term manifests as a discontinuity of the tangential component of $\vec{P}$ at $r = R$: $\vec{P} \cdot \widehat{\theta} = P\,\widehat{z} \cdot \widehat{\theta} = -P \sin\theta$. This nonzero value of $\widehat{s} \cdot \vec{P}$ is what makes $\widehat{s} \cdot \vec{D} \neq 0$ and thus we cannot apply the usual symmetry arguments to use Gauss's Law for $\vec{D}$ (which would otherwise tell us $\vec{D} = 0$ because $\rho_f = 0$ everywhere). For $\vec{D}$ to vanish completely, one not only needs the free charge to vanish but there must also be no nontrivial boundary conditions on $\vec{D}$.

Linear Dielectrics

So far, we have considered situations where $\vec{P}$ has been specified for us. But, it is usually caused by an external field, and so what we really want to do is figure out what observed potential and field arise by summing the externally applied potential/field and that due to the polarization of the dielectric in response to that external potential/field. For most substances, at least at low fields, the relation between the two is linear and there is a simple scalar constant of proportionality:

$$\vec{P} = \epsilon_o \, \chi_e \, \vec{E} \tag{4.28}$$

where χ_e is the *electric susceptibility*. Such materials are called *linear dielectrics*. An immediate implication of the above is:

$$\vec{D} = \epsilon_o \, \vec{E} + \vec{P} = \epsilon_o \, (1 + \chi_e) \, \vec{E} \equiv \epsilon \, \vec{E} \tag{4.29}$$

where $\epsilon \equiv \epsilon_o \, (1 + \chi_e)$ is the *permittivity* of the material and $\epsilon_r \equiv 1 + \chi_e$ is the *relative permittivity* or *dielectric constant* of the material.

A very important point is that $\vec{E}$ above is the *total* field, not just the externally applied field. You can think of polarization as an iterative process: an applied field $\vec{E}_0$ causes polarization $\vec{P}_0$, which creates its own field $\vec{E}_1$, which the polarization responds to by adding a contribution $\vec{P}_1$, which creates its own field $\vec{E}_2$, and so on. The process converges to the final total electric field $\vec{E}$ and polarization $\vec{P}$.

Example 4.2: Conducting sphere with dielectric shell around it

Consider a conducting sphere of radius a with (free) charge Q on it surrounded by a (thick) shell of dielectric ϵ with inner and outer radii a and b. Because the system is spherically symmetric and contains a linear dielectric, we know that $\vec{E}$, $\vec{D}$, and $\vec{P}$ all have the form

$$\vec{E} = E(r)\,\hat{r} \qquad \vec{D} = D(r)\,\hat{r} \qquad \vec{P} = P(r)\,\hat{r} \tag{4.30}$$

This ensures that the curl of all three vanish and that, at the boundaries, we have no tangential components of $\vec{D}$ and $\vec{P}$. *We have now satisfied all the conditions required for us to be able to derive $\vec{D}$ directly from the free charge by Gauss's Law*, which yields

$$\vec{D}(\vec{r}) = \frac{Q}{4\,\pi\,r^2}\,\hat{r} \qquad r > a \tag{4.31}$$

($\vec{D} = \vec{E} = \vec{P} = 0$ for $r < a$.) Then we just apply the relation between $\vec{D}$ and $\vec{E}$:

$$\vec{E}(\vec{r}) = \frac{Q}{4\,\pi\,\epsilon(r)\,r^2}\,\hat{r} = \begin{cases} \left(Q/4\,\pi\,\epsilon\,r^2\right)\hat{r} & a < r < b \\ \left(Q/4\,\pi\,\epsilon_o\,r^2\right)\hat{r} & b < r \end{cases} \tag{4.32}$$

The electric field is *screened* (reduced) inside the dielectric and unchanged outside.

Let's calculate the polarization vector and bound charge density:

$$\vec{P}(\vec{r}) = \epsilon_o \, \chi_e(r) \, \vec{E}(\vec{r}) = (\epsilon(r) - \epsilon_o) \, \vec{E}(\vec{r}) = \frac{\epsilon(r) - \epsilon_o}{\epsilon(r)} \, \frac{Q}{4\,\pi\,r^2} \, \widehat{r}$$

$$= \begin{cases} \frac{\epsilon - \epsilon_o}{\epsilon} \, \frac{Q}{4\,\pi\,r^2} \, \widehat{r} & a < r < b \\ 0 & b < r \end{cases} \tag{4.33}$$

$$\rho_b = -\vec{\nabla} \cdot \vec{P} = 0 \tag{4.34}$$

$$\sigma_b = \begin{cases} -\widehat{r} \cdot \vec{P}(r = a) = -\frac{\epsilon - \epsilon_o}{\epsilon} \, \frac{Q}{4\,\pi\,a^2} & r = a \\ \widehat{r} \cdot \vec{P}(r = b) = \frac{\epsilon - \epsilon_o}{\epsilon} \, \frac{Q}{4\,\pi\,b^2} & r = b \end{cases} \tag{4.35}$$

Note the ϵ in the denominator! We see that $\vec{P}$ is radially outward and decreasing with r like $1/r^2$ as $\vec{E}$ does. Note that, even though $\vec{P}$ is position-dependent, its divergence vanishes, so there is no bound charge density. There is surface charge density, negative at $r = a$ and positive at $r = b$. This is to be expected, as the dielectric polarizes so the negative ends of the dipoles are attracted to Q on the conducting sphere and the positive ends are repelled, leaving uncancelled layers of negative charge on the inner boundary and positive charge on the outer boundary.

The electric field is reduced inside the dielectric because the negative charge on the inner boundary screens (generates a field that partially cancels) the field of the free charge on the conducting sphere: the total surface charge density $\sigma_f + \sigma_b$ at $r = a$ is less than $Q/4\,\pi\,a^2$, and it is the *total* charge that determines $\vec{E}$.

Note that, because of the neutrality of the dielectric, the total surface charge on the outer boundary cancels that on the inner boundary, so the net charge enclosed inside a sphere of radius $r > b$ is just Q: outside the dielectric, no screening effect is present.

It is worth thinking about the above a bit: it occurs both because the dielectric has no net charge *and* the problem is spherically symmetric. In contrast, we will see a dielectric sphere can polarize in an external field and generate a field outside itself in spite of having no net charge, which is possible because spherical symmetry is broken in that case. But there is no monopole field, only a dipole field.

Note also that, once you have calculated σ_b and ρ_b, you can ignore the presence of the dielectric: as we stated earlier, the total field is sourced by the sum of the free and bound charge densities and the dielectric has no further effect, which one can see here from noticing that $\vec{E}$ in the dielectric is what one would have calculated if one had been given $\sigma_f + \sigma_b$ at $r = a$.

Finally, let's calculate the electric potential from $\vec{E}$ (not $\vec{D}$!):

$$V(\vec{r}) = -\int_{\infty}^{\vec{r}} d\vec{s}\,' \cdot \vec{E}(\vec{r}\,') = -\int_{\infty}^{r} dr'\, E(r')$$

$$V(r > b) = -\frac{Q}{4\pi}\left[\int_{\infty}^{r} dr'\,\frac{1}{\epsilon_o\, r^2}\right] = \frac{Q}{4\pi}\,\frac{1}{\epsilon_o\, r} \tag{4.36}$$

$$V(a < r < b) = -\frac{Q}{4\pi}\left[\int_{\infty}^{b} dr'\,\frac{1}{\epsilon_o\, r^2} + \int_{b}^{r} dr'\,\frac{1}{\epsilon\, r^2}\right]$$

$$= \frac{Q}{4\pi}\left[\frac{1}{\epsilon_o\, r}\bigg|_{\infty}^{b} + \frac{1}{\epsilon\, r}\bigg|_{b}^{r}\right] = \frac{Q}{4\pi}\left[\frac{1}{b}\left(\frac{1}{\epsilon_o} - \frac{1}{\epsilon}\right) + \frac{1}{\epsilon\, r}\right] \tag{4.37}$$

$$V(r < a) = V(r = a) = \frac{Q}{4\pi}\left[\frac{1}{b}\left(\frac{1}{\epsilon_o} - \frac{1}{\epsilon}\right) + \frac{1}{\epsilon\, a}\right] \tag{4.38}$$

where V is constant for $r < a$ because $r < a$ is occupied by a conductor.

A final comment: if one takes the $\epsilon \to \infty$ limit, one can see that one recovers the behavior one would have if the entire region $r < b$ were filled with conductor. A conductor can be considered to be an infinitely polarizable dielectric, with $\vec{E} = 0$ inside, which requires $\chi_e \to \infty$.

Example 4.3: Parallel plate capacitor with dielectric

You all know from Ph1b that filling the volume between the plates of a parallel-plate capacitor increases the capacitance to $C = \epsilon_r\, C_{vac}$ where C_{vac} is the capacitance with vacuum between the plates. We remind you why this is true.

Let the capacitor plates lie parallel to the xy-plane at $z = 0$ (negative plate) and $z = a$ (positive plate) so $\widehat{z}$ is the unit vector pointing from the negative plate to the positive one. In such a geometry, we know from symmetry that $\vec{E}$, $\vec{D}$, and $\vec{P}$ are all parallel to $\widehat{z}$ and independent of xy, assuming we ignore the capacitor edges. Thus, at the interfaces at $z = 0$ and $z = a$, all these vectors are normal to the interface and so no tangential components are present. These features of the fields imply that we can apply Gauss's Law to the free charge density to find $\vec{D}$.

The free charge density is $\sigma_f = \pm Q/A$ where Q is the charge on the plates ($+Q$ at $z = a$ and $-Q$ at $z = 0$) and A is the plate area. Gauss's Law for an infinite sheet of charge (Griffiths Example 2.5) tells us that the field of a single sheet is $E = \sigma/2\,\epsilon_o$. Therefore, we have for this case

$$\vec{D} = \begin{cases} -\frac{Q}{A}\,\widehat{z} & 0 < z < a \\ 0 & z < 0,\ z > a \end{cases} \tag{4.39}$$

because the fields of the two plates cancel for $z < 0$ and $z > a$ but add for $0 < z < a$, and there is no ϵ_o because we are calculating $\vec{D}$, not $\vec{E}$.

This implies:

$$\vec{E} = \begin{cases} -\frac{1}{\epsilon}\,\frac{Q}{A}\,\widehat{z} & 0 < z < a \\ 0 & z < 0,\, z > a \end{cases} \qquad \vec{P} = \begin{cases} -\frac{\epsilon - \epsilon_o}{\epsilon}\,\frac{Q}{A}\,\widehat{z} & 0 < z < a \\ 0 & z < 0,\, z > a \end{cases} \tag{4.40}$$

$$\rho_b = -\vec{\nabla}\cdot\vec{P} = 0 \tag{4.41}$$

$$\sigma_b = \widehat{n}\cdot\vec{P} = \begin{cases} \widehat{z}\cdot\vec{P}(z=a) & z = a \\ -\widehat{z}\cdot\vec{P}(z=0) & z = 0 \end{cases} = \begin{cases} -\frac{\epsilon-\epsilon_o}{\epsilon}\,\frac{Q}{A} & z = a \\ \frac{\epsilon-\epsilon_o}{\epsilon}\,\frac{Q}{A} & z = 0 \end{cases} \tag{4.42}$$

The electric field is, like in Example 4.2, screened inside the dielectric, with its amplitude reduced by a factor $\epsilon_r = \epsilon/\epsilon_o$. The bound surface charge densities near the plates are the source of this screening, contributing a field opposite to the field of the free charge. The electric field inside the dielectric is the field one expects from surface charge densities $\sigma_f + \sigma_b = \pm(\epsilon_o/\epsilon)\,(Q/A)$. Finally, the voltage is

$$V(0 < z < a) = -\int_0^z d\vec{s}\,'\cdot\vec{E}(\vec{r}\,') = -\int_0^z dz'\left(-\frac{1}{\epsilon}\,\frac{Q}{A}\right) = \frac{1}{\epsilon}\,\frac{Q}{A}\,z \tag{4.43}$$

From this, we can calculate the capacitance, which comes out as expected:

$$C = \frac{Q}{\Delta V} = \frac{Q}{(1/\epsilon)\,(Q/A)\,a} = \epsilon\,\frac{A}{a} = \epsilon_r\,C_{vac} \tag{4.44}$$

C is increased because ΔV is reduced because of the screening inside the dielectric.

Example 4.4: Parallel plate capacitor with two-layer dielectric

Let's repeat, but now with a capacitor that has two slabs of dielectric with different ϵ: ϵ_1 for $0 < z < a$ and ϵ_2 for $a < z < b$, where the top plate is now at $z = b$. Because the interface is normal to $\vec{P}$, we can apply Gauss's Law for $\vec{D}$ as we did before, yielding no change in $\vec{D}$, but now the ϵ quantities in $\vec{E}$ and $\vec{P}$ depend on z.

The volume bound charge density vanishes again. The surface charge density at the top and bottom has the same expression, but again with ϵ being evaluated for the particular value of z. The surface bound charge density at the $z = a$ interface is

$$\sigma_b(z=a) = \hat{n}_1 \cdot \vec{P}_1 + \hat{n}_2 \cdot \vec{P}_2 = \hat{z} \cdot \vec{P}_1 - \hat{z} \cdot \vec{P}_2 = \frac{Q}{A}\left(-\frac{\epsilon_1 - \epsilon_o}{\epsilon_1} + \frac{\epsilon_2 - \epsilon_o}{\epsilon_2}\right) \quad (4.45)$$

Depending on which dielectric constant is greater, this can be positive or negative. Of course, it vanishes if $\epsilon_1 = \epsilon_2$. The potential and capacitance are

$$V(0 < z < a) = \frac{1}{\epsilon_1}\frac{Q}{A}z \qquad V(a < z < b) = \frac{1}{\epsilon_1}\frac{Q}{A}a + \frac{1}{\epsilon_2}\frac{Q}{A}(z-a) \quad (4.46)$$

$$C = \frac{Q}{\Delta V} = \left(\frac{a}{\epsilon_1} + \frac{b-a}{\epsilon_2}\right)^{-1} A = \epsilon_{eff}\frac{A}{b} = \epsilon_{eff,r}\, C_{vac} \quad (4.47)$$

where $1/\epsilon_{eff} = [a/\epsilon_1 + (b-a)/\epsilon_2]/b$ is the thickness-weighted inverse mean of the dielectric constants and $\epsilon_{eff,r} = \epsilon_{eff}/\epsilon_o$. This is the same as two capacitors in series, which is not surprising since that problem has the same equipotential surfaces. The total field is that of three sheets of surface charge $\sigma_f + \sigma_b$, with $\sigma_f = 0$ at the interface between the two dielectrics.

Lecture 13:

Electrostatics in Matter II:
Linear Dielectrics (cont.)
Boundary Value Problems with Linear Dielectrics
Electrostatic Energy for Linear Dielectrics

Date Revised: 2024/02/20 06:30
Revised lecture break
Date Given: 2024/02/15

Example 4.5: Capacitor with two side-by-side (parallel) dielectrics

Now, allow the capacitor to have plate spacing a but with two different dielectrics side-by-side, with ϵ_1 occupying A_1 and $\mathcal{V}_1$ and ϵ_2 occupying A_2 and $\mathcal{V}_2$. It is a reasonable guess that one should treat this as two capacitors in parallel so that

$$C = C_1 + C_2 = \frac{1}{a}\left(\epsilon_1 A_1 + \epsilon_2 A_2\right) \tag{4.48}$$

But let's derive this from scratch, appreciating the subtlety at the interface.

Because the voltage difference between the two plates is independent of ϵ (they are equipotentials), it is reasonable to guess that $\vec{E}$ is the same in ϵ_1 and ϵ_2: this is the key insight! Because the dielectrics are uniform in z, it is also reasonable to assume the field is independent of z as one would have in the single-dielectric case. So, our guess for the form of the fields is:

$$\vec{E} = -E_0\,\hat{z} \quad \vec{D} = \begin{cases} -\epsilon_1 E_0\,\hat{z} & \text{in } \mathcal{V}_1 \\ -\epsilon_2 E_0\,\hat{z} & \text{in } \mathcal{V}_2 \end{cases} \qquad \vec{P} = \begin{cases} -\left(\epsilon_1 - \epsilon_o\right) E_0\,\hat{z} & \text{in } \mathcal{V}_1 \\ -\left(\epsilon_2 - \epsilon_o\right) E_0\,\hat{z} & \text{in } \mathcal{V}_2 \end{cases} \tag{4.49}$$

We see this form respects the tangential boundary conditions at the interface between the two dielectrics, as it has to:

$$\hat{z}\cdot\left(\vec{E}_2 - \vec{E}_1\right) = 0 \qquad \hat{z}\cdot\left(\vec{D}_2 - \vec{D}_1\right) = \left(\epsilon_1 - \epsilon_2\right) E_0 = \hat{z}\cdot\left(\vec{P}_2 - \vec{P}_1\right) \tag{4.50}$$

Because $\vec{D}$ and $\vec{P}$ are different in the two volumes, we must allow the free (and bound) charge densities to be different. This provides us a set of equations to solve for E_0:

$$\epsilon_1 E_0 = \sigma_{f,1} \qquad \epsilon_2 E_0 = \sigma_{f,2} \qquad A_1 \sigma_{f,1} + A_2 \sigma_{f,2} = Q \tag{4.51}$$

$$\Longrightarrow \qquad E_0 = \frac{1}{\epsilon_{eff}} \frac{Q}{A} \qquad \epsilon_{eff} = \frac{\epsilon_1 A_1 + \epsilon_2 A_2}{A_1 + A_2} \qquad A = A_1 + A_2 \tag{4.52}$$

$$C = \frac{Q}{\Delta V} = \frac{Q}{a E_0} = \epsilon_{eff} \frac{A}{a} = \epsilon_{eff,r} C_{vac} \tag{4.53}$$

which matches our parallel-capacitor expectation. The displacement field, polarization field, and free and bound charge densities are

$$\vec{D} = \begin{cases} -\frac{\epsilon_1}{\epsilon_{eff}} \frac{Q}{A} \hat{z} & \text{in } \mathcal{V}_1 \\ -\frac{\epsilon_2}{\epsilon_{eff}} \frac{Q}{A} \hat{z} & \text{in } \mathcal{V}_2 \end{cases} \qquad \vec{P} = \begin{cases} -\frac{\epsilon_1 - \epsilon_o}{\epsilon_{eff}} \frac{Q}{A} \hat{z} & \text{in } \mathcal{V}_1 \\ -\frac{\epsilon_2 - \epsilon_o}{\epsilon_{eff}} \frac{Q}{A} \hat{z} & \text{in } \mathcal{V}_2 \end{cases} \qquad \rho_b = -\vec{\nabla} \cdot \vec{P} = 0 \tag{4.54}$$

$$|\sigma_f| = \begin{cases} \frac{\epsilon_1}{\epsilon_{eff}} \frac{Q}{A} & \text{in } \mathcal{V}_1 \\ \frac{\epsilon_2}{\epsilon_{eff}} \frac{Q}{A} & \text{in } \mathcal{V}_2 \end{cases} \qquad |\sigma_b| = \begin{cases} \frac{\epsilon_1 - \epsilon_o}{\epsilon_{eff}} \frac{Q}{A} \hat{z} & \text{in } \mathcal{V}_1 \\ \frac{\epsilon_2 - \epsilon_o}{\epsilon_{eff}} \frac{Q}{A} \hat{z} & \text{in } \mathcal{V}_2 \end{cases} \tag{4.55}$$

σ_b always has the opposite sign as σ_f. For $Q > 0$, the sign of σ_f is positive at $z = a$ and negative at $z = 0$. Note that, because $\vec{P}$ is different in $\mathcal{V}_1$ and $\mathcal{V}_2$, so too does σ_b differ between the two dielectrics.

Finally, if one calculates the total charge density $\sigma_f + \sigma_b$ at $z = 0$ or $z = a$, one gets

$$\sigma_{t,1} = \sigma_{f,1} + \sigma_{b,1} = \left(\frac{\epsilon_1}{\epsilon_{eff}} - \frac{\epsilon_1 - \epsilon_o}{\epsilon_{eff}} \right) \frac{Q}{A} = \frac{\epsilon_o}{\epsilon_{eff}} \frac{Q}{A} \tag{4.56}$$

$$\sigma_{t,2} = \sigma_{f,2} + \sigma_{b,2} = \left(\frac{\epsilon_2}{\epsilon_{eff}} - \frac{\epsilon_2 - \epsilon_o}{\epsilon_{eff}} \right) \frac{Q}{A} = \frac{\epsilon_o}{\epsilon_{eff}} \frac{Q}{A} \tag{4.57}$$

This makes sense: since the electric field is the same in $\mathcal{V}_1$ and $\mathcal{V}_2$, the total (free + bound) surface charge density sourcing it should be the same. The total charge density is a factor $\epsilon_o/\epsilon_{eff}$ smaller than would be present in the absence of dielectrics because the bound charge density screens the free charge density. The free charge density is different in the two regions because the opposite-sign bound charge density is different because of the different dielectric constants. *In contrast to our naive expectation, the free charge density is not uniform on the conductor; rather, it redistributes itself so the fundamental condition, that the conductors be equipotentials, is satisfied when one includes the effect of the dielectric. Instead, the total charge density is uniform, which yields a field independent of (x, y), which is what ensures the equipotential condition is met.*

Boundary Value Problems with Linear Dielectrics

General Conditions for Linear, Homogeneous Dielectrics

In linear, homogeneous dielectrics,

$$\rho_b = -\vec{\nabla}\cdot\vec{P} = -\vec{\nabla}\cdot\left(\frac{\epsilon-\epsilon_0}{\epsilon}\,\vec{D}\right) = -\left(\frac{\epsilon-\epsilon_0}{\epsilon}\right)\vec{\nabla}\cdot\vec{D} = -\left(\frac{\epsilon-\epsilon_0}{\epsilon}\right)\rho_f \qquad (4.58)$$

(Homogeneity is required so the gradient does not act on ϵ.) Therefore, if there is no free charge density in a linear, homogeneous dielectric, there is no bound charge density either. Thus, the dielectric volume satisfies Laplace's Equation. All our machinery for solving Laplace's Equation applies here.

We always need boundary conditions, though, and we can use the ones we derived earlier (the tangential $\vec{E}$ and $\vec{D}$ conditions will yield the same condition on V, so we start with the simpler one):

$$\widehat{n} \cdot \left[\vec{D}_2 - \vec{D}_1\right] = \sigma_f \qquad \widehat{s} \cdot \left[\vec{E}_2 - \vec{E}_1\right] = 0 \tag{4.59}$$

Writing this in terms of the potential, we have

$$\widehat{n} \cdot \left[\epsilon_2 \vec{\nabla} V_2 - \epsilon_1 \vec{\nabla} V_1\right] = -\sigma_f \qquad \widehat{s} \cdot \left[\vec{\nabla} V_2 - \vec{\nabla} V_1\right] = 0 \tag{4.60}$$

And, we always require $V_1 = V_2$: the potential must be continuous. While we have three conditions, in general the continuity and tangential gradient conditions will be redundant: the normal gradient condition must be independent because it depends on the free surface charge density while the two others do not. The continuity condition is simpler and so is the one that should be used.

Example 4.6: Spherical cavity in a dielectric medium with uniform field applied

Let's apply the above to a spherical cavity of radius R in a medium with permittivity ϵ with a uniform field $\vec{E} = E_0 \hat{z}$ applied. There is no free charge anywhere. Our boundary conditions therefore are

$$V(r \to \infty) = -E_0\, z = -E_0\, r\, P_1(\cos\theta) \tag{4.61}$$

and, with $V_{in}(r) = V(r < R)$ and $V_{out}(r) = V(r > R)$,

$$\epsilon_o \left.\frac{\partial V_{in}}{\partial r}\right|_{r=R} = \epsilon \left.\frac{\partial V_{out}}{\partial r}\right|_{r=R} \qquad \text{and} \qquad V_{in}(r = R) = V_{out}(r = R) \tag{4.62}$$

We also choose the zero of the potential to be at $z = 0$, $V(z = 0) = 0$, by symmetry as in the case of the conducting sphere in a uniform electric field.

As usual, we begin by writing our generic solutions to Laplace's Equation in spherical coordinates:

$$V_{in}(r) = \sum_{\ell=0}^{\infty} A_\ell^{in} r^\ell P_\ell(\cos\theta) \qquad V_{out}(r) = \sum_{\ell=0}^{\infty} \left(A_\ell^{out} r^\ell + \frac{B_\ell^{out}}{r^{\ell+1}} \right) P_\ell(\cos\theta) \tag{4.63}$$

where we have applied the requirement that V be finite at the origin to eliminate the $1/r^{\ell+1}$ terms for V_{in}. Recall that we cannot eliminate the r^ℓ terms for V_{out} because the potential does not vanish at infinity.

Let's first apply the $r \to \infty$ condition. We did this before in the case of a metal sphere in a uniform field, and we found

$$A_1^{out} = -E_0 \qquad A_{\ell\neq 1}^{out} = 0 \tag{4.64}$$

Next, we apply the continuity condition at $r = R$, making use of orthonormality of the P_ℓ:

$$A_1^{in} R = -E_0 R + \frac{B_1^{out}}{R^2} \qquad A_{\ell\neq 1}^{in} R^\ell = \frac{B_{\ell\neq 1}^{out}}{R^{\ell+1}} \tag{4.65}$$

Finally, let's take the radial derivative and apply the matching condition on it, again using orthonormality:

$$\epsilon_0 A_1^{in} = -\epsilon \left(E_0 + \frac{2}{R^3} B_1^{out} \right) \qquad \epsilon_0 A_{\ell \neq 1}^{in} \, \ell \, R^{\ell - 1} = -\epsilon \, \frac{B_{\ell \neq 1}^{out}}{R^{\ell + 2}} \, (\ell + 1) \tag{4.66}$$

Doing the algebra, we find

$$A_{\ell \neq 1}^{in} = B_{\ell \neq 1}^{out} = 0 \qquad B_1^{out} = -\frac{\epsilon - \epsilon_o}{2\,\epsilon - \epsilon_o} \, E_0 \, R^3 \qquad A_1^{in} = -\frac{3\,\epsilon}{2\,\epsilon + \epsilon_o} \, E_0 \tag{4.67}$$

Thus, the potential is

$$V_{in}(r) = V(r < R) = -\frac{3\,\epsilon}{2\,\epsilon + \epsilon_o} \, E_0 \, r \, \cos\theta = -\frac{3\,\epsilon}{2\,\epsilon + \epsilon_o} \, E_0 \, z \tag{4.68}$$

$$\begin{aligned} V_{out}(r) = V(r > R) &= -E_0 \, r \, \cos\theta - \frac{\epsilon - \epsilon_o}{2\,\epsilon + \epsilon_o} \, E_0 \frac{R^3}{r^2} \, \cos\theta \\ &= -E_0 \, z + \frac{\vec{p} \cdot \widehat{r}}{4\,\pi\,\epsilon_o \, r^2} \quad \text{with} \quad \vec{p} = -\frac{4\,\pi}{3} \, R^3 \, E_0 \frac{3\,\epsilon_o}{2\,\epsilon + \epsilon_o} \, (\epsilon - \epsilon_o) \, \widehat{z} \end{aligned} \tag{4.69}$$

The potential inside the cavity is that of a uniform electric field in the same direction as the applied field but multiplied by the factor $3\,\epsilon/(2\,\epsilon+\epsilon_o) > 1$, while the potential outside is that of the uniform field plus that of a dipole whose orientation is *opposite* the uniform field and whose magnitude is given above. It is as if the *cavity* acquired a polarization density in the negative z direction, though of course that cannot happen because $\chi_e(r<R)=0$ there and thus $\vec{P}(r<R)=\epsilon_o\chi_e(r<R)\vec{E}(r<R)=0$. The polarization density outside the cavity is just the *total* (not the applied uniform) field times $\epsilon-\epsilon_o$ (which is not particularly illuminating).

The (bound) surface charge density is

$$\begin{aligned}\sigma_b &= \widehat{n}\cdot\vec{P}(r=R) = \widehat{n}\cdot(\epsilon-\epsilon_o)\,\vec{E}_{out}(r=R)\\ &= (\epsilon-\epsilon_o)\left(-\widehat{r}\cdot E_0\,\widehat{z} - \frac{\partial}{\partial r}\,\frac{\epsilon-\epsilon_o}{2\,\epsilon+\epsilon_o}\,E_0\,\frac{R^3}{r^2}\,\cos\theta\bigg|_{r=R}\right) = -3\,\epsilon_o\,\frac{\epsilon-\epsilon_o}{2\,\epsilon+\epsilon_o}\,E_0\cos\theta\end{aligned}$$

(Notice that $\widehat{n}=-\widehat{r}$ because $\widehat{n}$ is taken to point *out of* the dielectric medium in the definition of σ_b.) We see the boundary of the cavity acquires a surface charge density with the same magnitude and cosine dependence as the bound charge on the surface of a uniformly polarized sphere, though with *opposite* sign (so there is negative charge at the $+z$ end and positive charge at the $-z$ end). The sign follows naturally from our arguments about cancellation of dipole charge.

The field is enhanced in the cavity for two reasons: first, there is no polarizable material to screen the electric field, and, second there is surface charge density on the cavity's boundary that creates an additional field in the direction of the uniform field.

For reference, we note that the solution for a dielectric sphere (Griffiths Example 4.7) in a uniform field looks very similar:

$$V(r<R) = -\frac{3\,\epsilon_o}{2\,\epsilon_o+\epsilon}\,E_0 z \qquad V(r>R) = -E_0 z + \frac{\vec{p}\cdot\widehat{r}}{4\,\pi\,\epsilon_o\,r^2} \tag{4.70}$$

$$\text{with} \quad \vec{p} = \frac{4\,\pi}{3}\,R^3\,E_0\,\frac{3\,\epsilon_o}{2\,\epsilon_o+\epsilon}\,(\epsilon-\epsilon_o)\,\widehat{z} \equiv \frac{4\,\pi}{3}\,R^3\,\vec{P}(r<R) \tag{4.71}$$

$$\sigma_b = 3\,\epsilon_o\,\frac{\epsilon-\epsilon_o}{2\,\epsilon_o+\epsilon}\,E_0\cos\theta \tag{4.72}$$

Basically, exchange ϵ_o and ϵ everywhere to go between the two results. In this case, the sphere acquires a polarization density $3\,\epsilon_o(\epsilon-\epsilon_o)/(2\,\epsilon_o+\epsilon)$, now in the direction of the applied field. The surface charge density is also of same form as the cavity case with the $\epsilon \leftrightarrow \epsilon_o$ exchange. That exchange flips the sign so that the $+z$ end acquires a positive charge, again as expected from the dipole charge cancellation argument. The field amplitude is reduced (screened) in the dielectric.

From the polarized sphere, one can recover the case of a conducting sphere in an external uniform field by taking $\epsilon \to \infty$ as noted earlier.

Electrostatic Energy in and Forces on Linear Dielectrics

Electrostatic Potential Energy due to an Assembly of Free Charge in the Presence of Dielectrics

It turns out that electrostatic potential energy in the presence of dielectrics is a subtle topic because of the existence of the charges forming the dielectric. There are different kinds of electrostatic potential energy: that needed to assemble the free and bound charge distributions versus that needed to assemble the free charge distribution and polarize the preexisting dielectric. It is generally the latter we are interested in, so we consider that case.

Suppose we have a system in which an electric field $\vec{E}(\vec{r})$ and its potential $V(\vec{r})$ have already been set up and we want to bring in additional *free charge* $\delta\rho_f$ from infinity (assuming the potential vanishes at infinity). In this case, the change in potential energy is

$$\delta U = \int_{\mathcal{V}} d\tau' \left[\delta\rho_f(\vec{r}')\right] V(\vec{r}') \tag{4.73}$$

The free charge density is related to the displacement field by $\vec{\nabla}\cdot\vec{D} = \rho_f$, so a change $\delta\rho_f$ corresponds to a change in the divergence of the displacement field $\delta\left(\vec{\nabla}\cdot\vec{D}\right)$. Linearity of the divergence lets us rewrite this as $\delta\rho_f = \vec{\nabla}\cdot\delta\vec{D}$.

Then, we may integrate by parts and apply the divergence theorem:

$$\begin{aligned}\delta U &= \int_{\mathcal{V}} d\tau' \left[\vec{\nabla}\cdot\delta\vec{D}(\vec{r}\,')\right] V(\vec{r}\,') \\ &= \int_{\mathcal{V}} d\tau' \vec{\nabla}\cdot\left[V(\vec{r}\,')\,\delta\vec{D}(\vec{r}\,')\right] - \int_{\mathcal{V}} d\tau' \left[\delta\vec{D}(\vec{r}\,')\right]\cdot\vec{\nabla}V(\vec{r}\,') \\ &= \oint_{\mathcal{S}(\mathcal{V})} da'\,\hat{n}(\vec{r}\,')\cdot\left[V(\vec{r}\,')\,\delta\vec{D}(\vec{r}\,')\right] + \int_{\mathcal{V}} d\tau' \left[\delta\vec{D}(\vec{r}\,')\right]\cdot\vec{E}(\vec{r}\,') \qquad (4.74)\end{aligned}$$

Assuming the potential falls off at infinity, the surface term can be taken out to infinity to vanish. (Recall, $V \sim 1/r$ and $D \sim 1/r^2$ while $da'\ r^2$, so the integral falls off as $1/r$.) So, we are then left with

$$U = \int_0^{\vec{D}} \int_{\mathcal{V}} d\tau'\,\vec{E}(\vec{r}\,')\cdot d\vec{D}(\vec{r}\,') \qquad (4.75)$$

There are two integrals here, one over volume and one over the value of $\vec{D}$ from zero to its final value. $\vec{E}$ is of course tied to $\vec{D}$ and they vary together.

For the case of a linear (but perhaps not homogeneous) dielectric, we may use $\vec{D}(\vec{r}) = \epsilon(\vec{r})\vec{E}(\vec{r})$ and therefore

$$\begin{aligned} U &= \int_0^{\vec{E}} \int_{\mathcal{V}} d\tau' \, \epsilon(\vec{r}') \, \vec{E}(\vec{r}') \cdot d\vec{E}(\vec{r}') \\ &= \frac{1}{2} \int_0^{\vec{E}} \int_{\mathcal{V}} d\tau' \epsilon(\vec{r}') \, d \left[\vec{E}(\vec{r}') \cdot \vec{E}(\vec{r}') \right] \\ &= \frac{1}{2} \int_{\mathcal{V}} d\tau' \, \epsilon(\vec{r}') \, E^2(\vec{r}') = \frac{1}{2} \int_{\mathcal{V}} d\tau' \, \vec{E}(\vec{r}') \cdot \vec{D}(\vec{r}') \end{aligned} \tag{4.76}$$

If the medium is linear *and* homogeneous, one can pull ϵ outside the integral at any point, yielding

$$U = \frac{\epsilon}{2} \int_{\mathcal{V}} d\tau' \, \left| \vec{E}(\vec{r}') \right|^2 = \frac{1}{2\,\epsilon} \int_{\mathcal{V}} d\tau' \, \left| \vec{D}(\vec{r}') \right|^2 \tag{4.77}$$

We may infer that the energy density, neglecting the energy density intrinsic to the creation of the dipoles, is

$$u(\vec{r}) = \frac{\epsilon}{2} \, \left| \vec{E}(\vec{r}) \right|^2 = \frac{1}{2\,\epsilon} \, \left| \vec{D}(\vec{r}) \right|^2 \tag{4.78}$$

By contrast, if we wanted to know the total electrostatic potential energy stored in the assembly of the free and bound charge, we would just do the usual volume integral of E^2 with ϵ_o instead of ϵ. That energy is smaller because $\epsilon > \epsilon_o$. The reason for this difference is that assembling the medium in the first place, which consists of bringing positive and negative charges together, creates a system with negative potential energy, and thus the total potential energy of the system would be lower if we accounted for the energy of assembling the medium. But we will never pull the dielectric apart, so it is natural to treat that component of the potential energy as an offset that is inaccessible and neglect it in the electrostatic potential energy.

Energy of a Dielectric in an External Field

A topic naturally related to the above is the electrostatic energy of a polarizable material in an external field.

Suppose we start with a system with a free charge distribution ρ_f that sources a field $\vec{E}_1$ in a dielectric medium ϵ_1, yielding a displacement $\vec{D}_1 = \epsilon_1 \vec{E}_1$. The initial energy is

$$U_1 = \frac{1}{2} \int d\tau \, \vec{E}_1 \cdot \vec{D}_1 \tag{4.79}$$

Now, *with the charges sourcing $\vec{E}_1$ held fixed*, let's introduce a piece of dielectric occupying the volume $\mathcal{V}_2$ and having dielectric constant ϵ_2, replacing the dielectric of dielectric constant ϵ_1 there. The remainder of space outside $\mathcal{V}_2$ is occupied by ϵ_1 in both configurations. The electric field and displacement field everywhere change to $\vec{E}_2$ and $\vec{D}_2$, where $\vec{D}_2(\vec{r}) = \epsilon(\vec{r}) \, \vec{E}_2(\vec{r})$. *Note that $\vec{E}_1$ and $\vec{E}_2$ are not identical outside $\mathcal{V}_2$, and the same is true for $\vec{D}_1$ and $\vec{D}_2$. The dielectric affects the field everywhere, not just inside $\mathcal{V}_2$.* The energy is now

$$U_2 = \frac{1}{2} \int d\tau \, \vec{E}_2 \cdot \vec{D}_2 \tag{4.80}$$

The difference in energy between the two configurations is therefore

$$U_2 - U_1 = \frac{1}{2} \int d\tau \left[\vec{E}_2 \cdot \vec{D}_2 - \vec{E}_1 \cdot \vec{D}_1 \right] \tag{4.81}$$

Let us rewrite the energy difference as

$$U_2 - U_1 = \frac{1}{2}\int d\tau \left[\vec{E}_2 \cdot \vec{D}_1 - \vec{E}_1 \cdot \vec{D}_2\right] + \frac{1}{2}\int d\tau \left[\vec{E}_2 + \vec{E}_1\right] \cdot \left[\vec{D}_2 - \vec{D}_1\right] \tag{4.82}$$

It holds that $\vec{\nabla} \times \left[\vec{E}_2 + \vec{E}_1\right] = 0$, so it can be derived from a potential V, so the second integral becomes

$$-\frac{1}{2}\int d\tau \left(\vec{\nabla} V\right) \cdot \left[\vec{D}_2 - \vec{D}_1\right] \tag{4.83}$$

We integrate by parts (the surface term vanishes because it depends on $\vec{D}_2 - \vec{D}_1$, which should vanish as one goes far from the dielectric) to obtain

$$\frac{1}{2}\int d\tau \, V \, \vec{\nabla} \cdot \left[\vec{D}_2 - \vec{D}_1\right] \tag{4.84}$$

This divergence vanishes because the free charge has not changed between the two configurations (recall, $\vec{\nabla} \cdot \vec{D} = \rho_f$).

So the second term in the energy vanishes, leaving

$$U_2 - U_1 = \frac{1}{2}\int d\tau \left[\vec{E}_2 \cdot \vec{D}_1 - \vec{E}_1 \cdot \vec{D}_2\right] \tag{4.85}$$

Now, outside $\mathcal{V}_2$, it holds that $\vec{D}_2 = \epsilon_1 \vec{E}_2$ (remember, ϵ only changed inside $\mathcal{V}_2$), and recall also $\vec{D}_1 = \epsilon_1 \vec{E}_1$ everywhere, so the two terms cancel each other there and the integrand vanishes outside $\mathcal{V}_2$. Therefore, we can restrict the integral to $\mathcal{V}_2$:

$$U_2 - U_1 = -\frac{1}{2}\int_{\mathcal{V}_2} d\tau \, (\epsilon_2 - \epsilon_1)\, \vec{E}_2 \cdot \vec{E}_1 \tag{4.86}$$

This is already interesting — even though the field changes in all of space, we need only look at the before and after fields in the volume $\mathcal{V}_2$ rather than the entire system. If $\epsilon_1 = \epsilon_o$ (vacuum outside $\mathcal{V}_2$ and in $\mathcal{V}_2$ before the introduction of ϵ_2), then we can use $\vec{P} = (\epsilon_2 - \epsilon_o)\, \vec{E}_2$ to rewrite as

$$W = U_2 - U_1 = -\frac{1}{2}\int_{\mathcal{V}_2} d\tau \, \vec{P} \cdot \vec{E}_1 \qquad \Longleftrightarrow \qquad w = -\frac{1}{2}\, \vec{P} \cdot \vec{E}_1 \tag{4.87}$$

where we recall that $\vec{E}_1$ *is the electric field in the absence of the dielectric* and $\vec{P}$ is the polarization density of the dielectric, and w refers to an energy density. This is just like the energy of a dipole in an external electric field, except that the factor of $1/2$ accounts for the integration from zero field to actual field, from the fact that the dielectric polarizes in response to the applied field. *We see that the introduction of the dielectric into an existing electric field in vacuum, holding the source charges fixed, reduces the overall electrostatic energy.*

Why is the integrand not $\vec{P} \cdot \vec{E}_2$ or $\vec{D}_2 \cdot \vec{E}_2$? Because we are asking for the *difference* in energy between the field configurations without and with the dielectric. There was field in $\mathcal{V}_2$ before the dielectric was placed there, so we have to subtract off that original field energy density, and we also need to consider the field energy density difference between the two configurations *outside* the dielectric. It turns out that the above integrand correctly accounts for the differencing relative to the no-dielectric starting condition. We can see this by trying to evaluate the potential alternate expressions:

$$-\frac{1}{2}\int_{\mathcal{V}_2} d\tau\, \vec{P} \cdot \vec{E}_2 = -\frac{1}{2}\int_{\mathcal{V}_2} d\tau\, \left(\vec{D}_2 - \epsilon_o \vec{E}_2\right) \cdot \vec{E}_2 = \frac{1}{2}\int_{\mathcal{V}_2} d\tau\, \left[\epsilon_o \left|\vec{E}_2\right|^2 - \vec{D}_2 \cdot \vec{E}_2\right] \tag{4.88}$$

This is some sort of difference between the total electrostatic potential energy in $\mathcal{V}_2$ and the electrostatic potential energy neglecting that associated with the assembly of the dielectric medium. The expression has two problems: there is no differencing with the initial configuration, and it neglects the energy stored in $\mathcal{V}_1$. It is part of the energy difference we are interested in, but not all of. The use of $\vec{D}_2 \cdot \vec{E}_2$ would suffer the same problems.

Lecture 14:

Electrostatics in Matter III:
Electrostatic Forces and Torques for Linear Dielectrics

Magnetostatics I:
Lorentz Force, Biot-Savart Law, Fields and Forces

Date Revised: 2024/02/20 06:30
Date Given: 2024/02/22

Force and Torque on a Linear, Homogeneous Dielectric in an External Field with Free Charge Fixed

Let us first consider the force on the dielectric in the case that the free charge is held fixed. There are no batteries involved, so we need only consider the electrostatic energy of the field. We take the negative of its gradient with respect to some generalized displacement ξ to find the generalized force F_ξ:

$$F_\xi\big|_Q = -\left(\frac{\partial W}{\partial \xi}\right)_Q = -\left(\frac{\partial W}{\partial C}\right)_Q \frac{\partial C}{\partial \xi} \tag{4.89}$$

where we made the second step because, if Q is held fixed, the variation of the system energy is given entirely by the variation of the capacitance. ξ can be a spatial displacement coordinate like x, y, or z, or it can be an angular orientation coordinate, in which case the generalized force is actually a torque.

Any system of conductors can be reduced to a capacitance matrix, so the above can also be written using Equation 2.82 (recall, $\underline{\underline{D}} = \underline{\underline{C}}^{-1}$)

$$F_\xi\big|_Q = -\left.\frac{\partial}{\partial \xi}\frac{1}{2}\sum_{i,j=1}^{N} Q_i\, Q_j\, D_{ij}\right|_Q = -\frac{1}{2}\sum_{i,j=1}^{N} Q_i\, Q_j\, \frac{\partial D_{ij}}{\partial \xi} = -\frac{1}{2}\underline{Q}^T\left[\frac{\partial}{\partial \xi}\underline{\underline{C}}^{-1}\right]\underline{Q} \tag{4.90}$$

(We have intentionally avoided using the confusing notation C_{ij}^{-1}, using D_{ij} instead.)

Example 4.7: Force on a Dielectric Slab in a Parallel Plate Capacitor, Free Charge Fixed

Let's consider a parallel-plate capacitor with plate separation d, plate side dimensions ℓ and w, and with a slab of linear, homogeneous dielectric partially inserted between the plates, with vacuum from 0 to x and dielectric from x to ℓ with $0 < x < \ell$.

Let's do this by calculating the total energy of the slab in the capacitor, with E dependent on the position of the slab. The energy is (using the calculation of C from the earlier example)

$$W = \frac{1}{2}\frac{Q^2}{C} \qquad \text{with} \qquad C = \frac{\epsilon_o\, w\, x + \epsilon\, w\,(\ell - x)}{d} \tag{4.91}$$

Therefore,

$$F_x|_Q = -\left(-\frac{1}{2}\frac{Q^2}{C^2}\right)\frac{dC}{dx} = \frac{1}{2}\frac{Q^2}{C^2}\frac{(\epsilon_o - \epsilon)\, w}{d} = -\frac{1}{2}\, V^2\,(\epsilon - \epsilon_o)\,\frac{w}{d} \tag{4.92}$$

which matches Griffiths Equation 4.65 (recall, $\epsilon_o \chi_e = \epsilon - \epsilon_o$).

Intuitively, the dielectric is pulled in because it lowers the energy of the configuration: the field energy density is proportional to $\epsilon|\vec{E}|^2$, and $|\vec{E}| \propto \epsilon^{-1}$, so the field energy density is $\propto \epsilon^{-1}$: larger ϵ implies lower energy.

Microscopically, what is happening is that the fringing field of the capacitor polarizes the dielectric, leading to bound charge on the surface. The bound charge on the surface is attracted to the free charge on the capacitor plates, causing the dielectric to be pulled in. It's a runaway effect, with the movement of the dielectric into the capacitor leading to greater polarization of the fringing field region, increasing the bound surface charge density and thereby leading to a greater attractive force. The system only reaches equilibrium when the dielectric is maximally contained in the capacitor. (It would be interesting to calculate the trajectory, in particular the harmonic oscillations that would occur around the equilibrium position because the slab will have been accelerated and thus have some kinetic energy when it gets to the equilibrium position.)

Force and Torque on a Linear, Homogeneous Dielectric in an External Field with Voltages Fixed

In general, we do not encounter the above situation. Rather, we hold the voltages constant on a set of electrodes while we consider the work done during a virtual displacement $d\xi$.

Before we get into it, though, let's ask ourselves what we expect to have happen. Should the force change depending on whether we hold the voltage or the charge fixed? No, because the force is due to the arrangement of charges on the conductors and the dielectric at the current instant in time, not at some point in the future that is affected by whether the charges or voltages are kept constant.

Let's model the fixed voltage situation in two steps, first disconnecting the batteries and holding the charge fixed while we move the dielectric as we did above, then reconnecting the batteries so that charge flows on to or off of the electrodes and restores them to their original potentials.

Since we are now focusing on a situation with voltages on electrodes, it makes sense to think about a set of electrodes $i = 1...N$ with voltages V_i and charges Q_i. The electrodes have a capacitance matrix $\underline{\underline{C}}$. Let's first consider the change in electrostatic energy for the first step with the charges held fixed (again, using $\underline{\underline{D}} = \underline{\underline{C}}^{-1}$):

$$dW_{field}|_Q = d\left[\frac{1}{2}\sum_{i,j=1}^{N} Q_i Q_j D_{ij}\right]_Q = \frac{1}{2}\sum_{i,j=1}^{N} Q_i Q_j \, dD_{ij} \tag{4.93}$$

The change in the inverse capacitance matrix results in a change in the voltages on the electrodes given by

$$dV_i|_Q = \sum_{j=1}^{N} dD_{ij} Q_j \tag{4.94}$$

Now, let's return the voltages to their original values by allowing charge to flow on/off the electrodes from batteries while holding the dielectrics fixed (*i.e.*, D_{ij} held constant). The charge transfer required to undo the above voltage changes is

$$dQ_k|_V = \sum_{i=1}^{N} C_{ki}\,(-dV_i)_Q = -\sum_{i,j=1}^{N} C_{ki}\,Q_j\,dD_{ij} \tag{4.95}$$

The change in the electrostatic energy of the configuration (energy flowing out of the battery into the field) due to this flow of charge is

$$\begin{aligned} dW^{bat}_{field}\Big|_V &= \sum_{k=1}^{N} V_k\,dQ_k|_V = -\sum_{i,j,k=1}^{N} V_k\,C_{ki}\,Q_j\,dD_{ij} = -\sum_{i,j=1}^{N} Q_i\,Q_j\,dD_{ij} \\ &= -2\,dW_{field}|_Q \end{aligned} \tag{4.96}$$

where we used $C_{ki} = C_{ik}$ and $\sum_{k=1}^{N} V_k C_{ik} = Q_i$.

Therefore, the total infinitesimal change in energy is

$$dW_{field}|_V = dW_{field}|_Q + dW^{bat}_{field}\Big|_V = dW_{field}|_Q - 2\, dW_{field}|_Q = -\, dW_{field}|_Q \quad (4.97)$$

As we explained earlier, the force cannot depend on whether the charge is held fixed or the voltage is held fixed. To ensure we get the same force in the two cases, we therefore must conclude

$$F_\xi|_V = \left(\frac{\partial W_{field}}{\partial \xi}\right)_V = -\left(\frac{\partial W_{field}}{\partial \xi}\right)_Q = F_\xi|_Q \quad (4.98)$$

That is, when the battery is involved, we must consider the energy of the entire system and take the positive gradient of the field energy rather than considering only the energy of the field and taking the negative gradient of that energy. The reason these two gradients are different, with a sign between them, is because the derivative is calculationally different depending on whether V or Q is held fixed.

We can see this works mathematically by trying it:

$$\left(\frac{\partial W_{field}}{\partial \xi}\right)_V = \frac{\partial}{\partial \xi}\left[\frac{1}{2}\sum_{i,j=1}^{N} V_i\, V_j\, C_{ij}\right]_V = \frac{1}{2}\sum_{i,j=1}^{N} V_i\, V_j\, \frac{\partial C_{ij}}{\partial \xi}$$
$$= \frac{1}{2}\,\underline{V}^T\left[\frac{\partial}{\partial \xi}\underline{\underline{C}}\right]\underline{V} \tag{4.99}$$

Since $\partial \underline{\underline{C}}^{-1}/\partial\xi = -\underline{\underline{C}}^{-1}[\partial \underline{\underline{C}}/\partial\xi]\underline{\underline{C}}^{-1}$ (one can see this by evaluating $\partial[\underline{\underline{C}}\,\underline{\underline{C}}^{-1}]/\partial\xi = \partial \underline{\underline{1}}/\partial\xi = \underline{\underline{0}}$), this form yields Equation 4.90 for $F_\xi\big|_Q$. Thus,

$$F_\xi\big|_V = \left(\frac{\partial W_{field}}{\partial \xi}\right)_V = -\left(\frac{\partial W_{field}}{\partial \xi}\right)_Q = F_\xi\big|_Q \tag{4.100}$$

One can check this result using the parallel plate capacitor example by starting with $W = C\,V^2/2$ instead of $W = Q^2/2\,C$. Taking the positive derivative at fixed V gives the same result as taking the negative derivative at fixed Q because C is in the numerator in the first case while C is in the denominator in the second.

Section 5
Magnetostatics

Study Guidelines

As with basic electrostatics, you have seen much of the material in this section before in Ph1c. As with electrostatics, we will use more rigor here. We will also consider some more advanced topics such as the multipole expansion of the magnetic vector potential, off-axis fields for azimuthally symmetric configurations, etc. As with basic electrostatics, we won't do any examples in lecture or the notes where they would duplicate Ph1c. But you should be review the examples in Griffiths Chapter 5 and make sure you are comfortable with them.

Lorentz Forces and Current Densities

Force on a Moving Point Charge in a Magnetic Field

The magnetic force on a point charge q moving with velocity $\vec{v}$ in a magnetic field $\vec{B}$ is given by the *Lorentz Force Law*:

$$\boxed{\vec{F}_{mag} = q\left(\vec{v} \times \vec{B}\right)} \tag{5.1}$$

If an electric field is present, the total electrostatic and magnetostatic force on q is

$$\vec{F} = q\left(\vec{E} + \vec{v} \times \vec{B}\right) \tag{5.2}$$

Note that the electrostatic force on q is not modified by the fact that it is moving.

See the nice examples in Griffiths of cyclotron and cycloid motion (Examples 5.1 and 5.2). These are at the level of Ph1c, so we do not spend time in lecture on them.

Magnetic Forces Do No Work

Because $\vec{F}_{mag} \propto \vec{v} \times \vec{B}$, it holds that $\vec{F}_{mag} \perp \vec{v}$. Since the differential of work done by a force is $dW = \vec{F} \cdot d\vec{\ell} = \vec{F} \cdot \vec{v}\, dt$, we thus see that $dW = 0$ for magnetic forces. This may seem counterintuitive. In cases where it appears work is being done, there is usually a battery involved that is doing the work, while the magnetic force is redirecting the force doing the work (in the same way that a constraint force in mechanics does no work).

The one exception to this is the case of intrinsic magnetic moments of fundamental particles, which emerge from quantum field theory. In such cases, the magnetic moment is not identified with a current loop, it is just an intrinsic property of the particle. Since our proof above requires the Lorentz Force Law, and such moments are not assocated with a current that experiences the Lorentz Force, the proof does not apply. In cases concerning such moments, work can be done by the field of the moment or on the magnetic moment by an external magnetic field because no battery is required to maintain the magnetic moment.

Line Currents

A current carried by a wire can be modeled as a constant line charge density λ that is moving at fixed speed v:

$$\mathrm{I} = \lambda\, v \tag{5.3}$$

For the sake of the generalizations we will consider below, let us write this as a position-dependent vector

$$\vec{\mathrm{I}}(\vec{r}) = \lambda(\vec{r})\, \vec{v}(\vec{r}) \tag{5.4}$$

where $\vec{v}(\vec{r})$ is a function of position and its direction follows the wire. By conservation of charge, the only position dependence of $\vec{\mathrm{I}}(\vec{r})$ can be its direction. This implies that any position dependence in $\lambda(\vec{r})$ must be canceled by the position dependence of the magnitude of $\vec{v}(\vec{r})$. If λ is position-independent, then only the direction of $\vec{v}$ may change with position.

For magnetostatics, we assume that such a line current, and the surface and volume current densities that follow below, are time-independent, or **steady**: they were set up an infinitely long time ago and have been flowing at their current values since then. We also ignore the discretization of the charge density (in this case λ) and consider it to be a continuous quantity. This is called the **steady-state** assumption or approximation.

Force on a Line Current

It is straightforward to calculate the force on a line current by integrating the Lorentz Force Law over the wire:

$$\vec{F}_{mag} = \int dq \left[\vec{v}(\vec{r}) \times \vec{B}(\vec{r})\right] = \int_{\mathcal{C}} d\ell\, \lambda \left[\vec{v}(\vec{r}) \times \vec{B}(\vec{r})\right] \tag{5.5}$$

$$\boxed{\vec{F}_{mag} = \int_{\mathcal{C}} d\ell \left[\vec{\mathrm{I}}(\vec{r}) \times \vec{B}(\vec{r})\right]} \tag{5.6}$$

where we have used the fact that $d\vec{\ell}$, $\vec{v}$, and $\vec{\mathrm{I}}$ are all in the same direction at any point on the wire because the current flows in the wire. Now, realizing that I is independent of position along the wire (due to conservation of charge as noted above), we can pull it out in front of the integral, yielding

$$\vec{F}_{mag} = \mathrm{I} \int_{\mathcal{C}} \left[d\vec{\ell} \times \vec{B}(\vec{r})\right] \tag{5.7}$$

Griffiths Example 5.3 is a nice example of calculating the force on a current loop and also illustrates the point of the battery supplying the energy to do the work that appears to be done by the magnetic field. The magnetic field acts like a constraint force to redirect that work.

Current Densities

Just as we generalized point charges to line, surface, and volume charge densities, we can generalize single moving point charges to line, surface, and volume current densities. We have already made the first generalization, which is straightforward to understand since one intuitively thinks of a current as an ensemble of point charges moving through a wire.

A *surface current density* is a current flowing in a sheet; think of water flowing over the surface of an object. The surface current density $\vec{K}$ is defined by

$$d\vec{\mathrm{I}}(\vec{r}) = \vec{K}(\vec{r})\, d\ell_\perp = \left|\widehat{K}(\vec{r}) \times d\vec{\ell}\right| \vec{K}(\vec{r}) \tag{5.8}$$

where $d\ell_\perp$ is an infinitesimal length perpendicular to $\vec{K}$ and $d\vec{\ell}$ is an arbitrary infinitesimal length. The cross-product takes the projection of $d\vec{\ell}$ perpendicular to $\vec{K}$.

If one thinks about the surface current density as a moving distribution of a surface charge density, then

$$\vec{K}(\vec{r}) = \sigma(\vec{r})\, \vec{v}(\vec{r}) \tag{5.9}$$

where $\sigma(\vec{r})$ is the surface charge density at $\vec{r}$ and $\vec{v}(\vec{r})$ is the velocity of the surface charge density at $\vec{r}$.

A *volume current density* is a current flowing in a bulk volume; think of water flowing in a pipe or in a river. The volume current density $\vec{J}$ is defined by

$$d\vec{\mathrm{I}}(\vec{r}) = \vec{J}(\vec{r})\, da_\perp = \left|\widehat{J}(\vec{r}) \cdot \widehat{n}\right| da\, \vec{J}(\vec{r}) \tag{5.10}$$

where $\widehat{n}$ is the normal to the area element da. (If we had defined a normal $\widehat{n}$ to the line element $d\vec{\ell}$ in the plane of the sheet, we could have used a dot product instead of a cross product in the definition of the surface current density. But it is conventional to do it as we have done it.)

If one thinks about the volume current density as a moving distribution of a volume charge density, then

$$\vec{J}(\vec{r}) = \rho(\vec{r})\, \vec{v}(\vec{r}) \tag{5.11}$$

where $\rho(\vec{r})$ is the volume charge density at $\vec{r}$ and $\vec{v}(\vec{r})$ is the velocity of the volume charge density at $\vec{r}$.

Forces on Current Densities

We can integrate the force over the current densities just as we did for the line current:

$$\vec{F}_{mag} = \int dq \left[\vec{v}(\vec{r}) \times \vec{B}(\vec{r})\right] = \int_{\mathcal{S}} da\, \sigma(\vec{r}) \left[\vec{v}(\vec{r}) \times \vec{B}(\vec{r})\right] \tag{5.12}$$

$$\boxed{\vec{F}_{mag} = \int_{\mathcal{S}} da \left[\vec{K}(\vec{r}) \times \vec{B}(\vec{r})\right]} \tag{5.13}$$

$$\vec{F}_{mag} = \int dq \left[\vec{v}(\vec{r}) \times \vec{B}(\vec{r})\right] = \int_{\mathcal{V}} d\tau\, \rho(\vec{r}) \left[\vec{v}(\vec{r}) \times \vec{B}(\vec{r})\right] \tag{5.14}$$

$$\boxed{\vec{F}_{mag} = \int_{\mathcal{V}} d\tau \left[\vec{J}(\vec{r}) \times \vec{B}(\vec{r})\right]} \tag{5.15}$$

It should be clear that we could have considered Equation 5.15 to be the fundamental statement of the Lorentz Force Law and derived the lower-dimensional versions by inclusion of appropriate delta functions in the definition of ρ or $\vec{J}$. Such a reduction would be cumbersome because the sheet or line carrying the current may not be easy to parameterize, but the reduction is conceptually straightforward.

Conservation of Charge and the Continuity Equation

We defined the current densities above in terms of the infinitesimal current passing through an infinitesimal line element (for a surface current density) or through an infinitesimal area element (for a volume current density). Let's integrate the latter over a surface to obtain the total current passing through that surface:

$$\mathrm{I}_{\mathcal{S}} = \int_{\mathcal{S}} da\, \hat{n}(\vec{r}) \cdot \vec{J}(\vec{r}) \tag{5.16}$$

If we take $\mathcal{S}$ to be a closed surface, we may apply the divergence theorem to the above:

$$\oint_{\mathcal{S}} da\, \hat{n}(\vec{r}) \cdot \vec{J}(\vec{r}) = \int_{\mathcal{V}(\mathcal{S})} d\tau\, \vec{\nabla} \cdot \vec{J}(\vec{r}) \tag{5.17}$$

where $\mathcal{V}(\mathcal{S})$ is the volume enclosed by $\mathcal{S}$. By conservation of charge, the current is just the time derivative of the charge enclosed by $\mathcal{S}$, with the sign such that if a positive current is exiting $\mathcal{S}$, then the charge enclosed must be decreasing, assuming that the surface itself is time-independent. With this, we have

$$\mathrm{I}_{\mathcal{S}} = -\frac{d}{dt}\, Q_{\mathcal{V}(\mathcal{S})} = -\frac{d}{dt} \int_{\mathcal{V}(\mathcal{S})} d\tau\, \rho(\vec{r}) = -\int_{\mathcal{V}(\mathcal{S})} d\tau\, \frac{\partial \rho(\vec{r})}{\partial t} \tag{5.18}$$

Thus, we have

$$\int_{\mathcal{V}(\mathcal{S})} d\tau\, \vec{\nabla} \cdot \vec{J}(\vec{r}) = -\int_{\mathcal{V}(\mathcal{S})} d\tau\, \frac{\partial \rho(\vec{r})}{\partial t} \tag{5.19}$$

Since the surface $\mathcal{S}$ is arbitrary, it must hold that the integrands are equal everywhere:

$$\boxed{\vec{\nabla} \cdot \vec{J}(\vec{r}) = -\frac{\partial \rho(\vec{r})}{\partial t}} \tag{5.20}$$

This is the *continuity equation* and is effectively the differential version of conservation of charge.

With this equation, we can define our *steady-state assumption* more mathematically: it corresponds to $\partial\rho/\partial t = 0$, which then implies $\vec{\nabla} \cdot \vec{J} = 0$. The interpretation is that the charge density at any point cannot change with time, which implies that the net current flow into or out of any point vanishes.

Fields of and Magnetic Forces between Currents

Biot-Savart Law

For a *steady-state current distribution* — one in which the current densities are time-independent — it is an empirical observation, validated by the Lorentz force that moving charges or currents experience, that the magnetic field at $\vec{r}$ due to the current distribution is given by

$$\boxed{\vec{B}(\vec{r}) = \frac{\mu_o}{4\,\pi}\int d\ell'\,\frac{\vec{\mathrm{I}}(\vec{r}\,')\times(\vec{r}-\vec{r}\,')}{|\vec{r}-\vec{r}\,'|^3} = \frac{\mu_o}{4\,\pi}\,\mathrm{I}\int\frac{d\vec{\ell}\,'(\vec{r}\,')\times(\vec{r}-\vec{r}\,')}{|\vec{r}-\vec{r}\,'|^3}} \tag{5.21}$$

$\mu_o = 4\,\pi\,\times\,10^{-7}\ \mathrm{N\ A^{-2}}$ is the *permeability of free space*. The magnetic field carries units of *teslas*, $\mathrm{T} = \mathrm{N}/(\mathrm{A}\cdot\mathrm{m})$. The Biot-Savart Law is the analogue in magnetostatics of Coulomb's Law in electrostatics, and it has the same $1/r^2$ dependence.

You are well aware of the result that the field of a straight wire along the z-axis carrying current I at a transverse distance s from the wire is

$$\vec{B}(\vec{r}) = \frac{\mu_o}{2\,\pi}\,\frac{\mathrm{I}}{s}\,\widehat{\phi} \tag{5.22}$$

where $\widehat{\phi}$ is the azimuthal unit vector in cylindrical coordinates. The field forms circles around the wire with orientation set by the right-hand rule. This is derived in Griffiths Example 5.5, which we will not repeat here since you saw it in Ph1c.

Force between Two Current-Carrying Wires

We can combine the Lorentz Force Law and the Biot-Savart Law to calculate the force between two current-carrying wires; this force is the empirical basis for magnetostatics, as it is much easier to measure the force between two wires than it is to create ideal test charges and measure their motion in the magnetic field of a wire. We just plug the Biot-Savart Law into the Lorentz Force Law for a line current distribution, Equation 5.6, to find the force on the first wire due to the field of the second wire:

$$\vec{F}_{mag} = \mathrm{I}_1 \int_{\mathcal{C}_1} d\vec{\ell} \times \vec{B}(\vec{r}) \tag{5.23}$$

$$= \frac{\mu_o}{4\,\pi}\,\mathrm{I}_1\mathrm{I}_2 \int_{\mathcal{C}_1}\int_{\mathcal{C}_2} \frac{d\vec{\ell}(\vec{r}) \times \left[d\vec{\ell}\,'(\vec{r}\,') \times (\vec{r}-\vec{r}\,')\right]}{|\vec{r}-\vec{r}\,'|^3} \tag{5.24}$$

Consider the special case of both wires running parallel to the z axis separated by $s\,\widehat{s}$ in the xy-plane, with the first wire on the z-axis itself. Then $d\vec{\ell} = \widehat{z}\,dz$, $d\vec{\ell}\,' = \widehat{z}\,dz'$, $\vec{r} = z\,\widehat{z}$, $\vec{r}\,' = s\,\widehat{s} + z'\,\widehat{z}$.

Therefore,

$$d\vec{\ell}(\vec{r}) \times \left[d\vec{\ell}\,'(\vec{r}\,') \times (\vec{r} - \vec{r}\,') \right] = dz\, dz'\, \widehat{z} \times \left[\widehat{z} \times \left((z - z')\, \widehat{z} - s\, \widehat{s} \right) \right] \tag{5.25}$$

$$= dz\, dz'\, s\, \widehat{s} \tag{5.26}$$

$$\text{and} \qquad |\vec{r} - \vec{r}\,'|^3 = \left[(z - z')^2 + s^2 \right]^{3/2} \tag{5.27}$$

Thus,

$$\vec{F}_{mag} = \frac{\mu_o}{4\,\pi}\, \mathrm{I}_1 \mathrm{I}_2\, s\, \widehat{s} \int_{-\infty}^{\infty} dz \int_{-\infty}^{\infty} dz' \, \frac{1}{\left[(z - z')^2 + s^2 \right]^{3/2}} \tag{5.28}$$

$$= \frac{\mu_o}{4\,\pi}\, \mathrm{I}_1 \mathrm{I}_2\, s\, \widehat{s} \int_{-\infty}^{\infty} dz \left[\frac{z' - z}{s^2 \left[(z - z')^2 + s^2 \right]^{1/2}} \right] \Bigg|_{-\infty}^{\infty} \tag{5.29}$$

$$= \frac{\mu_o}{4\,\pi}\, \mathrm{I}_1 \mathrm{I}_2\, s\, \widehat{s} \int_{-\infty}^{\infty} dz \, \frac{2}{s^2} = \frac{\mu_o}{2\,\pi}\, \frac{\mathrm{I}_1 \mathrm{I}_2}{s}\, \widehat{s} \int_{-\infty}^{\infty} dz \tag{5.30}$$

where we did the integral using the trigonometric substitution $z' - z = s \tan\theta$. The total force is infinite, but we can abstract out of the above expression the force per unit length on the first wire, which is attractive (pointing towards the second wire) if the currents flow in the same direction:

$$\vec{f}_{mag} = \frac{\mu_o}{2\,\pi}\, \frac{\mathrm{I}_1 \mathrm{I}_2}{s}\, \widehat{s} \tag{5.31}$$

General Expressions for Fields due to Current Densities

The obvious generalizations of the Biot-Savart Law are

$$\boxed{\vec{B}(\vec{r}) = \frac{\mu_o}{4\,\pi}\int da'\,\frac{\vec{K}(\vec{r}\,')\times(\vec{r}-\vec{r}\,')}{|\vec{r}-\vec{r}\,'|^3} \qquad \vec{B}(\vec{r}) = \frac{\mu_o}{4\,\pi}\int d\tau'\,\frac{\vec{J}(\vec{r}\,')\times(\vec{r}-\vec{r}\,')}{|\vec{r}-\vec{r}\,'|^3}} \tag{5.32}$$

Griffiths notes that a line current distribution is the lowest-dimensional current distribution one can have because the zero-dimensional version — a point charge moving with velocity $\vec{v}$ — does not constitute a steady-state current: the charge passing a given point in space is time-dependent.

As with the Lorentz Force Law, it should also be clear that one could consider the volume version to be the fundamental statement of the Biot-Savart Law and one can derive the lower-dimensional versions by including delta functions in the definition of $\vec{J}$. This does not apply to a reduction to zero dimensionality, as noted above.

There are good examples of the use of the Biot-Savart Law in Griffiths. Again, these are at the level of Ph1c, so we do not spend time in lecture on them.

Another Form for the Biot-Savart Law

We begin by using Equation 2.53 to rewrite the Biot-Savart Law expression for the magnetic field:

$$\vec{B}(\vec{r}) = \frac{\mu_o}{4\,\pi}\int_{\mathcal{V}} d\tau'\,\frac{\vec{J}(\vec{r}\,')\times(\vec{r}-\vec{r}\,')}{|\vec{r}-\vec{r}\,'|^3} = -\frac{\mu_o}{4\,\pi}\int_{\mathcal{V}} d\tau'\,\vec{J}(\vec{r}\,')\times\vec{\nabla}_{\vec{r}}\left(\frac{1}{|\vec{r}-\vec{r}\,'|}\right) \quad (5.33)$$

We use one of the product rules for the curl, $\vec{\nabla}\times(f\vec{a}) = f(\vec{\nabla}\times\vec{a}) - \vec{a}\times(\vec{\nabla}f)$, and notice that $\vec{\nabla}_{\vec{r}}\times\vec{J}(\vec{r}\,') = 0$ because $\vec{J}(\vec{r}\,')$ is a function of $\vec{r}\,'$ while $\vec{\nabla}_{\vec{r}}$ is with respect to $\vec{r}$, to obtain

$$\boxed{\vec{B}(\vec{r}) = \vec{\nabla}\times\frac{\mu_o}{4\,\pi}\int_{\mathcal{V}} d\tau'\,\frac{\vec{J}(\vec{r}\,')}{|\vec{r}-\vec{r}\,'|}} \quad (5.34)$$

where we have brought $\vec{\nabla}_{\vec{r}}$ outside the integral over $\vec{r}\,'$ because it acts with respect to $\vec{r}$. We also dropped the $\vec{r}$ subscript since now, being outside the integral, it must act only on $\vec{r}$. This form is obviously suggestive of the idea of $\vec{B}$ being derived from a vector potential, which we will return to shortly.

We note that, while our derivation of this equation did not appear to require any assumptions about the way the current behaves at infinity, we will see later that the steady-state assumption does imply the net current through any sphere must vanish.

Lecture 15:

Magnetostatics II:
Ampere's Law
Divergence of B
Vector Potential
Uniqueness Theorem
Magnetostatic Scalar Potential

Date Revised: 2024/02/25 08:15
Date Given: 2024/02/25

Curl and Divergence of the Magnetic Field; Ampere's Law

Curl of the Magnetic Field

From the field of a current-carring wire, Equation 5.22, we get the clear impression that $\vec{B}$ has curl and that the curl is related to the current sourcing the field. Here, we explicitly calculate this curl from the Biot-Savart Law. Griffiths Section 5.3.2 provides one technique for this; we use Jackson's technique instead to avoid duplication.

We take the curl of Equation 5.34 and apply the $BAC - CAB$ rule for the triple vector product, $\vec{\nabla} \times (\vec{\nabla} \times \vec{a}) = \vec{\nabla}(\vec{\nabla} \cdot \vec{a}) - \nabla^2 \vec{a}$, writing the coordinate that $\vec{\nabla}$ acts on explicitly:

$$\vec{\nabla}_{\vec{r}} \times \vec{B}(\vec{r}) = \vec{\nabla}_{\vec{r}} \times \left[\vec{\nabla}_{\vec{r}} \times \frac{\mu_o}{4\,\pi} \int_{\mathcal{V}} d\tau' \, \frac{\vec{J}(\vec{r}\,')}{|\vec{r} - \vec{r}\,'|} \right] \tag{5.35}$$

$$= \frac{\mu_o}{4\,\pi} \left[\vec{\nabla}_{\vec{r}} \left(\vec{\nabla}_{\vec{r}} \cdot \int_{\mathcal{V}} d\tau' \, \left(\frac{\vec{J}(\vec{r}\,')}{|\vec{r} - \vec{r}\,'|} \right) \right) - \nabla^2_{\vec{r}} \int_{\mathcal{V}} d\tau' \, \left(\frac{\vec{J}(\vec{r}\,')}{|\vec{r} - \vec{r}\,'|} \right) \right] \tag{5.36}$$

$$= \frac{\mu_o}{4\,\pi} \left[\vec{\nabla}_{\vec{r}} \int_{\mathcal{V}} d\tau' \, \vec{\nabla}_{\vec{r}} \cdot \left(\frac{\vec{J}(\vec{r}\,')}{|\vec{r} - \vec{r}\,'|} \right) - \int_{\mathcal{V}} d\tau' \, \nabla^2_{\vec{r}} \left(\frac{\vec{J}(\vec{r}\,')}{|\vec{r} - \vec{r}\,'|} \right) \right] \tag{5.37}$$

We were able to bring $\vec{\nabla}_{\vec{r}}$ and $\nabla^2_{\vec{r}}$ inside the integrals because $\vec{\nabla}_{\vec{r}}$ is with respect to $\vec{r}$ and the integral is over $\vec{r}\,'$. Similarly, because $\vec{\nabla}_{\vec{r}}$ is with respect to $\vec{r}$ and $\vec{J}$ is a function of $\vec{r}\,'$, $\vec{J}$ passes through the divergence in the first term and the Laplacian in the second one, preserving the necessary dot product in the first term and the vectorial nature of the second term:

$$\vec{\nabla}_{\vec{r}} \times \vec{B}(\vec{r}) = \frac{\mu_o}{4\,\pi} \left[\vec{\nabla}_{\vec{r}} \int_{\mathcal{V}} d\tau' \, \vec{J}(\vec{r}\,') \cdot \vec{\nabla}_{\vec{r}} \left(\frac{1}{|\vec{r} - \vec{r}\,'|} \right) - \int_{\mathcal{V}} d\tau' \, \vec{J}(\vec{r}\,') \, \nabla^2_{\vec{r}} \left(\frac{1}{|\vec{r} - \vec{r}\,'|} \right) \right] \tag{5.38}$$

We know from electrostatics that

$$\vec{\nabla}_{\vec{r}}\left(\frac{1}{|\vec{r}-\vec{r}'|}\right) = -\vec{\nabla}_{\vec{r}'}\left(\frac{1}{|\vec{r}-\vec{r}'|}\right)$$
$$\nabla^2_{\vec{r}}\left(\frac{1}{|\vec{r}-\vec{r}'|}\right) = -4\,\pi\,\delta(\vec{r}-\vec{r}')$$

The first equation may seem surprising if one considers the exchange $\vec{r} \leftrightarrow \vec{r}'$, but one can see it is true by simply evaluating the gradient on both sides or by defining $\vec{s} = \vec{r} - \vec{r}'$ and applying the offset and inversion techniques we used in electrostatics. The second is Equation 3.35 with the exchange $\vec{r} \leftrightarrow \vec{r}'$ (where here there is no sign flip because the Laplacian is quadratic in the derivatives and the delta function is symmetric in its argument). Applying them, we obtain

$$\vec{\nabla} \times \vec{B}(\vec{r}) = \frac{\mu_o}{4\,\pi}\left[-\vec{\nabla}_{\vec{r}}\int_{\mathcal{V}} d\tau'\,\vec{J}(\vec{r}')\cdot\vec{\nabla}_{\vec{r}'}\left(\frac{1}{|\vec{r}-\vec{r}'|}\right) + 4\,\pi\int_{\mathcal{V}} d\tau'\,\vec{J}(\vec{r}')\,\delta(\vec{r}-\vec{r}')\right] \tag{5.39}$$

The second term just becomes $4\,\pi\,\vec{J}(\vec{r})$, yielding

$$\vec{\nabla} \times \vec{B}(\vec{r}) = \frac{\mu_o}{4\,\pi}\left[-\vec{\nabla}_{\vec{r}}\int_{\mathcal{V}} d\tau'\,\vec{J}(\vec{r}')\cdot\vec{\nabla}_{\vec{r}'}\left(\frac{1}{|\vec{r}-\vec{r}'|}\right)\right] + \mu_o\,\vec{J}(\vec{r}') \tag{5.40}$$

We can apply the product rule $\vec{\nabla} \cdot (f\vec{a}) = \vec{a} \cdot \vec{\nabla} f + f \vec{\nabla} \cdot \vec{a}$ to rewrite the first term:

$$\int_{\mathcal{V}} d\tau' \, \vec{J}(\vec{r}\,') \cdot \vec{\nabla}_{\vec{r}\,'} \left(\frac{1}{|\vec{r} - \vec{r}\,'|} \right) = \int_{\mathcal{V}} d\tau' \, \vec{\nabla}_{\vec{r}\,'} \cdot \left(\frac{\vec{J}(\vec{r}\,')}{|\vec{r} - \vec{r}\,'|} \right) - \int_{\mathcal{V}} d\tau' \, \frac{\vec{\nabla}_{\vec{r}\,'} \cdot \vec{J}(\vec{r}\,')}{|\vec{r} - \vec{r}\,'|} \tag{5.41}$$

$$= \oint_{\mathcal{S}(\mathcal{V})} da' \, \hat{n}(\vec{r}\,') \cdot \left(\frac{\vec{J}(\vec{r}\,')}{|\vec{r} - \vec{r}\,'|} \right) = 0 \tag{5.42}$$

The second term vanishes because $\vec{\nabla}_{\vec{r}\,'} \cdot \vec{J}(\vec{r}\,') = 0$ under the steady-state assumption by the continuity equation with $\partial\rho/\partial t = 0$. We used the divergence theorem to transform the first term into a surface integral, and then we take the surface to infinity. Assuming the currents are localized, the integrand vanishes on that surface, causing the first term to vanish. Thus, we obtain, *under the steady-state assumption*,

$$\boxed{\vec{\nabla} \times \vec{B}(\vec{r}) = \mu_o \, \vec{J}(\vec{r})} \tag{5.43}$$

This equation is the differential version of *Ampere's Law*, which we will return to shortly.

Let's discuss some subtleties in the above derivation connected to the vanishing of the $\vec{\nabla}(\vec{\nabla}\cdot\vec{a})$ term. There are two points to make:

- When we get to the definition of the vector potential $\vec{A}$, we will be able to interpret the vanishing of that term as implying $\vec{\nabla}\cdot\vec{A}=0$ *for the form of the vector potential implied by Equation 5.34.* $\vec{\nabla}\cdot\vec{A}$ *will not vanish for any other form of the vector potential that yields the same field.* Just keep this point in mind, we'll provide more explanation later.
- We assumed that the currents are localized (confined to a finite volume) to make the surface term vanish. This is not the minimal condition required. We only need the integral to vanish. If we let the surface go off to infinity while keeping the point $\vec{r}$ at which we want to know the field at finite distance from the origin, then $1/|\vec{r}-\vec{r}\,'|\to 1/r'$. Thus, we can also make the integral vanish by simply requiring that the net flux of $\vec{J}$ through a surface of radius r' vanishes. Griffiths notes this subtlety in Footnote 14 in §5.3.2. It explains how Ampere's Law works for an infinitely long wire: for any sphere at large radius, as much current flows in as out of that sphere, so the integral vanishes.

  Do we have to make this requirement? It may seem that we do not; we would just get a nonstandard Ampere's Law if we did not. But we do have to make it to be self-consistent with our steady-state assumption. If there were a net current through some sphere, then the charge contained in that volume would be changing with time, violating our steady-state assumption. This is the point we made in connection to Equation 5.34.

Divergence of the Magnetic Field

The vector identity $\vec{\nabla} \cdot (\vec{\nabla} \times \vec{a}) = 0$ combined with Equation 5.34 immediately implies

$$\boxed{\vec{\nabla} \cdot \vec{B}(\vec{r}) = 0} \tag{5.44}$$

The magnetic field has no divergence. This immediately implies *there are no magnetic point charges*: magnetic fields are sourced by currents only. It should be realized that this apparent fact is really an assumption inherent in the Biot-Savart Law. If we had added to the Biot-Savart Law a second term that looks like Coulomb's Law, due to magnetic monopoles, then the above divergence would have yielded that density of magnetic charge on the right side. It is *an empirical observation* that there are no magnetic monopoles, and hence we *assume* that magnetic fields are only sourced by currents via the Biot-Savart Law. That magnetic fields are sourced by currents at all is also *an empirical observation*; the Biot-Savart Law simplify codifies that observation.

General Thoughts on the Curl and Divergence of the Electric and Magnetic Field

Considering the corresponding expressions for electrostatics, we recognize that the electric field has divergence equal to the charge density because of the empirical observation of Coulomb's Law describing the electric field. It has a vanishing curl because of the empirical absence of a current that sources electric fields in the way that electric currents source magnetic fields; if there were a Biot-Savart-like term that added to Coulomb's Law, then the electric field would have curl. (In fact, when we consider time-varying fields, we will see that such a term comes into existence, proportional to $\partial\vec{B}/\partial t$.) We can in fact guess that, if magnetic monopoles existed, moving magnetic monopoles would generate an electric field in the same way that moving electric monopoles generate a magnetic field.

The key point in all of the above is that the nature of the divergence and the curl of the electric and magnetic fields reflect *empirical observations* about the way these fields are generated. These are not *derivable* results: they are inherent in the formulae we wrote down for the electric and magnetic fields, which themselves are based on *observations*.

We will see later that we can replace the assumption of Coulomb's Law and the Biot-Savart Law with assumptions about the potentials from which the electric and magnetic fields can be derived. But, again, we can only make those assumptions because they yields the correct empirical relations, Coulomb's Law and the Biot-Savart Law.

Integral form of Ampere's Law

We obtained the differential version of Ampere's Law above by taking the curl of the Biot-Savart Law for the magnetic field. We may obtain the integral form of Ampere's Law from it. We begin by integrating over an open surface $\mathcal{S}$ with normal $\hat{n}(\vec{r})$:

$$\int_{\mathcal{S}} da\, \hat{n}(\vec{r}) \cdot \left[\vec{\nabla} \times \vec{B}(\vec{r})\right] = \mu_o \int_{\mathcal{S}} da\, \hat{n} \cdot \vec{J}(\vec{r}) \tag{5.45}$$

The left side can be transformed using Stokes' Theorem into a line integral around the edge of $\mathcal{S}$, which we denote by the closed contour $\mathcal{C}(\mathcal{S})$, while the right side is just total current passing through $\mathcal{C}(\mathcal{S})$, I_{encl}:

$$\boxed{\oint_{\mathcal{C}(\mathcal{S})} d\vec{\ell} \cdot \vec{B}(\vec{r}) = \mu_o \mathrm{I}_{encl}} \tag{5.46}$$

yielding the integral version of Ampere's Law.

As before, there are a number of examples in Griffiths that are at the level of Ph1c, so we do not spend time on them here.

Magnetic Vector Potential

Form for the Magnetic Vector Potential

We saw (Equations 5.43 and 5.44) that the magnetic field has no divergence and has curl. You know from vector calculus (Griffiths §1.6) that this implies the magnetic field can be written purely as the curl of a *vector potential*. Equation 5.34 gave us its form:

$$\boxed{\vec{B}(\vec{r}) = \vec{\nabla} \times \vec{A}(\vec{r}) \qquad \vec{A}(\vec{r}) = \frac{\mu_o}{4\,\pi} \int_{\mathcal{V}} d\tau' \, \frac{\vec{J}(\vec{r}\,')}{|\vec{r} - \vec{r}\,'|}} \tag{5.47}$$

But this form, implied by the Biot-Savart Law, is not the only form. We had freedom with the electrostatic potential to add an offset. Here, we can add any curl-less function to $\vec{A}$ without affecting $\vec{B}$. The form above corresponds to the additional condition

$$\vec{\nabla} \cdot \vec{A}(\vec{r}) = 0 \tag{5.48}$$

If one tries to test this requirement on the above form for $\vec{A}$, one will find oneself doing the same manipulations needed to derieve Ampere's Law, Equation 5.43. In repeating those manipulations, *which is possible for this form of $\vec{A}$ only*, one sees that $\vec{\nabla} \cdot \vec{A} = 0$ is the representation of the steady-state assumption and that the net current through a surface of any radius vanishes (and also how the latter implies the former). For a different choice of $\vec{A}$ (and thus of $\vec{\nabla} \cdot \vec{A}$), the mathematical manifestation of this physical requirement will be different. In fact, it must be, because $\vec{\nabla} \cdot \vec{A} = 0$ is unique to this form.

Explicit Proof that $\vec{\nabla} \cdot \vec{A} = 0$ Can Always Be Obtained

It is interesting to prove "mechanically" that the choice $\vec{\nabla} \cdot \vec{A}$ is possible even if one, for some reason, started out with a form that did not satisfy this condition. Suppose one has a vector potential $\vec{A}_0$ that is not divergenceless. We need to add to it a function that makes the result divergenceless. For reasons we will see below, let's add a function $\vec{\nabla}\lambda(\vec{r})$:

$$\vec{A} = \vec{A}_0 + \vec{\nabla}\lambda \tag{5.49}$$

Then

$$\vec{\nabla} \cdot \vec{A} = \vec{\nabla} \cdot \vec{A}_0 + \nabla^2 \lambda \tag{5.50}$$

If we require the left side to vanish, then we have a version of Poisson's Equation:

$$\nabla^2 \lambda = -\vec{\nabla} \cdot \vec{A}_0 \tag{5.51}$$

One thus sees one of the motivations for the assumed form $\vec{\nabla}\lambda$.

Let's choose boundary conditions that place the boundary at infinity with the field falling off at infinity. For these boundary conditions, we know from Coulomb's Law that the solution to Poisson's Equation is

$$\lambda(\vec{r}) = \frac{1}{4\,\pi} \int_{\mathcal{V}} d\tau' \frac{\vec{\nabla} \cdot \vec{A}_0(\vec{r}')}{|\vec{r} - \vec{r}'|} \tag{5.52}$$

The vector calculus identity $\vec{\nabla} \times \vec{\nabla}\lambda = 0$ implies that $\vec{\nabla} \times \vec{A} = \vec{\nabla} \times \vec{A}_0$ and thus the magnetic field is the same for the two vector potentials (our second motivation for the choice to add $\vec{\nabla}\lambda$). We thus have an explicit formula for the term that has to be added to $\vec{A}_0$ so that the resulting form $\vec{A}$ is divergenceless while leaving the magnetic field unchanged.

The above explicit formula may not be valid if we assume different boundary conditions, but we know Poisson's Equation always has a solution, so we are guaranteed that the desired function $\lambda(\vec{r})$ exists.

Let us make a final point about how the above relates to the connection between $\vec{\nabla} \cdot \vec{A} = 0$ and the behavior of the currents at infinity. It is *not true* that starting with $\vec{\nabla} \cdot \vec{A}_0 \neq 0$ corresponds to a different physical assumption about the currents at infinity: changing $\vec{\nabla} \cdot \vec{A}$ has no effect on the fields and thus can have no effect on the currents. Our standard formula for $\vec{A}$ is only valid under the assumption $\vec{\nabla} \cdot \vec{A} = 0$, and so the relation between $\vec{\nabla} \cdot \vec{A}$ and the assumption about how the currents behave is only valid for that form. If one assumes a different form for $\vec{A}$, one that has $\vec{\nabla} \cdot \vec{A} \neq 0$, then taking its divergence will not necessarily result in the particular expressions that we encountered before in deriving the differential form of Ampere's Law, so the interpretation of $\vec{\nabla} \cdot \vec{A} = 0$ will be different, and the mathematical manifestation of the currents vanishing at infinity will also change. One benefit of the choice $\vec{\nabla} \cdot \vec{A} = 0$ is that this mathematical manifestation is simple.

Alternate Proof of the Form for the Magnetic Vector Potential

We can arrive at Equation 5.47 via a slightly different path, which makes uses of Ampere's Law and the same triple vector identity we used to prove Ampere's Law, $\vec{\nabla} \times (\vec{\nabla} \times \vec{a}) = \vec{\nabla}(\vec{\nabla} \cdot \vec{a}) - \nabla^2 \vec{a}$:

$$\text{Ampere's Law:} \qquad \vec{\nabla} \times (\vec{\nabla} \times \vec{A}) = \vec{\nabla} \times \vec{B} = \mu_o \vec{J} \tag{5.53}$$

$$\text{use vector identity:} \qquad \vec{\nabla}(\vec{\nabla} \cdot \vec{A}) - \nabla^2 \vec{A} = \mu_o \vec{J} \tag{5.54}$$

$$\text{set } \vec{\nabla} \cdot \vec{A} = 0: \qquad \nabla^2 \vec{A} = -\mu_o \vec{J} \tag{5.55}$$

Note that the vector components of $\vec{A}$ and $\vec{J}$ line up. Thus, the last equation is a component-by-component Poisson's Equation. Again, under the assumption that the currents are localized and for appropriate boundary conditions (as we assumed in providing the alternate version of the Biot-Savart Law that we previously used to define $\vec{A}$), we know the solution:

$$\boxed{\nabla^2 \vec{A}(\vec{r}) = -\mu_o \vec{J}(\vec{r}) \quad \overset{\text{localized currents}}{\Longleftrightarrow} \quad \vec{A}(\vec{r}) = \frac{\mu_o}{4\pi} \int_{\mathcal{V}} d\tau' \, \frac{\vec{J}(\vec{r}\,')}{|\vec{r} - \vec{r}\,'|}} \tag{5.56}$$

This is just Equation 5.47 again. Essentially, we can think of the three components of the current density as sourcing the three components of the vector potential in the same way that the electric charge density sources the electric potential.

The Vector Potential for Line and Surface Currents

We can consider the specific cases of line and surface current densities as volume current densities that include delta functions specifying the localization to a line or sheet. When one does the volume integral, the delta function reduces the three-dimensional integral over the volume to one- or two-dimensional integrals over a line or sheet, yielding:

$$\boxed{\vec{A}(\vec{r}) = \frac{\mu_o}{4\,\pi}\int_{\mathcal{C}} d\ell\,\frac{\vec{\mathrm{I}}(\vec{r}\,')}{|\vec{r}-\vec{r}\,'|} \qquad \vec{A}(\vec{r}) = \frac{\mu_o}{4\,\pi}\int_{\mathcal{S}} da'\,\frac{\vec{K}(\vec{r}\,')}{|\vec{r}-\vec{r}\,'|}} \tag{5.57}$$

Note that the units of the vector potential are unchanged: the change in the units of the current densities are canceled by the change in the units of the measure of integration.

Example 5.1: Spinning Sphere of Charge (Griffiths Example 5.11)

The calculation of the vector potential for a spinning spherical shell of charge is a straightforward application of the definition of the vector potential. The only complication is the vector arithmetic.

Center the sphere of radius R at the origin. Let the surface change density be σ and the angular rotation frequency be $\vec{\omega} = \omega\,\widehat{z}$. Then the surface current density is

$$\vec{K}(\vec{r}) = \sigma\,\vec{v}(\theta) = \sigma\,\vec{\omega}\times\vec{r} = \widehat{\phi}\,\sigma\,\omega\,R\,\sin\theta \tag{5.58}$$

We need to calculate

$$\vec{A}(\vec{r}) = \frac{\mu_o}{4\,\pi}\,R^2\int_0^{2\,\pi} d\phi'\int_0^{\pi} d\theta'\,\sin\theta'\,\frac{\sigma\,\omega\,\sin\theta'\,\widehat{\phi}}{|\vec{r}-\vec{r}\,'|} \tag{5.59}$$

$$= \frac{\mu_o}{4\,\pi}\,\sigma\,\omega\,R^3\int_0^{2\,\pi} d\phi'\int_0^{\pi} d\theta'\,\sin\theta'\,\frac{\sin\theta'\,\left(-\widehat{x}\,\sin\phi' + \widehat{y}\,\cos\phi'\right)}{|\vec{r}-\vec{r}\,'|} \tag{5.60}$$

(The R^2 out front is because an area integral, not just a solid angle integral, needs to be done.) This is done in Griffiths Example 5.11 via explicit integration. We will use a different technique benefiting from our knowledge of the spherical harmonics and the Addition Theorem.

We use Equation 3.169, the Spherical Harmonic Addition Theorem Corollary, which expands $|\vec{r}-\vec{r}\,'|^{-1}$ in terms of spherical harmonics, recognizing $|\vec{r}\,'| = R$ because the integral is over the sphere of radius R:

$$\frac{1}{|\vec{r}-\vec{r}\,'|} = 4\,\pi \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} \frac{1}{2\,\ell+1} \frac{r_<^{\ell}}{r_>^{\ell+1}}\, Y_{\ell m}^*(\theta\,',\phi\,')\, Y_{\ell m}(\theta,\phi) \tag{5.61}$$

Let's consider the $\widehat{x}$ piece of the above angular integral; the other term will be similar in spirit. We will write the numerator in terms of spherical harmonics and use the expansion. We abbreviate $\int_0^{2\,\pi} d\phi\,' \int_0^{\pi} d\theta\,' \sin\theta\,' = \int d\Omega'$ and recall $Y_{\ell,-m} = (-1)^m Y_{\ell,m}^*$. Applying these facts yields

$$\int d\Omega'\, \frac{\sin\theta\,'\,(-\sin\phi\,')}{|\vec{r}-\vec{r}\,'|} = \tag{5.62}$$

$$\int d\Omega'\, \sqrt{\frac{8\,\pi}{3}}\, \frac{Y_{1,1}(\theta\,',\phi\,') + Y_{1,-1}(\theta\,',\phi\,')}{2\,i} \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} \frac{4\,\pi}{2\,\ell+1} \frac{r_<^{\ell}}{r_>^{\ell+1}}\, Y_{\ell m}^*(\theta\,',\phi\,')\, Y_{\ell m}(\theta,\phi)$$

The integral over Ω' gives $\delta_{\ell,1}\delta_{m,1}$ and $\delta_{\ell,1}\delta_{m,-1}$, eliminating the sum and yielding

$$\int d\Omega' \, \frac{\sin\theta' \, (-\sin\phi')}{|\vec{r} - \vec{r}\,'|} = \frac{4\,\pi}{2\,i} \sqrt{\frac{8\,\pi}{3}} \, \frac{1}{3} \, \frac{r_<^\ell}{r_>^{\ell+1}} \, [Y_{1,1}(\theta,\phi) + Y_{1,-1}(\theta,\phi)] \tag{5.63}$$

$$= -\frac{4\,\pi}{3} \, \frac{r_<}{r_>^2} \, \sin\theta \, \sin\phi \tag{5.64}$$

where the $1/3$ came from $1/(2\,\ell + 1)$. We get back the same type of angular dependence, but with the $1/|\vec{r} - \vec{r}\,'|$ turned into the prefactor shown, which has the correct dimensions.

We can repeat the same kind of manipulation for the $\widehat{y}$ term, yielding

$$\int d\Omega' \, \frac{\sin\theta' \, (\cos\phi')}{|\vec{r} - \vec{r}\,'|} = \frac{4\,\pi}{3} \, \frac{r_<}{r_>^2} \, \sin\theta \, \cos\phi \tag{5.65}$$

Therefore,

$$\vec{A}(\vec{r}) = \frac{\mu_o}{4\,\pi} \, \sigma \, \omega \, R^3 \, \frac{4\,\pi}{3} \, \frac{r_<}{r_>^2} \, \sin\theta \, [-\widehat{x} \, \sin\phi + \widehat{y} \cos\phi] = \frac{\mu_o}{4\,\pi} \, \sigma \, \omega \, \frac{4\,\pi\,R^3}{3} \, \frac{r_<}{r_>^2} \, \sin\theta \, \widehat{\phi} \tag{5.66}$$

where $r_< = \min(r, R)$ and $r_> = \max(r, R)$. While the vector potential's r dependence change from r inside the sphere to $1/r^2$ outside the sphere, the direction and angular dependence are always $\widehat{\phi} \, \sin\theta$. One can see that $\vec{A}$ points in the same direction as $\vec{K}$.

Example 5.2: Solenoid (Griffiths Example 5.12)

The calculation of the vector potential for a solenoid, which is the equivalent of a spinning cylinder of charge if one ignores the small axial current contribution, is more interesting because one need not do it by explicit integration as done for the spinning sphere of charge. Instead, one can use some intuition along with the combination of Stokes' Theorem and the relation between $\vec{B}$ and $\vec{A}$:

$$\oint_{\mathcal{C}(\mathcal{S})} d\vec{\ell} \cdot \vec{A} = \int_{\mathcal{S}} da\, \widehat{n} \cdot \vec{\nabla} \times \vec{A} = \int_{\mathcal{S}} da\, \widehat{n} \cdot \vec{B} \tag{5.67}$$

The intuition part is to recognize that, because $\vec{B}$ is along the z-axis inside the solenoid and vanishing outside and because $\vec{A}$ "wraps around" $\vec{B}$, it is natural to assume $\vec{A}$ is along $\widehat{\phi}$. (This directionality can also be seen from the inherent coalignment of $\vec{A}$ and $\vec{J}$ from the integral formula for $\vec{A}$.) Then one can do the calculation in the same way as one applies the integral form of Ampere's Law, except that instead of current through a surface ("enclosed current"), we have enclosed magnetic flux, and, instead of a line integral of magnetic field around the edge of the surface, we have a line integral of vector potential. Please study the details in Griffiths, as a variant on this problem will be given in homework.

Uniqueness Theorem for Magnetic Fields

This is Griffiths Problem 5.56.

Just as we did for electric fields, we can show that, given a current distribution and a well-defined set of boundary conditions, the magnetic field obtained is unique. We assume that a current distribution $\vec{J}(\vec{r})$ in a volume $\mathcal{V}$ is specified. We will see later how specific we must be about the boundary conditions.

First, we need something analogous to the Green's Identities we used in the case of electrostatics. Using the vector identity $\vec{\nabla} \cdot (\vec{a} \times \vec{b}) = \vec{b} \cdot \vec{\nabla} \times \vec{a} - \vec{a} \cdot \vec{\nabla} \times \vec{b}$, letting $\vec{u}$ and $\vec{v}$ be two arbitrary vector fields, and applying the identity with $\vec{a} = \vec{u}$ and $\vec{b} = \vec{\nabla} \times \vec{v}$, we may write

$$\int_{\mathcal{V}} d\tau \, \vec{\nabla} \cdot (\vec{u} \times (\vec{\nabla} \times \vec{v})) = \int_{\mathcal{V}} d\tau \left[(\vec{\nabla} \times \vec{v}) \cdot (\vec{\nabla} \times \vec{u}) - \vec{u} \cdot (\vec{\nabla} \times (\vec{\nabla} \times \vec{v})) \right] \quad (5.68)$$

Since the expression on the left-hand side is a divergence, we may turn it into a surface integral using the divergence theorem:

$$\oint_{\mathcal{S}(\mathcal{V})} da \, \hat{n} \cdot (\vec{u} \times (\vec{\nabla} \times \vec{v})) = \int_{\mathcal{V}} d\tau \left[(\vec{\nabla} \times \vec{u}) \cdot (\vec{\nabla} \times \vec{v}) - \vec{u} \cdot (\vec{\nabla} \times (\vec{\nabla} \times \vec{v})) \right] \quad (5.69)$$

We will use this below.

Now, suppose that we have two different magnetic field configurations $\vec{B}_1 \neq \vec{B}_2$, derived from two different magnetic vector potentials $\vec{A}_1 \neq \vec{A}_2$, that both satisify Ampere's Law for the same current distribution: $\vec{\nabla} \times \vec{B}_1 = \vec{\nabla} \times \vec{B}_2 = \mu_o \vec{J}$. Let $\vec{A}_3 = \vec{A}_2 - \vec{A}_1$ and $\vec{B}_3 = \vec{B}_2 - \vec{B}_1$. We apply the above vector identity with $\vec{u} = \vec{v} = \vec{A}_3$:

$$\oint_{\mathcal{S}(\mathcal{V})} da\, \hat{n} \cdot (\vec{A}_3 \times (\vec{\nabla} \times \vec{A}_3)) = \int_{\mathcal{V}} d\tau \left[(\vec{\nabla} \times \vec{A}_3) \cdot (\vec{\nabla} \times \vec{A}_3) - \vec{A}_3 \cdot (\vec{\nabla} \times (\vec{\nabla} \times \vec{A}_3)) \right] \tag{5.70}$$

We have that $\vec{\nabla} \times (\vec{\nabla} \times \vec{A}_3) = \vec{\nabla} \times \vec{B}_3 = \vec{\nabla} \times \vec{B}_2 - \vec{\nabla} \times \vec{B}_1 = \mu_o(\vec{J} - \vec{J}) = 0$ by Ampere's Law and the assumption that both field configurations are sourced by the same current distribution, so the second term on the right side vanishes. Exchanging the two sides, plugging in $\vec{B}_3 = \vec{\nabla} \times \vec{A}_3$, and using the cyclic property of the triple scalar product, $\vec{a} \cdot (\vec{b} \times \vec{c}) = \vec{c} \cdot (\vec{a} \times \vec{b}) = \vec{b} \cdot (\vec{c} \times \vec{a})$, we have

$$\int_{\mathcal{V}} d\tau \left| \vec{B}_3 \right|^2 = \oint_{\mathcal{S}(\mathcal{V})} da\, \hat{n} \cdot (\vec{A}_3 \times \vec{B}_3) = \oint_{\mathcal{S}(\mathcal{V})} da\, \vec{B}_3 \cdot (\hat{n} \times \vec{A}_3) \tag{5.71}$$

$$= \oint_{\mathcal{S}(\mathcal{V})} da\, \vec{A}_3 \cdot (\vec{B}_3 \times \hat{n}) = - \oint_{\mathcal{S}(\mathcal{V})} da\, \vec{A}_3 \cdot (\hat{n} \times \vec{B}_3) \tag{5.72}$$

From the above equation, we can see what (minimal) boundary condition information we must have to obtain uniqueness of $\vec{B}$: we must have that, at any given point on the surface, $\vec{A}$, $\vec{B}$, $\hat{n} \times \vec{A}$, or $\hat{n} \times \vec{B}$ is specified. If this is true, then $\vec{A}_3 = \vec{A}_2 - \vec{A}_1 = 0$ where $\vec{A}$ is specified, $\vec{B}_3 = \vec{B}_2 - \vec{B}_1 = 0$ where $\vec{B}$ is specified, $\hat{n} \times \vec{A}_3 = \hat{n} \times (\vec{A}_2 - \vec{A}_1) = 0$ where $\hat{n} \times \vec{A}$ is specified, and $\hat{n} \times \vec{B}_3 = \hat{n} \times (\vec{B}_2 - \vec{B}_1) = 0$ where $\hat{n} \times \vec{B}$ is specified. Requiring one of these four conditions at every point on $\mathcal{S}(\mathcal{V})$ ensures the integrand on the right side vanishes at every point on $\mathcal{S}(\mathcal{V})$ and thus the right side vanishes. Since the integrand on the left side is nonnegative, it must therefore vanish everywhere: $\vec{B}_3 = 0$. Hence, $\vec{B}_1 = \vec{B}_2$ and the fields are identical and the field solution is unique.

Specifying $\vec{A}$ is like a Dirichlet boundary condition where we specify the electrostatic potential on the boundary, and specifying $\hat{n} \times \vec{B} = \hat{n} \times (\vec{\nabla} \times \vec{A})$ is a lot like a Neumann boundary condition where we specify the normal gradient of the electrostatic potential $\hat{n} \cdot \vec{\nabla} V$ (which is proportional to the normal component of the electric field, $\hat{n} \cdot \vec{E}$). In fact, we will see via Ampere's Law that this is equivalent to specifying the surface current density flowing on the boundary. The other two types of conditions, specifying $\hat{n} \times \vec{A}$ or specifying $\vec{B}$, have no obvious analogue.

Uniqueness of the Vector Potential?

We have already discussed how the $\vec{A}$ that generates a particular $\vec{B}$ is unique up to the gradient of an additional function if its divergence is left unspecified. The above theorem for the uniqueness of the magnetic field therefore now tells us that specification of $\vec{J}$ in the volume and of $\vec{A}$, $\vec{B}$, $\widehat{n} \times \vec{A}$, or $\widehat{n} \times \vec{B}$ on the boundary gives a vector potential that is unique up to the gradient of an additional function if its divergence is unspecified. But what do we need to know to completely determine the vector potential?

Obtaining a unique vector potential is the equivalent of being able to also know the λ function (up to an offset). We showed that λ satisfies Poisson's Equation with $\vec{\nabla} \cdot \vec{A}$ as the source, Equation 5.51. So, clearly, to obtain a unique $\vec{A}$, we would need to specify $\vec{\nabla} \cdot \vec{A}$. We also would need appropriate boundary conditions for this Poisson Equation. We may conclude from our proof of the uniquess of the scalar potential (up to an offset) that we must either specify λ or $\widehat{n} \cdot \vec{\nabla}\lambda$ on the boundary to obtain a unique λ (again, up to an offset) and thus a unique $\vec{A}$.

Which of the above conditions provide the necessary boundary condition on λ? Only specification of $\vec{A}$ on the boundary is certain to be sufficient. This gives $\vec{\nabla}\lambda$ and thus $\widehat{n} \cdot \vec{\nabla}\lambda$, a Neumann boundary condition for λ and thus sufficient to render λ unique.

We can see specifying $\hat{n} \times \vec{A}$ would only be sufficient in special cases. Doing so specifies $\hat{n} \times \vec{\nabla}\lambda$, which gives the component of $\vec{\nabla}\lambda$ tangent to the boundary. If the boundary is either at infinity or is a single, closed boundary, it seems likely one could then construct λ on the boundary by doing the line integral of $\hat{n} \times \vec{\nabla}\lambda$, much like one constructs the scalar potential from its gradient, the electric field. (It is ok that we would only know the component of $\vec{\nabla}\lambda$ tangent to the boundary, as $\hat{n} \cdot \vec{\nabla}\lambda$ will have zero dot product with the line element $d\vec{\ell}$ involved in the line integral.) As with the scalar potential, the offset is not important. However, if the boundary is not simply connected, then there is no way to connect λ on different pieces of the boundary without specifying its value on at least one point on each of those pieces. But we do not specify λ anywhere if we are given $\hat{n} \times \vec{A}$ and thus $\hat{n} \times \vec{\nabla}\lambda$ on the boundary. So specifying $\hat{n} \times \vec{A}$ (and $\vec{\nabla} \cdot \vec{A}$) is sufficient to make $\vec{A}$ unique only if the boundary is simply connected.

We can be assured that specifying $\vec{B}$ or $\hat{n} \times \vec{B}$ is entirely insufficient: because $\vec{B}$ is unaffected by λ, providing information about $\vec{B}$ cannot give us any information about λ.

Lastly, we remind the reader that, even if $\vec{A}$ is specified on the boundary, one also needs to know $\vec{\nabla} \cdot \vec{A}$ in the volume. Providing the former without the latter is equivalent to having a boundary condition but no differential equation to solve: the source term in the latter is unspecified.

The Magnetostatic Scalar Potential

If one considers current-free regions, then we have $\vec{\nabla} \times \vec{B} = 0$ and the magnetic field should be derivable from a scalar potential:

$$\vec{B}(\vec{r}) = -\vec{\nabla} U(\vec{r}) \tag{5.73}$$

One must take some care, though: in addition to being current-free, the region under consideration must be *simply connected*. Griffiths Problem 5.29 shows a situation where the current in a region may vanish but $\vec{\nabla} \times \vec{B} \neq 0$ because the region is not simply connected and the enclosed volume outside the region contains current.

With the above assumptions, and noting $\vec{\nabla} \cdot \vec{B} = 0$, we can infer that U satisfies Laplace's Equation:

$$\nabla^2 U(\vec{r}) = -\vec{\nabla} \cdot \vec{B}(\vec{r}) = 0 \tag{5.74}$$

Our usual assumption of simple boundary conditions — everything falls off to zero at infinity — yields a trivial result here, $U(\vec{r}) = 0$, so we must assume less trivial boundary conditions to obtain a nonzero U. We will return to the use of the magnetostatic scalar potential in connection with magnetically polarizable materials.

Lecture 16:

Magnetostatics III:
Boundary Conditions on Magnetic Field and Vector Potential
Magnetic Multipoles

Date Revised: 2024/02/27 06:00
Date Given: 2024/02/27

Boundary Conditions on Magnetic Field and Vector Potential

We will use techniques similar to those we used in determining the boundary conditions on the electric field. We will not immediately apply these conditions to boundary value problems for currents in vacuum because there are no nontrivial boundary-value problems of this type. That is because there is no way to directly set the vector potential, unlike for the electostatic potential. There is also no equivalent to the perfect conductor, which yields equipotential surfaces in electrostatics. One only has Neumann boundary conditions, with current densities on surfaces, from which one can calculate the field directly via the Biot-Savart Law rather than solving Laplace's or Poisson's Equation. We will find the boundary conditions more useful in the context of magnetically polarizable materials.

Boundary Conditions on the Magnetic Field

Recall that Gauss's Law, $\vec{\nabla} \cdot \vec{E} = \rho/\epsilon_o$, implied that the normal component of the electric field satisfied Equation 2.59

$$\hat{n}(\vec{r}) \cdot \left[\vec{E}_2(\vec{r}) - \vec{E}_1(\vec{r})\right] = \frac{1}{\epsilon_o}\,\sigma(\vec{r}) \tag{5.75}$$

Since $\vec{\nabla} \cdot \vec{B} = 0$, we can conclude by analogy that

$$\boxed{\hat{n}(\vec{r}) \cdot \left[\vec{B}_2(\vec{r}) - \vec{B}_1(\vec{r})\right] = 0} \tag{5.76}$$

That is, the normal component of the magnetic field is continuous at any boundary.

For the tangential component, we return to the derivation leading to Equation 2.61. In that case, we considered a contour $\mathcal{C}$ that consisted of two legs $\mathcal{C}_1$ and $\mathcal{C}_2$ parallel to the interface and to each other and two legs normal to the interface whose length would be shrunk to zero. We saw

$$\oint_{\mathcal{C}} d\vec{\ell} \cdot \vec{E}(\vec{r}) = -\int_{\mathcal{C}_1, \vec{r}_a - \widehat{n}(\vec{r})\frac{dz}{2}}^{\vec{r}_b - \widehat{n}(\vec{r})\frac{dz}{2}} \vec{E}_1(\vec{r}) \cdot d\vec{\ell} + \int_{\mathcal{C}_2, \vec{r}_a + \widehat{n}(\vec{r})\frac{dz}{2}}^{\vec{r}_b + \widehat{n}(\vec{r})\frac{dz}{2}} \vec{E}_2(\vec{r}) \cdot d\vec{\ell} \tag{5.77}$$

$$\stackrel{dz \to 0}{\longrightarrow} \int_{\mathcal{C}_2, \vec{r}_a}^{\vec{r}_b} \left[\vec{E}_2(\vec{r}) - \vec{E}_1(\vec{r})\right] \cdot d\vec{\ell} \tag{5.78}$$

where the ends of the loop are near $\vec{r}_a$ and $\vec{r}_b$, $\widehat{n}$ is the normal to the surface (parallel to the short legs of the loop), $\widehat{t}$ is the normal to the loop area, $\widehat{s} = \widehat{t} \times \widehat{n}$ is the unit vector parallel to the long legs of the loop, and ds is a line element along $\widehat{s}$. In the electric field case, the left side of the above expression vanished. In the case of the magnetic field, Ampere's Law tells us that it is the current enclosed flowing in the direction $\widehat{t}$. Therefore, the magnetic field version of the above equation is:

$$\mu_o \int_{\mathcal{C}_2} ds\, \widehat{t}(\vec{r}) \cdot \vec{K}(\vec{r}) = \int_{\mathcal{C}_2, \vec{r}_a}^{\vec{r}_b} \left[\vec{B}_2(\vec{r}) - \vec{B}_1(\vec{r})\right] \cdot d\vec{\ell} \tag{5.79}$$

where $\mathcal{C}_1 \to \mathcal{C}_2$ in the plane of the interface as $dz \to 0$. We neglect any volume current density passing through the area enclosed by the contour $\mathcal{C}$ because the integral of that volume current density vanishes as $dz \to 0$.

Since the contour $\mathcal{C}_2$ is arbitrary, the integrands must be equal

$$\left[\vec{B}_2(\vec{r}) - \vec{B}_1(\vec{r})\right] \cdot \widehat{s}(\vec{r}) = \mu_o \, \widehat{t}(\vec{r}) \cdot \vec{K}(\vec{r}) \tag{5.80}$$

Next, we use $\widehat{t} = \widehat{n} \times \widehat{s}$:

$$\left[\vec{B}_2(\vec{r}) - \vec{B}_1(\vec{r})\right] \cdot \widehat{s}(\vec{r}) = \mu_o \left[\widehat{n}(\vec{r}) \times \widehat{s}(\vec{r})\right] \cdot \vec{K}(\vec{r}) \tag{5.81}$$

Finally, using the cyclic nature of triple vector products ,
$\vec{a} \cdot (\vec{b} \times \vec{c}) = \vec{c} \cdot (\vec{a} \times \vec{b}) = \vec{b} \cdot (\vec{c} \times \vec{a})$:

$$\boxed{\left[\vec{B}_2(\vec{r}) - \vec{B}_1(\vec{r})\right] \cdot \widehat{s}(\vec{r}) = \mu_o \left[\vec{K}(\vec{r}) \times \widehat{n}(\vec{r})\right] \cdot \widehat{s}(\vec{r})} \tag{5.82}$$

Note that this condition holds for any $\widehat{s}$ tangential to the interface. To give some intuition, $\widehat{n} \times \vec{K}$ has the magnitude of $\vec{K}$ (because $\widehat{n} \perp \vec{K}$ always) but points in a direction perpendicular to $\vec{K}$ while still tangent to the interface. The sign is set by the cross-product right-hand rule.

We can combine the conditions on the normal and tangential components of $\vec{B}$ to obtain one compact expression for the boundary condition on the magnetic field. By the definition of the cross product, $\vec{K} \times \hat{n}$ is always perpendicular to $\hat{n}$ and thus has no component along $\hat{n}$. Therefore, the expression

$$\boxed{\vec{B}_2(\vec{r}) - \vec{B}_1(\vec{r}) = \mu_o \, \vec{K}(\vec{r}) \times \hat{n}(\vec{r})} \tag{5.83}$$

captures both boundary conditions: the projection of $\vec{B}$ normal to the interface (along $\hat{n}$) is continuous because the projection of the right side along that direction vanishes, and the projection of $\vec{B}$ along any $\hat{s}$ parallel to the interface can be discontinuous by the projection of $\mu_o \, \vec{K} \times \hat{n}$ along that direction. This is a very nice relation: given $\vec{K}$, it provides a way to calculate the change in the entire magnetic field across the interface, not just the change of a component.

We can rewrite the above in another way. Take the cross product of both sides with $\widehat{n}(\vec{r})$ from the left. The right side becomes a triple vector product, which we can rewrite using the $BAC - CAB$ rule, $\vec{a} \times (\vec{b} \times \vec{c}) = \vec{b}(\vec{a} \cdot \vec{c}) - \vec{c}(\vec{a} \cdot \vec{b})$. The second term has $\widehat{n} \cdot \vec{K}$, which vanishes, while the first term has $\widehat{n} \cdot \widehat{n} = 1$. Thus, we have

$$\boxed{\widehat{n}(\vec{r}) \times \left[\vec{B}_2(\vec{r}) - \vec{B}_1(\vec{r})\right] = \mu_o \, \vec{K}(\vec{r})} \tag{5.84}$$

The earlier form is more useful when $\vec{K}$ is specified, and the second form would more easily yield $\vec{K}$ if the fields are specified. Note, however, that this form does not preserve the information about the normal component of $\vec{B}$ because the contribution of that component to the left side vanishes.

Boundary Conditions on the Vector Potential

As one might expect by analogy to the electrostatic case, the vector potential itself has to be continuous across a boundary:

$$\boxed{\vec{A}_2(\vec{r}) - \vec{A}_1(\vec{r}) = 0} \tag{5.85}$$

This is seen easily:

- We have chosen the divergence of $\vec{A}$ to vanish, so the normal component of $\vec{A}$ must be continuous, just as we found the normal component of $\vec{B}$ is continuous for the same reason.
- The curl of $\vec{A}$ does not vanish, $\vec{\nabla} \times \vec{A} = \vec{B}$. This implies the line integral of $\vec{A}$ around the contour $\mathcal{C}$ used above is nonzero and equals $\Phi_{\mathcal{S}(\mathcal{C})} = \int_{\mathcal{S}(\mathcal{C})} da\, \hat{n} \cdot \vec{B}$, the *magnetic flux* of $\vec{B}$ through the surface $\mathcal{S}(\mathcal{C})$ defined by $\mathcal{C}$. But, as the area of the contour is shrunk to zero, the magnetic flux vanishes via an argument similar to the one we used to show that the flux of the electric field always goes to zero as the area through which it is calculated goes to zero: while the field can be quite singular ($1/r^2$), there are always cancellations that cause the flux to vanish. Therefore, the tangential component of $\vec{A}$ is also continuous.

While the vector potential itself is continuous, its derivatives are not necessarily continuous because its derivatives are related to $\vec{B}$, which is not necessarily continuous. Evaluating these discontinuities is a bit harder than in the case of the electric potential because the derivatives are not related in a trivial component-by-component way to the field. We need an expression involving second derivatives of $\vec{A}$ if we want to obtain boundary conditions on the first derivatives of $\vec{A}$. Let's use Equation 5.56:

$$\nabla^2 \vec{A}(\vec{r}) = -\mu_o \vec{J}(\vec{r}) \tag{5.86}$$

Consider a projection of this equation in Cartesian coordinates by taking the dot product with a Cartesian unit vector on the left and then passing it through the Laplacian, rewritten so the divergence is clear:

$$\vec{\nabla} \cdot \vec{\nabla} \left(\widehat{x} \cdot \vec{A}(\vec{r}) \right) = -\mu_o \, \widehat{x} \cdot \vec{J}(\vec{r}) \tag{5.87}$$

We have used Cartesian coordinates rather than a coordinate system using $\widehat{n}$, $\widehat{t}$, and $\widehat{s}$ because the latter vary in direction depending on where one is on the surface; their derivatives do not vanish, so we would not have been able to pull them inside the Laplacian as we did with $\widehat{x}$.

Given the above, we now apply the same kind of geometry we used to derive the boundary condition on the normal component of $\vec{E}$. That yields

$$\widehat{n} \cdot \left[\vec{\nabla} \left(\widehat{x} \cdot \vec{A}_2(\vec{r}) \right) - \vec{\nabla} \left(\widehat{x} \cdot \vec{A}_1(\vec{r}) \right) \right] = -\mu_o \, \widehat{x} \cdot \vec{K}(\vec{r}) \tag{5.88}$$

$$\widehat{n} \cdot \vec{\nabla} \left[\widehat{x} \cdot \vec{A}_2(\vec{r}) - \widehat{x} \cdot \vec{A}_1(\vec{r}) \right] = \tag{5.89}$$

where $\widehat{x} \cdot \vec{K}$ is what is left of $\widehat{x} \cdot \vec{J}$ as the Gaussian volume used in that proof shrinks to zero thickness in the direction normal to the interface, just as ρ reduced to σ in the case of the electric field.

The above argument holds for the $\widehat{y}$ and $\widehat{z}$ projections of $\vec{A}$ and $\vec{K}$ also, so we may combine them to obtain

$$\boxed{\widehat{n} \cdot \vec{\nabla} \left[\vec{A}_2(\vec{r}) - \vec{A}_1(\vec{r}) \right] = -\mu_o \, \vec{K}(\vec{r})} \tag{5.90}$$

Thus, we see that the normal derivative of each component of the vector potential has a discontinuity set by the surface current density in the direction of that component of the vector potential. This is a lot like the discontinuity in the normal gradient of the electric potential being determined by the surface charge density at the boundary.

We may derive, from the above, conditions in the normal and tangential directions by recognizing that

$$\left(\widehat{n}\cdot\vec{\nabla}\right)\widehat{n}=0 \qquad\qquad \left(\widehat{n}\cdot\vec{\nabla}\right)\widehat{s}=0 \tag{5.91}$$

These relations should be intuitively obvious: the direction of $\widehat{n}$, $\widehat{s}$, and $\widehat{t}$ change as one moves transversely along the surface (along $\widehat{s}$ or $\widehat{t}$), but they simply are not defined off the surface and thus they can have no derivative in that direction. This implies that the normal derivative of the normal component of $\vec{A}$ has no discontinuity since there can be no surface current in that direction:

$$\boxed{\widehat{n}\cdot\vec{\nabla}\left\{\widehat{n}\cdot\left[\vec{A}_2(\vec{r})-\vec{A}_1(\vec{r})\right]\right\}=0} \tag{5.92}$$

It also implies that the normal gradient of the vector potential in a particular direction parallel to the interface changes by the surface current density in that direction:

$$\boxed{\widehat{n}\cdot\vec{\nabla}\left\{\widehat{s}\cdot\left[\vec{A}_2(\vec{r})-\vec{A}_1(\vec{r})\right]\right\}=-\mu_o\,\widehat{s}\cdot\vec{K}(\vec{r})} \tag{5.93}$$

Next, let's consider the tangential derivatives of the vector potential. Here, we use the vector identity

$$\vec{\nabla} \times \vec{\nabla} \vec{A}(\vec{r}) = 0 \tag{5.94}$$

where again we consider each component of $\vec{A}$ as a scalar function and the above equation holds for all three components. If we again project by Cartesian components; *e.g.*

$$\vec{\nabla} \times \vec{\nabla} \left(\widehat{x} \cdot \vec{A}(\vec{r}) \right) = 0 \tag{5.95}$$

then we can apply the same type of argument as we applied for calculating the boundary condition on the tangential components of $\vec{E}$, which in this case yields

$$\widehat{s} \cdot \left[\vec{\nabla} \left(\widehat{x} \cdot \vec{A}_2(\vec{r}) \right) - \vec{\nabla} \left(\widehat{x} \cdot \vec{A}_1(\vec{r}) \right) \right] = 0 \tag{5.96}$$

$$\widehat{s} \cdot \vec{\nabla} \left[\widehat{x} \cdot \vec{A}_2(\vec{r}) - \widehat{x} \cdot \vec{A}_1(\vec{r}) \right] = \tag{5.97}$$

Since the argument again generalizes to any Cartesian component, we may combine the three expressions to obtain

$$\boxed{\widehat{s} \cdot \vec{\nabla} \left[\vec{A}_2(\vec{r}) - \vec{A}_1(\vec{r}) \right] = 0} \tag{5.98}$$

for any $\widehat{s}$ parallel to the interface: the tangential derivatives of $\vec{A}$ are continuous across an interface.

Magnetic Multipoles

Derivation of Magnetic Multipole Expansion

Since the vector potential is sourced by the current distribution in a manner similar to the way the charge distribution sources the electric potential, it is natural to develop the same multipole expansion. We follow Jackson for the sake of generality and variety; you can of course read the derivation in Griffiths, too. *We continue to make the steady-state assumption, and now we also make the assumption the currents are localized.* We start with the equation for the vector potential in terms of the current distribution:

$$\vec{A}(\vec{r}) = \frac{\mu_o}{4\,\pi} \int_{\mathcal{V}} d\tau' \, \frac{\vec{J}(\vec{r}\,')}{|\vec{r} - \vec{r}\,'|} \tag{5.99}$$

We recall Equation 3.149:

$$\frac{1}{|\vec{r} - \vec{r}\,'|} = \sum_{\ell=0}^{\infty} \frac{r_<^\ell}{r_>^{\ell+1}} \, P_\ell(\cos\gamma) \tag{5.100}$$

where $r_<$ and $r_>$ and the smaller and larger of r and $r\,'$.

As with the multipole expansion for the electrostatic potential, we will take $r \gg r'$: we want to know what the potential looks like far away from the current distribution. Therefore, $r_< = r'$ and $r_> = r$:

$$\vec{A}(\vec{r}) = \frac{\mu_o}{4\pi} \int_{\mathcal{V}} d\tau' \, \vec{J}(\vec{r}') \sum_{\ell=0}^{\infty} \frac{(r')^\ell}{r^{\ell+1}} P_\ell(\cos\gamma) \tag{5.101}$$

where $\cos\gamma = \hat{r} \cdot \hat{r}'$ is the angle between the two vectors.

There is a common $1/r$ we can factor out, leaving

$$\vec{A}(\vec{r}) = \frac{\mu_o}{4\pi} \frac{1}{r} \sum_{\ell=0}^{\infty} \frac{1}{r^\ell} \int_{\mathcal{V}} d\tau' \, \vec{J}(\vec{r}') \, (r')^\ell \, P_\ell(\cos\gamma) \tag{5.102}$$

Now, consider the first term, which is just the volume integral of the current density. Under the steady-state assumption, it is intuitively clear this integral must vanish. To prove this explicitly, we first use the vector identity $\vec{\nabla} \cdot (f\vec{a}) = f\vec{\nabla} \cdot \vec{a} + \vec{a}\vec{\nabla} f$ with $\vec{a} = \vec{J}$ and $f = r_i$ any of the Cartesian coordinates:

$$\vec{\nabla} \cdot (r_i \vec{J}) = r_i \vec{\nabla} \cdot \vec{J} + \vec{J} \cdot \vec{\nabla} r_i = 0 + \sum_{j=1}^{3} J_j \frac{\partial}{\partial r_j} r_i = \sum_{j=1}^{3} J_j \delta_{ij} = J_i \tag{5.103}$$

where the first term vanishes because of the steady-state assumption and so continuity implies $\vec{\nabla} \cdot \vec{J} = 0$. With this, we can compute the integral using the divergence theorem:

$$\int_{\mathcal{V}} d\tau' \, J_i(\vec{r}\,') = \int_{\mathcal{V}} d\tau' \, \vec{\nabla}' \cdot \left[r_i' \vec{J}(\vec{r}\,') \right] = \oint_{\mathcal{S}(\mathcal{V})} da' \, \hat{n}(\vec{r}\,') \cdot \left[r_i' \vec{J}(\vec{r}\,') \right] = 0 \tag{5.104}$$

where the surface integral in the last term vanishes because the current distribution is localized.

So, we are left with

$$\boxed{\vec{A}(\vec{r}) = \frac{\mu_o}{4\,\pi}\,\frac{1}{r}\sum_{\ell=1}^{\infty}\frac{1}{r^{\ell}}\int_{\mathcal{V}} d\tau'\,\vec{J}(\vec{r}\,')\,\left(r'\right)^{\ell}\,P_{\ell}(\cos\gamma)} \tag{5.105}$$

This is the *multipole expansion* of the vector potential of the current distribution. As with the multipole expansion of the electric potential, one can see that the successive terms fall off as successively higher powers of $1/r$.

It makes sense that there is no monopole term because $\vec{\nabla}\cdot\vec{B} = 0$: if there were a way to make a current distribution look like a monopole from far away, then one would have a field configuration with a nonzero Gauss's law integral of magnetic flux through a closed surface containing the current distribution, which is not allowed by $\vec{\nabla}\cdot\vec{B} = 0$.

The Magnetic Dipole Term

Let's consider the first nonzero term in more detail, which we subscript with a $_2$ because it will look like the electric dipole potential, and let's expand $\vec{J}$ in terms of its components so it is easier to work with:

$$\vec{A}_2(\vec{r}) = \frac{\mu_o}{4\,\pi}\,\frac{1}{r^2}\int_{\mathcal{V}} d\tau'\,\vec{J}(\vec{r}\,')\,r'\,P_2(\cos\gamma) = \frac{\mu_o}{4\,\pi}\,\frac{1}{r^3}\int_{\mathcal{V}} d\tau'\,\vec{J}(\vec{r}\,')\,\vec{r}\cdot\vec{r}\,' \tag{5.106}$$

$$= \frac{\mu_o}{4\,\pi}\,\frac{1}{r^3}\sum_{i,j=1}^{3}\widehat{r_i}\int_{\mathcal{V}} d\tau'\,J_i(\vec{r}\,')\,r_j\,r_j' \tag{5.107}$$

We must first prove an identity. We start with the same vector identity as before, now with $f = r_i\,r_j$ and $\vec{a} = \vec{J}$:

$$\vec{\nabla}\cdot(r_i\,r_j\,\vec{J}) = r_i\,r_j\,\vec{\nabla}\cdot\vec{J} + \vec{J}\cdot\vec{\nabla}(r_i\,r_j) = 0 + r_j\,\vec{J}\cdot\vec{\nabla}r_i + r_i\,\vec{J}\cdot\vec{\nabla}r_j \tag{5.108}$$

$$= r_j\,J_i + r_i\,J_j \tag{5.109}$$

where we have again used $\vec{\nabla}\cdot\vec{J} = 0$. We apply the same technique of integrating over volume and turning the left side into a surface term that vanishes, so we are left with

$$\int_{\mathcal{V}} d\tau'\,\left[r_i'\,J_j(\vec{r}\,') + r_j'\,J_i(\vec{r}\,')\right] = 0 \tag{5.110}$$

We can use this identity to rewrite the $\vec{A}_2$ term as:

$$\vec{A}_2(\vec{r}) = \frac{\mu_o}{4\,\pi}\,\frac{1}{r^3}\sum_{i,j=1}^{3}\widehat{r}_i\,r_j\int_{\mathcal{V}}d\tau'\,\frac{1}{2}\left[J_i(\vec{r}\,')\,r_j{}' - J_j(\vec{r}\,')\,r_i{}'\right] \tag{5.111}$$

where we split out half of the $J_i\,r_j'$ factor and used the identity to exchange the indices. You have learned in Ph106a and hopefully elsewhere that the cross-product can be written

$$(\vec{a}\times\vec{b})_k = \sum_{m,n=1}^{3}\epsilon_{kmn}\,a_m\,b_n \quad \text{with} \quad \epsilon_{kmn} = \begin{cases} 1 & \text{for cyclic index permutations} \\ -1 & \text{for anticyclic index permutations} \\ 0 & \text{when any two indices are identical} \end{cases} \tag{5.112}$$

where ϵ_{kmn} is the *Levi-Civita* symbol. Multiplying this definition by ϵ_{ijk} and summing over k gives

$$\sum_{k=1}^{3}\epsilon_{ijk}(\vec{a}\times\vec{b})_k = \sum_{k,m,n=1}^{3}\epsilon_{ijk}\,\epsilon_{kmn}\,a_m\,b_n = \sum_{k,m,n=1}^{3}\epsilon_{kij}\,\epsilon_{kmn}\,a_m\,b_n \tag{5.113}$$

There is an identity for the Levi-Civita symbol

$$\sum_{k=1}^{3} \epsilon_{kij}\, \epsilon_{kmn} = \delta_{im}\, \delta_{jn} - \delta_{in}\, \delta_{jm} \tag{5.114}$$

(this is the identity that produces the $BAC - CAB$ rule,
$\vec{a} \times (\vec{b} \times \vec{c}) = \vec{b}(\vec{a} \cdot \vec{c}) - \vec{c}(\vec{a} \cdot \vec{b})$) which lets us rewrite the above as

$$\sum_{k=1}^{3} \epsilon_{ijk} (\vec{a} \times \vec{b})_k = \sum_{m,n=1}^{3} a_m\, b_n \left(\delta_{im}\, \delta_{jn} - \delta_{in}\, \delta_{jm}\right) = a_i\, b_j - a_j\, b_i \tag{5.115}$$

This is exactly the expression we have inside the integral above.

Using the above identity, we may rewrite the $\vec{A}_2$ term as

$$\vec{A}_2(\vec{r}) = \frac{\mu_o}{4\,\pi}\, \frac{1}{r^3} \sum_{i,j,k=1}^{3} \widehat{r}_i\, r_j \int_{\mathcal{V}} d\tau'\, \frac{1}{2} \epsilon_{ijk} \left[J(\vec{r}\,') \times \vec{r}\,'\right]_k \tag{5.116}$$

$$= -\frac{\mu_o}{4\,\pi}\, \frac{1}{r^3}\, \frac{1}{2} \sum_{i}^{3} \widehat{r}_i \left\{ \vec{r} \times \int_{\mathcal{V}} d\tau'\, \left[\vec{r}\,' \times J(\vec{r}\,')\right] \right\}_i \tag{5.117}$$

$$= -\frac{\mu_o}{4\,\pi}\, \frac{1}{r^3}\, \frac{1}{2}\, \vec{r} \times \int_{\mathcal{V}} d\tau'\, \left[\vec{r}\,' \times J(\vec{r}\,')\right] \tag{5.118}$$

If we define the *magnetization density* $\vec{\mathcal{M}}(\vec{r})$ and the *magnetic dipole moment* $\vec{m}$ by

$$\boxed{\vec{\mathcal{M}}(\vec{r}) = \frac{1}{2}\,\vec{r} \times \vec{J}(\vec{r}) \qquad \text{and} \qquad \vec{m} = \int_{\mathcal{V}} d\tau'\,\vec{\mathcal{M}}(\vec{r}\,')} \tag{5.119}$$

then the $_2$ term is the *magnetic dipole vector potential*

$$\boxed{\vec{A}_2(\vec{r}) = \frac{\mu_o}{4\,\pi}\,\frac{\vec{m} \times \vec{r}}{r^3}} \tag{5.120}$$

Interestingly, this form has the same radial dependence as that of the electrostatic potential of a dipole, but the cross-product in the numerator differs from the dot product in the numerator of the electric dipole potential. However, because the magnetic field is obtained from the curl of the vector potential, while the electric field is obtained from the gradient of the electric potential, we will see that the two forms result in the same field configuration (up to normalization)!

Specialization of Magnetic Dipole Potential to a Current Loop

Now, let us consider a current loop. The only assumption we make is that the current throughout the loop is the same so that we can extract it from the integral. The volume integral reduces to a line integral over the loop contour:

$$\vec{A}_2(\vec{r}) = -\frac{\mu_o}{4\,\pi}\,\frac{1}{r^3}\,\frac{1}{2}\,\vec{r}\times\oint_{\mathcal{C}}\vec{r}\,'\times \mathrm{I}\,d\vec{\ell}\,'(\vec{r}\,') = -\frac{\mu_o}{4\,\pi}\,\frac{1}{r^3}\,\vec{r}\times \mathrm{I}\oint_{\mathcal{C}}\frac{\vec{r}\,'\times d\vec{\ell}\,'(\vec{r}\,')}{2} \tag{5.121}$$

The integral is now just a geometric quantity that has units of area. Separating out the magnetic moment, we have

$$\boxed{\vec{A}_2^{\,loop}(\vec{r}) = \frac{\mu_o}{4\,\pi}\,\frac{\vec{m}_{loop}\times\vec{r}}{r^3} \qquad \vec{m}_{loop} = \mathrm{I}\oint_{\mathcal{C}}\frac{\vec{r}\,'\times d\vec{\ell}\,'(\vec{r}\,')}{2}} \tag{5.122}$$

For the case of a loop confined to a plane that contains the origin, the quantity $\vec{r}' \times d\vec{\ell}'/2$ is the differential area element for the loop: it is the area of the triangle formed by $\vec{r}'$, the vector from the origin to a point on the loop, and $d\vec{\ell}'$, the line element tangent to the loop at $\vec{r}'$ and in the direction of the current, and this cross product has the standard right-hand-rule orientation. The integral thus calculates the area of the loop! Thus, for a planar loop, the above reduces to

$$\vec{A}_2(\vec{r}) = -\frac{\mu_o}{4\,\pi}\,\frac{1}{r^3}\,\vec{r}\times \mathrm{I}\,\widehat{n}\,a \tag{5.123}$$

where a is the loop area and $\widehat{n}$ is the normal to the loop with orientation defined by the current via the right-hand rule. Therefore, for this case, we have

$$\boxed{\vec{A}_2^{\,flat\ \ loop}(\vec{r}) = \frac{\mu_o}{4\,\pi}\,\frac{\vec{m}_{flat\ \ loop}\times\vec{r}}{r^3} \qquad \vec{m}_{flat\ \ loop} = \mathrm{I}\,\widehat{n}\,a} \tag{5.124}$$

Field of a Magnetic Dipole

If we let $\vec{m} = m\,\hat{z}$, then the dipole vector potential is

$$\vec{A}_2(\vec{r}) = \frac{\mu_o}{4\,\pi}\,\frac{m\,\sin\theta}{r^2}\,\hat{\phi} \equiv A_{2,\phi}\,\hat{\phi} \tag{5.125}$$

This form offers some intuition about how $\vec{A}_2(\vec{r})$ behaves. In general, $\vec{A}_2$ "circulates" around $\vec{m}$ using the right-hand rule in the same way that $\vec{A}$ "circulates" around $\vec{B}$ or $\vec{B}$ "circulates" around $\vec{J}$ using the right-hand rule. Since we are considering the distribution from far enough away that it is indistinguishable from a simple circular current loop in the xy-plane, the direction of $\vec{A}_2$ just results from the fact that $\vec{A}$ is the convolution of $\vec{J}$ with a scalar function: the direction of $\vec{A}$ always follows that of $\vec{J}$.

If we take the curl of this in spherical coordinates, we obtain

$$B_{2,r}(\vec{r}) = \frac{1}{r\,\sin\theta}\,\frac{\partial}{\partial\theta}(\sin\theta\,A_{2,\phi}) = 2\,\frac{\mu_o}{4\,\pi}\,\frac{m\,\cos\theta}{r^3} \tag{5.126}$$

$$B_{2,\theta}(\vec{r}) = -\frac{1}{r}\,\frac{\partial}{\partial r}(r\,A_{2,\phi}) = \frac{\mu_o}{4\,\pi}\,\frac{m\,\sin\theta}{r^3} \tag{5.127}$$

$$B_{2,\phi}(\vec{r}) = 0 \tag{5.128}$$

$$\text{or} \quad \vec{B}_2(\vec{r}) = \frac{\mu_o}{4\,\pi}\,\frac{m}{r^3}\left(2\,\hat{r}\,\cos\theta + \hat{\theta}\,\sin\theta\right) \tag{5.129}$$

which matches the form of Equation 3.239 for an electric dipole.

Let's derive the more generic result by releasing the condition $\vec{m} = m\,\widehat{z}$:

$$\vec{B}_2(\vec{r}) = \vec{\nabla} \times \vec{A} = \sum_{i,j,k=1}^{3} \epsilon_{ijk}\widehat{r}_i \frac{\partial A_k}{\partial r_j} = \frac{\mu_o}{4\,\pi} \sum_{i,j,k,\ell,m=1}^{3} \epsilon_{ijk}\widehat{r}_i \frac{\partial}{\partial r_j}\epsilon_{k\ell m}\left(\frac{m_\ell r_m}{r^3}\right) \tag{5.130}$$

$$= \frac{\mu_o}{4\,\pi} \sum_{i,j,k,\ell,m=1}^{3} \epsilon_{ijk}\epsilon_{k\ell m}\widehat{r}_i \left[\frac{m_\ell \delta_{jm}}{r^3} - \frac{3}{2}\frac{m_\ell r_m}{r^5}\left(2r_j\right)\right] \tag{5.131}$$

We use the cyclicity of the Levi-Civita symbol in its indices and the identities $\sum_{k=1}^{3}\epsilon_{kij}\epsilon_{k\ell m} = \delta_{i\ell}\delta_{jm} - \delta_{im}\delta_{j\ell}$ and $\sum_{j,k=1}^{3}\epsilon_{jki}\epsilon_{jk\ell} = 2\delta_{i\ell}$ to rewrite the above in a form identical to that of the electric dipole, Equation 3.242:

$$\vec{B}_2(\vec{r}) = \frac{\mu_o}{4\,\pi}\sum_{i=1}^{3}\widehat{r}_i\left[\frac{2m_i}{r^3} - \frac{3}{r^5}\left(m_i\sum_{j=1}^{3}r_j r_j - r_i\sum_{j=1}^{3}m_j r_j\right)\right] \tag{5.132}$$

$$= \frac{\mu_o}{4\,\pi}\sum_{i=1}^{3}\widehat{r}_i\frac{3\,r_i\left(\vec{m}\cdot\vec{r}\right) - m_i\left(\vec{r}\cdot\vec{r}\right)}{r^5} \tag{5.133}$$

$$\Longrightarrow \qquad \boxed{\vec{B}_2(\vec{r}) = \frac{\mu_o}{4\,\pi}\,\frac{3\left(\vec{m}\cdot\widehat{r}\right)\widehat{r} - \vec{m}}{r^3}} \tag{5.134}$$

Lecture 17:

Magnetostatics IV:
Magnetic Multipoles (cont'd)

Magnetostatics in Matter I:
Field of a Magnetized Object
Auxiliary Field $\vec{H}$

Date Revised: 2024/02/29 06:00
Date Given: 2024/02/29

Force on a Magnetic Dipole (à la Jackson)

As we did for electric multipoles, let's consider the problem of the force and torque on a magnetic dipole. However, because there is no magnetic potential energy function, we must begin from the Lorentz Force on the current distribution, which is given by

$$\vec{F}_{mag} = \int_{\mathcal{V}} d\tau \, \vec{J}(\vec{r}) \times \vec{B}(\vec{r}) \tag{5.135}$$

As we did in the case of the force on an electric multipole, we Taylor expand $\vec{B}(\vec{r})$. Again, as we did for electrostatics, we place the multipole at the origin and will generalize the result later. The expansion is

$$B_k(\vec{r}) = B_k(\vec{r} = \vec{0}) + \sum_{m=1}^{3} r_m \left. \frac{\partial B_k}{\partial r_m} \right|_{\vec{r}=\vec{0}} + \cdots \tag{5.136}$$

Thus, the Lorentz Force is

$$\vec{F}_{mag} = \sum_{i,j,k=1}^{3} \epsilon_{ijk}\, \widehat{r}_i \int_{\mathcal{V}} d\tau\, J_j(\vec{r})\, B_k(\vec{r}) \tag{5.137}$$

$$= \sum_{i,j,k=1}^{3} \epsilon_{ijk}\, \widehat{r}_i \left[B_k(\vec{0}) \int_{\mathcal{V}} d\tau\, J_j(\vec{r}) + \sum_{m=1}^{3} \left(\frac{\partial B_k}{\partial r_m}\bigg|_{\vec{0}} \right) \int_{\mathcal{V}} d\tau\, J_j\, r_m + \cdots \right] \tag{5.138}$$

We have done both these integrals before. The first one contains the monopole of the current distribution, which vanishes as in Equation 5.104. Since we will see that the second term is in general nonzero and is proportional to the magnetic dipole moment, let's call it $\vec{F}_{dip}$ and focus on it, dropping the higher-order terms. It is very similar in structure to what we encountered in calculating the dipole term in Equation 5.107. Applying the same tricks we used there to obtain Equation 5.116, we may rewrite it as

$$\vec{F}_{dip} = \sum_{i,j,k,m,n=1}^{3} \epsilon_{ijk}\, \widehat{r}_i \left(\frac{\partial B_k}{\partial r_m}\bigg|_{\vec{0}} \right) \int_{\mathcal{V}} d\tau\, \frac{1}{2}\, \epsilon_{jmn} \left[\vec{J}(\vec{r}) \times \vec{r} \right]_n \tag{5.139}$$

$$= - \sum_{i,j,k,m,n=1}^{3} \epsilon_{ijk}\epsilon_{jmn}\, \widehat{r}_i \left(\frac{\partial B_k}{\partial r_m}\bigg|_{\vec{0}} \right) m_n \qquad \text{with} \quad \vec{m} = \frac{1}{2} \int_{\mathcal{V}} d\tau \left[\vec{r} \times \vec{J}(\vec{r}) \right] \tag{5.140}$$

We use the vector identity Equation 5.114, $\sum_{j=1}^{3} \epsilon_{jik}\epsilon_{jmn} = \delta_{im}\delta_{kn} - \delta_{in}\delta_{km}$, and also use $\epsilon_{ijk} = -\epsilon_{jik}$ to adjust the indices to match this expression, yielding

$$\vec{F}_{dip} = \sum_{i,k,m,n=1}^{3} (\delta_{im}\delta_{kn} - \delta_{in}\delta_{km})\, \widehat{r}_i \left(\frac{\partial B_k}{\partial r_m}\bigg|_{\vec{0}} \right) m_n \tag{5.141}$$

$$= \sum_{i,k=1}^{3} \widehat{r}_i \left[\left(\frac{\partial B_k}{\partial r_i}\bigg|_{\vec{0}} \right) m_k - \left(\frac{\partial B_k}{\partial r_k}\bigg|_{\vec{0}} \right) m_i \right] \tag{5.142}$$

$$= \vec{\nabla}\left(\vec{m}\cdot\vec{B}\right)\Big|_{\vec{0}} - \vec{m}\left(\vec{\nabla}\cdot\vec{B}\right)\Big|_{\vec{0}} \tag{5.143}$$

The second term vanishes. Generalizing the first term to a dipole at an arbitrary position, we have

$$\boxed{\vec{F}_{dip} = \vec{\nabla}\left[\vec{m}\cdot\vec{B}(\vec{r})\right] \qquad \text{with} \quad \vec{m} = \frac{1}{2}\int_{\mathcal{V}} d\tau \left[\vec{r}\,' \times \vec{J}(\vec{r}\,')\right]} \tag{5.144}$$

The force causes the magnetic dipole to move to a local maximum of $\vec{m}\cdot\vec{B}$. Note how it is identical to the force on an electric dipole in an electric field, Equation 3.252. We'll address below the implication that the magnetic field can do work on the dipole.

Torque on a Magnetic Dipole (à la Jackson)

We may obtain from the Lorentz Force Law on a current distribution the corresponding torque:

$$\vec{N}_{mag} = \int_{\mathcal{V}} d\tau\, \vec{r} \times \left[\vec{J}(\vec{r}) \times \vec{B}(\vec{r}) \right] \tag{5.145}$$

where we have just added up the torque volume element by volume element in the same way we summed the force. When we Taylor expand the magnetic field, we have

$$\vec{N}_{mag} = \int_{\mathcal{V}} d\tau\, \vec{r} \times \left[\vec{J}(\vec{r}) \times \vec{B}(\vec{0}) \right] + \cdots \tag{5.146}$$

Because of the $\vec{r}\times$ inside the integrand, the zeroth-order term no longer vanishes and so we do not need to consider the next order term in the Taylor expansion. We will write the zeroth-order term as $\vec{N}_{dip}$ below for reasons that will become clear.

To get the above expression into a familiar form, we need to repeat the same kinds of vector arithmetic tricks we have used before. First, we apply the $BAC - CAB$ rule, $\vec{a} \times (\vec{b} \times \vec{c}) = \vec{b}(\vec{a} \cdot \vec{c}) - \vec{c}(\vec{a} \cdot \vec{b})$, which we can do without having to write things in terms of indices because there are no derivatives floating around:

$$\vec{N}_{dip} = \int_{\mathcal{V}} d\tau\, \vec{r} \times \left[\vec{J}(\vec{r}) \times \vec{B}(\vec{0})\right] = \int_{\mathcal{V}} d\tau\, \vec{J}(\vec{r}) \left[\vec{r} \cdot \vec{B}(\vec{0})\right] - \int_{\mathcal{V}} d\tau\, \vec{B}(\vec{0}) \left[\vec{r} \cdot \vec{J}(\vec{r})\right] \tag{5.147}$$

We can make the second term vanish by the same kinds of tricks we used earlier during the vector potential multipole expansion:

$$\vec{r} \cdot \vec{J}(\vec{r}) = \left[r\, \vec{\nabla} r\right] \cdot \vec{J}(\vec{r}) = \frac{1}{2} \left[\vec{\nabla} r^2\right] \cdot \vec{J}(\vec{r}) = \frac{1}{2} \left\{\vec{\nabla} \cdot \left[r^2 \vec{J}(\vec{r})\right] - r^2\, \vec{\nabla} \cdot \vec{J}(\vec{r})\right\} \tag{5.148}$$

In this expression, the second term vanishes under the steady-state assumption, and the first term can be turned into a surface integral with integrand $r^2 \vec{J}(\vec{r})$. Since we are considering a localized current distribution, the surface can be taken far enough out that $\vec{J}(\vec{r})$ vanishes on the surface.

The first term looks again like the expression we have encountered in Equation 5.107, which becomes apparent when we write it out in component form:

$$\vec{N}_{dip} = \sum_{i,j=1}^{3} \widehat{r}_i \, B_j(\vec{0}) \int_{\mathcal{V}} d\tau \, J_i(\vec{r}) \, r_j \tag{5.149}$$

We again apply the same tricks used to arrive at Equation 5.116:

$$\vec{N}_{dip} = \sum_{i,j=1}^{3} \widehat{r}_i \, B_j(\vec{0}) \int_{\mathcal{V}} d\tau \, \frac{1}{2} \, \epsilon_{ijk} \left[\vec{J}(\vec{r}) \times \vec{r} \right]_k = -\frac{1}{2} \, \vec{B}(\vec{0}) \times \int_{\mathcal{V}} d\tau \, \vec{r} \times \vec{J}(\vec{r}) \tag{5.150}$$

$$= -\vec{B}(\vec{0}) \times \vec{m} \qquad \text{with} \qquad \vec{m} = \frac{1}{2} \int_{\mathcal{V}} d\tau \left[\vec{r} \times \vec{J}(\vec{r}) \right] \tag{5.151}$$

Generalizing to a multipole distribution centered on an arbitrary point, the zeroth-order term in the torque is (and hence the $_{dip}$ subscript)

$$\boxed{\vec{N}_{dip} = \vec{m} \times \vec{B}(\vec{r}) \qquad \text{with} \qquad \vec{m} = \frac{1}{2} \int_{\mathcal{V}} d\tau' \left[\vec{r}' \times \vec{J}(\vec{r}') \right]} \tag{5.152}$$

The magnetic dipole feels a torque that tends to align it with the magnetic field (the torque vanishes when $\vec{m}$ is aligned with $\vec{B}$), again like the situation for an electric dipole in an electric field.

Potential Energy of a Magnetic Dipole

We can do the line integral of the force or the angular integral of the torque to determine that we can write a potential energy

$$\boxed{U(\vec{r}) = -\vec{m} \cdot \vec{B}(\vec{r})} \tag{5.153}$$

This form for the potential energy expresses two features of magnetic dipoles: they like to be aligned with the local magnetic field, and they seek the region of largest $\vec{m} \cdot \vec{B}$.

The thing that should be concerning about this expression is that we argued earlier that magnetic fields can do no work, yet here we have the possibility of such work. That is because we are assuming $\vec{m}$ is held fixed. For a finite current loop, there must be a battery doing work to keep the current fixed as $\vec{m}$ moves or turns relative to $\vec{B}$: such motion yields changing magnetic fields, which, as you know from Ph1c, generate voltages around the loop in which the current for $\vec{m}$ flows. The battery will be the thing doing the work to counter these voltages and keep the current flowing. If $\vec{m}$ is a property of a fundamental particle, then there is no explicit battery: it is simply an empirical fact that $|\vec{m}|$ cannot change, and one that we must incorporate as a postulate.

Section 6
Magnetostatics in Matter

Paramagnetism and Diamagnetism

See Griffiths Sections 6.1.1 and 6.1.3 and Purcell Sections 11.1 and 11.5 for discussions of paramagnetism and diamagnetism. This will be discussed in class briefly, but there is little to add to their discussions, but we briefly summarize that discussion here.

Diamagnetism arises due to the orbital angular momentum of electrons: when they execute their orbits in a magnetic field, the magnetic field adds to $(\vec{B}||\vec{\omega}|| - \vec{m})$ or partially cancels $(\vec{B}|| - \vec{\omega}||\vec{m})$ the Coulomb force that provides the centripetal force. This causes the electron velocity to, respectively, increase or decrease. In both cases, this change in speed yields a change $\Delta\vec{m}|| - \vec{B}$. We will see later that an ensemble of magnetic dipoles that change in this way yield an additional field that partially cancels the applied magnetic field, hence the term diamagnetism.

Paramagnetism arises from an entirely quantum-mechanical phenomenon: electrons have a magnetic moment antiparallel to their spin and, when electrons are unpaired in atomic orbitals, those magnetic moments are uncanceled. We saw in the previous section that magnetic dipoles align with a magnetic field, opposite to the effect seen in diamagnetism. We will see that an ensemble of such aligning magnetic moments yields a field in the same direction as the applied field, hence the term paramagnetism.

We remind you that unpaired electrons are present not just in atoms with odd numbers of electrons but also in electrons with even Z in which there are not enough electrons to fill a particular atomic orbital shell (ℓ) that has multiple ℓ_z values. Because of Coulomb repulsion, such electrons tend to spread out among different ℓ_z values, and then there is an "exchange" term that we will discuss later that causes the lower-energy state to be the one in which the spatial wavefunction is antisymmetric under exchange, which then implies the spin state must be symmetric under exchange in order for the entire state to be antisymmetric under exchange.

The Field of a Magnetized Object

Bound Currents

Suppose we have an object with a position-dependent macroscopic density of magnetic moments, or *macroscopic magnetization density* $\vec{M}(\vec{r})$, where the magnetic moment of an infinitesimal volume $d\tau$ is

$$d\vec{m} = \vec{M}(\vec{r})\, d\tau \tag{6.1}$$

$\vec{M}$ is not to be confused with the magnetization density $\mathcal{M}(\vec{r})$; the latter can be for some arbitrary current distribution, while the former is specifically to be considered to be a density of magnetic dipole moments. $\mathcal{M}(\vec{r})$ should give $\vec{M}(\vec{r})$ for this special case of pure dipoles. We will, confusingly, drop "macroscopic" from here on out. Assuming we are looking at the dipoles from a macroscopic enough scale that the dipole approximation is valid, we may use our expression for the vector potential of a magnetic dipole, Equation 5.120, to calculate the contribution to the vector potential at $\vec{r}$ due to the above infinitesimal volume at $\vec{r}\,'$:

$$d\vec{A}(\vec{r}) = \frac{\mu_o}{4\,\pi}\,\frac{d\vec{m}(\vec{r}\,') \times (\vec{r} - \vec{r}\,')}{|\vec{r} - \vec{r}\,'|^3} = \frac{\mu_o}{4\,\pi}\,\frac{d\tau'\,\vec{M}(\vec{r}\,') \times (\vec{r} - \vec{r}\,')}{|\vec{r} - \vec{r}\,'|^3} \tag{6.2}$$

Integrating over the volume containing the magnetization density, we have

$$\vec{A}(\vec{r}) = \frac{\mu_o}{4\,\pi} \int_{\mathcal{V}} d\tau' \, \frac{\vec{M}(\vec{r}') \times (\vec{r} - \vec{r}')}{|\vec{r} - \vec{r}'|^3} \tag{6.3}$$

Now, we use $(\vec{r} - \vec{r}')/|\vec{r} - \vec{r}'|^3 = \vec{\nabla}_{\vec{r}'} |\vec{r} - \vec{r}'|^{-1}$ (note that the gradient is with respect to $\vec{r}'$, not $\vec{r}$!), which allows us to apply the product rule for curl, $\vec{\nabla} \times (f\vec{a}) = f\vec{\nabla} \times \vec{a} - \vec{a} \times \vec{\nabla} f$:

$$\vec{A}(\vec{r}) = \frac{\mu_o}{4\,\pi} \int_{\mathcal{V}} d\tau' \, \vec{M}(\vec{r}') \times \vec{\nabla}_{\vec{r}'} \left(\frac{1}{|\vec{r} - \vec{r}'|} \right) \tag{6.4}$$

$$= \frac{\mu_o}{4\,\pi} \int_{\mathcal{V}} d\tau' \, \frac{\vec{\nabla}_{\vec{r}'} \times \vec{M}(\vec{r}')}{|\vec{r} - \vec{r}'|} - \frac{\mu_o}{4\,\pi} \int_{\mathcal{V}} d\tau' \, \vec{\nabla}_{\vec{r}'} \times \left(\frac{\vec{M}(\vec{r}')}{|\vec{r} - \vec{r}'|} \right) \tag{6.5}$$

$$= \frac{\mu_o}{4\,\pi} \int_{\mathcal{V}} d\tau' \, \frac{\vec{\nabla}_{\vec{r}'} \times \vec{M}(\vec{r}')}{|\vec{r} - \vec{r}'|} + \frac{\mu_o}{4\,\pi} \int_{\mathcal{S}(\mathcal{V})} da' \, \frac{\vec{M}(\vec{r}') \times \hat{n}(\vec{r}')}{|\vec{r} - \vec{r}'|} \tag{6.6}$$

where, in the last step, we have used a vector identity that we will prove on the following slide.

Let's prove the vector identity we just used, which is a corollary of the divergence theorem for the curl. Let $\vec{a}(\vec{r})$ be an arbitrary vector field and let $\vec{c}$ be an arbitrary constant vector. Then the divergence theorem tells us

$$\int_{\mathcal{V}} d\tau\, \vec{\nabla} \cdot [\vec{a}(\vec{r}) \times \vec{c}] = \oint_{\mathcal{S}(\mathcal{V})} da\, \hat{n}(\vec{r}) \cdot [\vec{a}(\vec{r}) \times \vec{c}] \tag{6.7}$$

Now, apply the cyclicity of triple scalar products (along with the fact that $\vec{c}$ is constant and can thus it can be moved past $\vec{\nabla}$) and bring $\vec{c}$ outside the integrals (since it is a constant vector):

$$\vec{c} \cdot \int_{\mathcal{V}} d\tau\, \left[\vec{\nabla} \times \vec{a}(\vec{r})\right] = \vec{c} \cdot \oint_{\mathcal{S}(\mathcal{V})} da\, [\hat{n}(\vec{r}) \times \vec{a}(\vec{r})] \tag{6.8}$$

Since $\vec{c}$ is arbitrary, the expression must hold for any $\vec{c}$ and thus:

$$\int_{\mathcal{V}} d\tau\, \left[\vec{\nabla} \times \vec{a}(\vec{r})\right] = \oint_{\mathcal{S}(\mathcal{V})} da\, [\hat{n}(\vec{r}) \times \vec{a}(\vec{r})] \tag{6.9}$$

which is what we wanted to prove.

Making some definitions, we recognize that the vector potential can be considered to be sourced by a *bound volume current density* $\vec{J}_b(\vec{r})$ and a *bound surface current density* $\vec{K}_b(\vec{r})$:

$$\boxed{\vec{J}_b(\vec{r}) = \vec{\nabla} \times \vec{M}(\vec{r}) \qquad \vec{K}_b(\vec{r}) = \vec{M}(\vec{r}) \times \hat{n}(\vec{r})} \tag{6.10}$$

$$\boxed{\vec{A}(\vec{r}) = \frac{\mu_o}{4\,\pi} \int_{\mathcal{V}} d\tau' \, \frac{\vec{J}_b(\vec{r}')}{|\vec{r} - \vec{r}'|} + \frac{\mu_o}{4\,\pi} \oint_{\mathcal{S}(\mathcal{V})} da' \, \frac{\vec{K}_b(\vec{r}')}{|\vec{r} - \vec{r}'|}} \tag{6.11}$$

The way in which these current densities source $\vec{A}$ is identical to the way in which free current densities do. Moreover, we can see the clear analogy to bound volume and surface charges in the case of polarized materials.

Griffiths Section 6.2.2 gives a nice discussion of the physical interpretation of bound currrents that will be presented in class, but there is not much to add here.

Example 6.1: Uniformly Magnetized Sphere

Center the sphere of radius R at the origin. Let $\vec{M} = M\,\hat{z}$. Then

$$\vec{J}_b(\vec{r}) = \vec{\nabla} \times M\,\hat{z} = 0 \qquad \vec{K}_b(\vec{r}) = M\,\hat{z} \times \hat{n} = M\,\hat{z} \times \hat{r} = \hat{\phi}\,M \sin\theta \tag{6.12}$$

We need to calculate

$$\vec{A}(\vec{r}) = \frac{\mu_o}{4\,\pi}\,R^2 \int_0^{2\,\pi} d\phi' \int_0^{\pi} d\theta'\,\sin\theta'\,\frac{M \sin\theta'\,\hat{\phi}}{|\vec{r} - \vec{r}\,'|} \tag{6.13}$$

$$= \frac{\mu_o}{4\,\pi}\,R^2 \int_0^{2\,\pi} d\phi' \int_0^{\pi} d\theta'\,\sin\theta'\,\frac{M \sin\theta'\,(-\hat{x}\,\sin\phi' + \hat{y}\,\cos\phi')}{|\vec{r} - \vec{r}\,'|} \tag{6.14}$$

(The R^2 out front is because an area integral, not just a solid angle integral, needs to be done.) Recall that we did this calculation when we considered the magnetic vector potential of a spinning spherical shell with a uniform charge density, Example 5.1. In that case the surface current density was $\vec{K} = \hat{\phi}\,\sigma\,\omega\,R\,\sin\theta$, so we can use that result with $\sigma\,\omega\,R$ replaced by M. The result is

$$\vec{A}(r \le R, \theta, \phi) = \frac{\mu_o}{3}\,M\,r\,\sin\theta\,\hat{\phi} \qquad \vec{A}(r \ge R, \theta, \phi) = \frac{\mu_o}{4\,\pi}\left(\frac{4\,\pi}{3}\,R^3 M\right)\frac{\sin\theta}{r^2}\,\hat{\phi} \tag{6.15}$$

Evaluating the curl of the first term to obtain the magnetic field, we have inside the sphere

$$\vec{B}(r \leq R) = \vec{\nabla} \times \vec{A}(r \leq R) = \frac{1}{3}\,\mu_o\,M\left[2\,\widehat{r}\,\cos\theta - \widehat{\theta}\,\sin\theta\right] = \frac{2}{3}\,\mu_o\,\vec{M} \qquad (6.16)$$

which is a uniform field pointing in the *same* direction as the magnetization.

For $r \geq R$, we have

$$\vec{A}(r \geq R) = \frac{\mu_o}{4\,\pi}\,\frac{\vec{m} \times \widehat{r}}{r^2} \qquad \vec{m} = \frac{4\,\pi}{3}\,R^3\vec{M} \qquad (6.17)$$

which is the vector potential (thus yielding the field of) a pure dipole with magnetic moment given by integrating the uniform magnetization density over the sphere. This form is exact for *all* $r \geq R$.

Let's compare to the case of a uniformly polarized dielectric sphere:

$$r \leq R \qquad \vec{E}(\vec{r}) = -\frac{1}{3\,\epsilon_o}\,\vec{P} \qquad \vec{B}(\vec{r}) = \frac{2}{3}\,\mu_o\,\vec{M} \tag{6.18}$$

$$r \geq R \qquad V(\vec{r}) = \frac{1}{4\,\pi\,\epsilon_o}\,\frac{\vec{p}\cdot\widehat{r}}{r^2} \qquad \vec{A}(\vec{r}) = \frac{\mu_o}{4\,\pi}\,\frac{\vec{m}\times\widehat{r}}{r^2} \tag{6.19}$$

$$\vec{p} = \frac{4\,\pi}{3}\,R^3\vec{P} \qquad \vec{m} = \frac{4\,\pi}{3}\,R^3\vec{M} \tag{6.20}$$

Inside the sphere, the difference is a factor of -2 and the exchange of $1/\epsilon_o$ for μ_o. Outside the sphere, the two potentials result in fields identical up to the replacement of $\vec{P}/\epsilon_o$ by $\mu_o\,\vec{M}$. The difference in the $r \leq R$ expressions reflects the fact that the magnetic field of the bound surface current (*i.e.*, of $\vec{M}$) is aligned with $\vec{M}$ while the electric field of the surface bound charge density (*i.e.*, of $\vec{P}$) is opposite to $\vec{P}$. This sign difference is a generic phenomenon, resulting in the very different behavior of electrostatic and magnetostatic fields in matter.

The Auxiliary Field $\vec{H}$ and Magnetic Permeability

Definition of the Auxiliary Field

We saw that $\vec{A}$ is sourced by the bound current density $\vec{J}_b = \vec{\nabla} \times \vec{M}$ in the same way it would be sourced by a *free* current density $\vec{J}_f$. Therefore, Ampere's Law is satisfied with the sum of the two currents:

$$\frac{1}{\mu_o} \vec{\nabla} \times \vec{B} = \vec{J}_f + \vec{J}_b = \vec{J}_f + \vec{\nabla} \times \vec{M} \tag{6.21}$$

If we want to write an Ampere's Law in terms of the free currents only, in the same way that we wanted to write Gauss's Law in terms of the free charges only, then we can define the *auxiliary field*

$$\boxed{\vec{H} \equiv \frac{\vec{B}}{\mu_o} - \vec{M}} \tag{6.22}$$

In contrast to electrostatics, where the displacement field was the sum of the electric field and the polarization density, here the auxiliary field is the difference of the magnetic field and the magnetization density. The sign flip comes from the differing signs in the definition of the bound charge and current densities: $\rho_b = -\vec{\nabla} \cdot \vec{P}$ while $\vec{J}_b = \vec{\nabla} \times \vec{M}$, which itself comes from the commutative vs. anticommutative natures of the dot and cross product.

With this definition of $\vec{H}$, we then have

$$\vec{\nabla} \times \vec{H} = \frac{1}{\mu_o} \vec{\nabla} \times \vec{B} - \vec{\nabla} \times \vec{M} = \vec{J}_f + \vec{J}_b - \vec{J}_b = \vec{J}_f \tag{6.23}$$

Therefore, we have a modified Ampere's Law

$$\boxed{\vec{\nabla} \times \vec{H} = \vec{J}_f \qquad \Longleftrightarrow \qquad \oint_{\mathcal{C}} d\vec{\ell} \cdot \vec{H}(\vec{r}) = \int_{\mathcal{S}(\mathcal{C})} da\, \hat{n}(\vec{r}) \cdot \vec{J}_f(\vec{r}) = I_{f,enc}} \tag{6.24}$$

Thus, as intended, we have an Ampere's Law in terms of the free currents only, which (partially) source $\vec{H}$. The fact that $\vec{H}$ satisfies Ampere's Law in the free current leads some to use the name *applied field* for it. That may be misleading, though, because the free current does not tell one everything one must know to determine $\vec{H}$ (in the same way that ρ_f does not completely determine the displacement field $\vec{D}$).

To fully specify $\vec{H}$, we need to know its divergence, which is given by applying $\vec{\nabla} \cdot \vec{B} = 0$:

$$\boxed{\vec{\nabla} \cdot \vec{H} = -\vec{\nabla} \cdot \vec{M}} \tag{6.25}$$

This nonvanishing of $\vec{\nabla} \cdot \vec{H}$ is analogous to the nonvanishing of $\vec{\nabla} \times \vec{D}$ in electrostatics.

There is an example of how to calculate $\vec{H}$ using the above Ampere's Law in Griffiths Example 6.2.

Boundary Conditions on $\vec{H}$

From the boundary conditions on $\vec{B}$ at an interface, we can derive boundary conditions on $\vec{H}$. The continuity of the normal component of the magnetic field (Equation 5.76) along with Equation 6.22 implies

$$\boxed{\hat{n}(\vec{r}) \cdot \left[\vec{H}_2(\vec{r}) - \vec{H}_1(\vec{r})\right] = -\hat{n}(\vec{r}) \cdot \left[\vec{M}_2(\vec{r}) - \vec{M}_1(\vec{r})\right]} \tag{6.26}$$

Applying the same arguments using Ampere's Law for $\vec{H}$ as we did using Ampere's Law for $\vec{B}$, we can also conclude the analogy of Equation 5.82:

$$\boxed{\left[\vec{H}_2(\vec{r}) - \vec{H}_1(\vec{r})\right] \cdot \hat{s}(\vec{r}) = \left[\vec{K}_f(\vec{r}) \times \hat{n}(\vec{r})\right] \cdot \hat{s}(\vec{r})} \tag{6.27}$$

where $\vec{K}_f$ is the free surface current density at the interface.

Recall that we found alternative forms of the corresponding boundary conditions for $\vec{B}$, Equations 5.83 and 5.84:

$$\vec{B}_2(\vec{r}) - \vec{B}_1(\vec{r}) = \mu_o \, \vec{K}(\vec{r}) \times \hat{n}(\vec{r})$$

$$\hat{n}(\vec{r}) \times \left[\vec{B}_2(\vec{r}) - \vec{B}_1(\vec{r}) \right] = \mu_o \, \vec{K}(\vec{r})$$

There is no trivial analogue of the first one because it relied on the normal component of $\vec{B}$ being continuous. However, we can obtain the analogue of the second equation, though we have to do it in a different way because, for $\vec{B}$, we used the first equation above to obtain the second one.

We start by using $\widehat{s} = \widehat{t} \times \widehat{n}$ and then applying the cyclicity of the triple scalar product on both sides:

$$\left[\vec{H}_2 - \vec{H}_1\right] \cdot \left[\widehat{t} \times \widehat{n}\right] = [\widehat{n} \times \widehat{s}] \cdot \vec{K}_f \tag{6.28}$$

$$\widehat{t} \cdot \left(\widehat{n} \times \left[\vec{H}_2 - \vec{H}_1\right]\right) = \widehat{t} \cdot \vec{K}_f \tag{6.29}$$

The same equation holds trivially with $\widehat{t}$ replaced by $\widehat{n}$: the left side vanishes because $\widehat{n}$ is perpendicular to any cross product involving $\widehat{n}$ and the right side vanishes because $\vec{K}_f$ is always perpendicular to $\widehat{n}$. This, combined with the fact that $\widehat{t}$ in the above can be any vector in the plane of the boundary, implies the more general statement

$$\boxed{\widehat{n}(\vec{r}) \times \left[\vec{H}_2(\vec{r}) - \vec{H}_1(\vec{r})\right] = \vec{K}_f(\vec{r})} \tag{6.30}$$

which is the analogue of the second equation on the previous slide. But note that this equation provides no information about the normal component of $\vec{H}$ because it is related to the normal component of $\vec{M}$.

What Sources $\vec{H}$? When Does It Vanish?

Considering the uniformly magnetized sphere example we just looked at, we see

$$\vec{H}(r < R) = \frac{\vec{B}(r \leq R)}{\mu_o} - \vec{M} = \frac{2}{3}\vec{M} - \vec{M} = -\frac{1}{3}\vec{M} \tag{6.31}$$

$$\vec{H}(r > R) = \frac{\vec{B}(r \geq R)}{\mu_o} = \frac{\text{field of the magnetic dipole } \vec{m} = \frac{4\pi}{3} R^3 \vec{M}}{\mu_o} \tag{6.32}$$

This example highlights the importance of the nonvanishing of $\vec{\nabla} \cdot \vec{H}$. There is no free current in this problem, so one might be inclined to think $\vec{H}$ vanishes by analogy to the fact $\vec{B}$ would vanish if there were no total current. But the nonzero nature of $\vec{\nabla} \cdot \vec{H}$ means that $\vec{H}$ has another sourcing term that is not captured by Ampere's Law alone. In this case, this sourcing term manifests as a discontinuity of the normal component of $\vec{M}$ at $r = R$. This is analogous to the way that, even if there is no free charge, there may be a displacement field $\vec{D}$, sourced by $\vec{\nabla} \times \vec{P}$ and/or a discontinuity in the tangential component of $\vec{P}$, as we saw in for the polarized sphere (Example 5.1) and the spherical cavity in a dielectric with uniform field applied (Example 5.6). To have $\vec{H}$ vanish identically, one needs to have $\vec{\nabla} \cdot \vec{M} = 0$ and also trivial boundary conditions on $\vec{M}$ (no change in $\hat{n} \cdot \vec{M}$).

This all makes sense given the Helmholtz theorem: since $\vec{\nabla} \cdot \vec{H}$ does not vanish, $\vec{H}$ is not just the curl of a vector potential, but must be the sum of the gradient of a scalar potential and the curl of a vector potential. Ampere's Law for $\vec{H}$ tells us that the free current density sources the vector potential, while $-\vec{\nabla} \cdot \vec{M}$ sources the scalar potential. We will see later that the latter point allows us to use our electrostatic boundary value problem techniques.

In particular, in the example of the uniformly magnetized sphere, we see that $\vec{H}$ is identical in form to $\vec{E}$ from the uniformly polarized sphere up to the replacement $\vec{P}/\epsilon_o \to \vec{M}$, so the scalar potential that yields $\vec{M}$ will have the same form, up to this replacement, as the scalar potential that yields $\vec{E}$. We'll pursue this analogy in detail when we discuss boundary value problems for magnetostatic systems.

We can make the same point about ρ_f not being the only source of $\vec{D}$; when $\vec{\nabla} \times \vec{P}$ is nonzero, then $\vec{D}$ receives an additional sourcing term. It was not convenient to make this point when we discussed $\vec{D}$ initially because we had not yet learned about vector potentials and how to discuss sourcing of $\vec{D}$ by a vector field, $\vec{\nabla} \times \vec{P}$. But now we do, and so it should be clear that $\vec{D}$ received a contribution that is sourced by $\vec{\nabla} \times \vec{P}$ in the same way that $\vec{H}$ receives a contribution that is sourced by $\vec{\nabla} \times \vec{H} = \vec{J}_f$.

In particular, in Example 4.5, the capacitor with two side-by-side dielectrics, we saw such a situation, manifested by the discontinuity in the tangential component of $\vec{P}$ at the interface between the two dielectrics.

Who Cares About $\vec{H}$?

Is $\vec{H}$ any more useful than $\vec{D}$ was?

The thing that limits the utility of $\vec{D}$ is that, in practice, one rarely controls free charge, the most obvious source for $\vec{D}$. In practice, one sets potentials using batteries or other voltage sources. Potentials specify $\vec{E}$, not $\vec{D}$. Consider the example of the parallel-plate capacitor with side-by-side dielectrics: σ_f ended being an output of the calculation after calculating $\vec{E}$ rather than an input that yielded $\vec{D}$.

On the other hand, $\vec{H}$ is sourced by the free currents, which is the thing one explicitly controls in the lab. For that reason alone, we expect $\vec{H}$ is of greater utility than $\vec{D}$. We will see this more clearly when we consider specific types of permeable materials.

The other reason we will find $\vec{H}$ more useful is that, in reality, we frequently come across ferromagnets, where $\vec{M}$ is provided and thus we are given the $\vec{\nabla} \cdot \vec{M}$ source for $\vec{H}$, but we rarely encounter ferroelectrics, where $\vec{P}$ and thus the $\vec{\nabla} \times \vec{P}$ source for $\vec{D}$ are provided. We would find $\vec{D}$ useful as a calculation tool if we were given a system in which $\vec{\nabla} \times \vec{P}$ were nonzero or, more likely, $\vec{P}$ were tangent to boundaries between a ferroelectric and vacuum or between different ferroelectrics. Then $\vec{\nabla} \times \vec{P}$ and any discontinuity in $\hat{n} \times \vec{P}$ would source $\vec{D}$ in the same way that $\vec{J}$ and a boundary $\vec{K}$ source $\vec{B}$.

Lecture 18:

Magnetostatics in Matter II:
Magnetic Permeability in Linear Systems
Boundary Value Problems in Magnetostatics

Date Revised: 2024/03/05 06:00
Date Given: 2024/03/05

Magnetic Permeability in Linear Materials

Many magnetic materials we will consider have a linear relationship between the field and the magnetization. The *magnetic susceptibility* of a material is defined to be the constant of proportionality between $\vec{M}$ and $\vec{H}$:

$$\vec{M} = \chi_m \vec{H} \tag{6.33}$$

(One can see why $\vec{H}$ is sometimes called the applied field!) Since $\vec{B} = \mu_o \left(\vec{H} + \vec{M}\right)$, we have

$$\vec{B} = \mu_o \left(\vec{H} + \vec{M}\right) = \mu_o \left(1 + \chi_m\right) \vec{H} \equiv \mu \vec{H} \tag{6.34}$$

where we have defined the *magnetic permeability* $\mu = \mu_o(1 + \chi_m)$. The quantity $\mu_r = 1 + \chi_m$ is the *relative permeability*. The definition of χ_m and μ follows a different convention than the definition of χ_e and ϵ. This is for the reason we discussed above: we experimentally control the free current and thus $\vec{H}$, whereas in electrostatics we control the voltages and thus $\vec{E}$. We define the permittivity and the permeability to be the constant of proportionality relating the thing we do control to the thing we do not control.

Paramagnetic materials have $\chi_m > 0$ because the magnetization is in the same direction as the field and so the field due to the free currents is added to by the field from the magnetization.

Diamagnetic materials have $\chi_m < 0$ because the magnetization is in the direction opposite the field and so the field due to the free currents is partially canceled by the field from the magnetization.

For electrostatics in matter, we were concerned entirely with *dielectric* materials: because every atom has some polarizability, every material is dielectric to some extent. In that case, the "di" prefix went with $\chi_e > 0$ (in contrast to $\chi_m < 0$ here) because of the different convention for the relation between $\vec{E}$ and $\vec{D}$.

Diamagnetic materials exist via the same kind of classical argument, now involving the response of currents in materials to applied fields.

The analogous *paraelectric* materials ($\chi_e < 0$) do not exist for the most part — it is hard to understand how one can get an electrically polarizable material to have $\chi_e < 0$. Metals can have negative permittivity at high frequencies (optical), but not DC.

Paramagnetic materials exist only because of quantum mechanics — the existence of magnetic moments not caused by an applied field. There are no such quantum-mechanics-caused electric dipole moments, mainly because such moments violate time-reversal symmetry while magnetic moments do not. They work differently because the "current" sourcing a magnetic dipole moment reverses sign under time reversal while the charges sourcing an electric dipole moment do not.

Example 6.2: Magnetizable Rod with Uniform Current (Griffiths 6.17)

Let's consider a rod of radius R whose axis is in the z direction and which carries a current $\vec{I} = I\,\hat{z}$ distributed uniformly across its cross section. Assume the material is linear with magnetic susceptibility χ_m. Let's find $\vec{H}$, $\vec{M}$, and $\vec{B}$.

Let's first see how far we can get without using χ_m. Ampere's Law for $\vec{H}$ tells us

$$\oint_{\mathcal{C}} \vec{H} \cdot d\vec{\ell} = \int_{\mathcal{S}(\mathcal{C})} da\,\hat{n} \cdot \vec{J}_f \tag{6.35}$$

This system has azimuthal symmetry as well as translational symmetry in z, so we can guess $\vec{H} = \vec{H}(s)$ where s is the radial coordinate in cylindrical coordinates. By the right-hand rule and the z translational symmetry, we expect $\vec{H} = H(s)\,\hat{\phi}$. This eliminates any concern about $\vec{\nabla} \cdot \vec{M}$ or $\hat{n} \cdot \vec{M}$: we know $\vec{M} = \chi_m \vec{H} \propto \vec{H}$, therefore we know, for the assumed form of $\vec{H}$, $\vec{\nabla} \cdot \vec{M} = 0$ inside the cylinder and $\hat{n} \cdot \vec{M} = 0$ at the surface of the cylinder. ($\vec{M} = 0$ outside the cylinder.) Adding in that we know $\vec{J}_f = \hat{z}\,I/\pi R^2$, Ampere's Law in $\vec{J}_f$ and $\vec{H}$ tells us

$$s \le R: \qquad 2\pi s H(s) = \pi s^2 \frac{I}{\pi R^2} \qquad \Longleftrightarrow \qquad \vec{H}(s) = \frac{I}{2\pi}\frac{s}{R^2}\hat{\phi} \tag{6.36}$$

$$s \ge R: \qquad 2\pi s H(s) = \pi R^2 \frac{I}{\pi R^2} \qquad \Longleftrightarrow \qquad \vec{H}(s) = \frac{I}{2\pi}\frac{1}{s}\hat{\phi} \tag{6.37}$$

If we do not know χ_m, we do not know $\vec{M}$ inside the material and so we cannot calculate $\vec{B}$ for $s \leq R$. For $s \geq R$, we have vacuum and so $\vec{M} = 0$ and $\vec{B} = \mu_o \vec{H}$:

$$\vec{B}(s \geq R) = \mu_o \vec{H}(s \geq R) = \frac{\mu_o \, \mathrm{I}}{2\,\pi\, s}\,\widehat{\phi} \tag{6.38}$$

Note that $\vec{B}(s \geq R)$ is unaffected by the presence of the magnetizable material — this is the same field we would have had with $\chi_m = 0$ inside the rod. We will see why below.

Next, if we use the linearity of the material with susceptibility χ_m, we have

$$\vec{M}(s \leq R) = \chi_m \vec{H}(s \leq R) = \chi_m \, \frac{\mathrm{I}}{2\,\pi} \, \frac{s}{R^2}\,\widehat{\phi} = \frac{\mu - \mu_o}{\mu_o} \, \frac{\mathrm{I}}{2\,\pi} \, \frac{s}{R^2}\,\widehat{\phi} \tag{6.39}$$

and therefore

$$\vec{B}(s \leq R) = \mu \vec{H}(s \leq R) = \frac{\mu \, \mathrm{I}}{2\,\pi} \, \frac{s}{R^2}\,\widehat{\phi} \tag{6.40}$$

All three fields are azimuthal inside and outside R. For paramagnetic materials, $\chi_m \geq 0$ ($\mu \geq \mu_o$), so $\vec{M}$ is parallel to $\vec{H}$ and $|\vec{B}| > \mu_o |\vec{H}|$ inside R. For diamagnetic materials, $\chi_m < 0$ ($\mu \leq \mu_o$), so $\vec{M}$ is antiparallel to $\vec{H}$ and $|\vec{B}| \leq \mu_o |\vec{H}|$ inside R.

Let's check the boundary conditions. All the fields are tangential at the boundary, so continuity of the normal component of $\vec{B}$ trivially satisfied (Equation 5.76), as is the equality of the change in the normal components of $\vec{H}$ and $-\vec{M}$ (Equation 6.26). There is no free surface current density, so we expect the tangential component of $\vec{H}$ to be continuous. We see this indeed holds, with value

$$\widehat{\phi} \cdot \vec{H}(s = R) = \frac{\mathrm{I}}{2\,\pi\,R} \tag{6.41}$$

The $\widehat{z}$ tangential component is trivially continuous since it vanishes.

For the sake of completeness, let's calculate the *bound* surface current and check that the tangential boundary condition on $\vec{B}$ is correct. The bound surface current is $\vec{K}_b = \vec{M} \times \widehat{n}$ (Equation 6.10). In this case, $\widehat{n} = \widehat{s}$, the radial unit vector in cylindrical coordinates, so

$$\vec{K}_b(s = R) = M(s = R)\,\widehat{\phi} \times \widehat{s} = -\chi_m \frac{\mathrm{I}}{2\,\pi\,R}\,\widehat{z} \tag{6.42}$$

For a paramagnetic materials ($\chi_m > 0$), the surface current points along $-\widehat{z}$ while, for diamagnetic materials ($\chi_m < 0$), it points along $+\widehat{z}$. One can see this physically by considering the direction of alignment of the dipoles and which direction the uncancelled current on the boundary flows. From the direction of this surface current, one can then see that the field of this surface current adds to the field of the free current for the paramagnetic case and partially cancels it for the diamagnetic case. (Note that we get this behavior even though the surface current is antiparallel to (parallel to) the direction of the volume current in the paramagnetic (diamagnetic) case. We get this behavior because the region of interest is *inside* the surface current density, not outside it. We'll look at the field outside next.) With the surface bound current in hand, let's check the boundary conditions on $\vec{B}$. It has no normal component in either region, so continuity of the normal component is trivially satisified. The discontinuity in the tangential component matches Equation 5.84:

$$\widehat{n} \times \left[\vec{B}_2 - \vec{B}_1\right] = \widehat{s} \times [\mu_o - \mu]\,\frac{\mathrm{I}}{2\,\pi\,R}\,\widehat{\phi} = -\mu_o \chi_m \frac{\mathrm{I}}{2\,\pi\,R}\,\widehat{z} = \mu_o\,\vec{K}_b \tag{6.43}$$

Let's also calculate the bound volume current density, $\vec{J}_b = \vec{\nabla} \times \vec{M}$ from Equation 6.10. It is

$$\vec{J}_b(\vec{r}) = \vec{\nabla} \times \vec{M} = \chi_m \vec{\nabla} \times \vec{H} = \chi_m \vec{J}_f = \chi_m \frac{\mathrm{I}}{\pi R^2} \widehat{z} \tag{6.44}$$

For paramagnetic materials, $\vec{J}_b$ is parallel to $\vec{J}_f$ and thus its field adds to the field of the free current, while, for diamagnetic materials, it is antiparallel and it partially cancels the free current's field.

Note that the integral of $\vec{J}_b$ over the cross section and the integral of $\vec{K}_b$ over the circumference are equal in magnitude ($\chi_m \, \mathrm{I}$) and opposite in sign, canceling perfectly. This is why the magnetic field outside the wire is only that due to the free current.

A modest extension to this problem would be to include a free surface current in the $\widehat{z}$ direction, which would then cause a discontinuity in the $\widehat{\phi}$ component of $\vec{H}$. You should try this on your own.

Boundary Value Problems in Magnetic Materials

Griffiths does not really consider boundary value problems in magnetostatics, so we follow Jackson §5.9–5.12.

General Conditions for Linear, Homogeneous Magnetic Materials

In linear, homogeneous dielectrics, we showed $\rho_b \propto \rho_f$. We just saw that a similar relation holds for linear, homogeneous magnetic materials, which we can derive generally:

$$\vec{J}_b = \vec{\nabla} \times \vec{M} = \vec{\nabla} \times \left(\frac{\mu - \mu_o}{\mu_o} \vec{H} \right) = \left(\frac{\mu - \mu_o}{\mu_o} \right) \vec{\nabla} \times \vec{H} = \left(\frac{\mu - \mu_o}{\mu_o} \right) \vec{J}_f \qquad (6.45)$$

In particular, if there is no free current in a linear, homogeneous magnetic material, then there is no bound current either. In such situations, the magnetic field is derivable from a scalar potential and Laplace's Equation holds everywhere there is no free current! Boundary conditions, and matching conditions between regions, will determine $\vec{H}$. We'll explore such situations shortly.

The General Technique

In general, it always holds that

$$\vec{B} = \vec{\nabla} \times \vec{A} \qquad \vec{H} = \vec{H}(\vec{B}) \qquad \vec{\nabla} \times \vec{H} = \vec{J}_f \tag{6.46}$$

Therefore, one can always write the differential equation

$$\vec{\nabla} \times \vec{H}(\vec{\nabla} \times \vec{A}) = \vec{J}_f \tag{6.47}$$

If the relation between $\vec{H}$ and $\vec{B}$ is not simple, the above equation may be difficult to solve.

For linear magnetic materials, though, the above reduces to

$$\vec{\nabla} \times \left(\frac{1}{\mu} \vec{\nabla} \times \vec{A} \right) = \vec{J_f} \tag{6.48}$$

If we further specify that μ is constant over some region, then in that region we have

$$\vec{\nabla} \times \left(\vec{\nabla} \times \vec{A} \right) = \vec{\nabla} \left(\vec{\nabla} \cdot \vec{A} \right) - \nabla^2 \vec{A} = \mu \vec{J_f} \tag{6.49}$$

Finally, if we specify $\vec{\nabla} \cdot \vec{A} = 0$, this simplifies to a component-by-component Poisson Equation:

$$\nabla^2 \vec{A} = -\mu \vec{J_f} \tag{6.50}$$

In principle, one can apply the same techniques as we used for solving Poisson's Equation in electrostatics to solve this component by component. Boundary conditions must be specified either directly (recall that we proved that if any one of $\vec{A}$, $\vec{B}$, $\widehat{n} \times \vec{A}$, or $\widehat{n} \times \vec{B}$ is specified at every point on the boundary, then the resulting field (though not necessarily the vector potential) is unique) or by matching using the conditions on the normal and tangential components at boundaries.

Another technical challenge associated with the above equation is that it only separates cleanly into component-by-component Poisson Equations in Cartesian coordinates. If the current distribution it not naturally represented in Cartesian coordinates (*e.g.*, even a simple circular current loop), then separation of variables may not be feasible. Method of images may work, or one may have to resort to other techniques or numerical solution. None of this technical complication takes away from the fact that there will be a unique solution for each component independently. The technical complication just makes it hard to actually obtain that solution.

Hard Ferromagnets ($\vec{M}$ fixed and $\vec{J}_f = 0$): Magnetostatic Scalar Potential

If there are no free currents, then $\vec{\nabla} \times \vec{H} = 0$ and we are assured that $\vec{H}$ can be derived from a magnetostatic scalar potential. Here, we use $\vec{B} = \mu_o \left(\vec{H} + \vec{M} \right)$ with $\vec{M}$ fixed. Then $\vec{\nabla} \cdot \vec{B} = 0$ gives

$$\vec{\nabla} \cdot \mu_o \left(\vec{H} + \vec{M} \right) = 0 \tag{6.51}$$

$$-\nabla^2 V_M + \vec{\nabla} \cdot \vec{M} = 0 \tag{6.52}$$

$$\nabla^2 V_M = -\rho_M \qquad \text{with} \quad \rho_M = -\vec{\nabla} \cdot \vec{M} \tag{6.53}$$

(note the canceling minus signs in the definitions!) where ρ_M is termed the *magnetostatic charge density*. Note the close similarity to the definition of the bound charge density $\rho_b = -\vec{\nabla} \cdot \vec{P}$ for dielectrics. This analogy also implies the existence of surface magnetostatic charge density $\sigma_M = \widehat{n} \cdot \vec{M}$, which will be related to the discontinuity in the normal gradient of V_M in a manner timilar to that for V. This equation can be solved by the standard techniques for solving Poisson's Equation.

Example 6.3: Uniformly Magnetized Sphere, Again

Let's apply the above kind of formalism for the uniformly magnetized sphere, which satisfies the hard ferromagnet condition. Again, $\vec{M} = M\,\hat{z}$. This implies $\rho_M = -\vec{\nabla}\cdot\vec{M} = 0$ and $\sigma_M = \hat{n}\cdot\vec{M} = M\cos\theta$. We solved this same problem before for the uniformly polarized dielectric sphere via separation of variables in spherical coordinates, which yielded Equation 4.15. Making the replacement $P/\epsilon_o \to M$ (because ρ/ϵ_o is replaced by ρ_M in Equation 6.53), we obtain

$$V_M(r \le R) = \frac{M\,z}{3} \qquad V_M(r \ge R) = \frac{\vec{m}\cdot\hat{r}}{4\,\pi\,r^2} \quad \text{with} \quad \vec{m} = \frac{4\,\pi}{3}\,\pi\,R^3\,\vec{M} \tag{6.54}$$

$$\vec{H} = -\vec{\nabla}V_M = \begin{cases} -\frac{\vec{M}}{3} & r \le R \\ \vec{H} \text{ field of a magnetic dipole } \vec{m} & r \ge R \end{cases} \tag{6.55}$$

$$\vec{B} = \mu_o\left(\vec{H} + \vec{M}\right) \quad \Longrightarrow \quad \vec{B}(r \le R) = \mu_o\left(-\frac{1}{3}\,\vec{M} + \vec{M}\right) = \frac{2}{3}\,\mu_o\,\vec{M} \tag{6.56}$$

$$\vec{B}(r \ge R) = \mu_o\,\vec{H} = \vec{B} \text{ field of a magnetic dipole } \vec{m} \tag{6.57}$$

This matches our previous solution for the magnetic field of this system that we obtained by calculating the vector potential of the bound surface current (Example 7.1).

Hard Ferromagnets ($\vec{M}$ fixed and $\vec{J}_f = 0$) via Vector Potential

We have already done this analysis, yielding Equations 6.10 and 6.11:

$$\vec{J}_b(\vec{r}) = \vec{\nabla} \times \vec{M}(\vec{r}) \qquad \vec{K}_b(\vec{r}) = \vec{M}(\vec{r}) \times \hat{n}(\vec{r})$$

$$\vec{A}(\vec{r}) = \frac{\mu_o}{4\pi} \int_{\mathcal{V}} d\tau' \, \frac{\vec{J}_b(\vec{r}')}{|\vec{r} - \vec{r}'|} + \frac{\mu_o}{4\pi} \oint_{\mathcal{S}(\mathcal{V})} da' \, \frac{\vec{K}_b(\vec{r}')}{|\vec{r} - \vec{r}'|}$$

We can, in fact, directly calculate the *field* from the bound currents using the Biot-Savart Law. The approach described above of using the magnetostatic scalar potential for such cases will in general be calculationally easier if the problem is amenable to the techniques for solving Poisson's Equation, but the Biot-Savart Law is certainly always guaranteed to work.

Example 6.4: Uniformly Magnetized Sphere, Again

We don't need to do this again: the above vector potential based on the bound current density (in this case, only a bound surface current density) is exactly how we solved this system before in Example 7.1.

No Free Currents, Linear Materials via Scalar Potential

If there are no free currents, then $\vec{\nabla} \times \vec{H} = 0$ and again we are assured that $\vec{H}$ can be derived from a magnetostatic scalar potential

$$\vec{H} = -\vec{\nabla} V_M(\vec{r}) \tag{6.58}$$

Again, if we know the relationship $\vec{B} = \vec{B}(\vec{H})$, then we can use the divergence equation:

$$\vec{\nabla} \cdot \vec{B} \left(-\vec{\nabla} V_M \right) = 0 \tag{6.59}$$

Again, if the relation between $\vec{H}$ and $\vec{B}$ is not simple, the above equation may be difficult to solve.

Again, though, for the case of linear magnetic materials, we have

$$\vec{\nabla} \cdot \left(\mu \, \vec{\nabla} \, V_M \right) = 0 \tag{6.60}$$

In a region where μ is constant, it can be passed through the divergence and we can reduce this to

$$\nabla^2 V_M = 0 \tag{6.61}$$

We now have Laplace's Equation. Again, boundary conditions and/or matching conditions will allow one to solve for V_M. In a region where μ is constant, we could equally well write $\vec{B} = -\vec{\nabla} U_M$ and solve $\nabla^2 U_M = 0$ with appropriate boundary conditions. In general, we will use V_M because its boundary conditions are simpler: the tangential gradient of U_M can have discontinuities due to bound surface currents, while we will see we can write boundary conditions on V_M without referencing the analogue for V_M, the bound magnetostatic surface charge density.

The importance of boundary conditions should be even more clear in such cases: since there is no source term in the equation, the boundary conditions *entirely* determine the solution.

Example 6.5: Magnetically Permeable Sphere in External Field

This is now a "soft," linear material, where we cannot take $\vec{M}$ to be fixed. But it is a situation with no free currents, so Laplace's Equation holds (except at the $r = R$ boundary, but we develop matching conditions there).

Fortunately, we do not need to solve the boundary value problem from scratch because this problem is directly analogous the case of a dielectrically polarizable sphere in an external electric field. We have the following correspondence:

$$\epsilon_o \vec{E} = -\epsilon_o \vec{\nabla} V \qquad \vec{H} = -\vec{\nabla} V_M \tag{6.62}$$

$$\epsilon_o \nabla^2 V = 0 \qquad \nabla^2 V_M = 0 \tag{6.63}$$

$$\vec{P} = \frac{\epsilon - \epsilon_o}{\epsilon_o} \vec{E} \qquad \vec{M} = \frac{\mu - \mu_o}{\mu_o} \vec{H} \tag{6.64}$$

$$\vec{D} = \epsilon_o \vec{E} + \vec{P} \qquad \vec{B}/\mu_o = \vec{H} + \vec{M} \tag{6.65}$$

$$\epsilon_o \vec{E} \overset{r\to\infty}{\longrightarrow} \epsilon_o \vec{E}_0 \qquad \vec{H} \overset{r\to\infty}{\longrightarrow} \vec{B}_0/\mu_o \tag{6.66}$$

$$\vec{D} \overset{r\to\infty}{\longrightarrow} \epsilon_o \vec{E}_0 \qquad \vec{B}/\mu_o \overset{r\to\infty}{\longrightarrow} \vec{B}_0/\mu_o \tag{6.67}$$

We have carefully avoided making correspondences in the above between ρ_b and ρ_M and between σ_b and σ_M because, in both cases, these quantities are not specified ahead of time: there is not permanent polarization, there is only polarization in response to applied field.

Let's also compare the matching conditions. We want to use the matching conditions that incorporate only the free charge densities because we do not know the bound charge densities ahead of time. For the electrostatic case, we used

$$\widehat{n} \cdot \left[\vec{D}_{>}(R) - \vec{D}_{<}(R) \right] = \sigma_f = 0 \tag{6.68}$$

$$\widehat{s} \cdot \left[\epsilon_o \, \vec{E}_{>}(R) - \epsilon_o \, \vec{E}_{<}(R) \right] = 0 \tag{6.69}$$

The corresponding matching conditions for the magnetic case are

$$\widehat{n} \cdot \left[\frac{\vec{B}_{>}(R)}{\mu_o} - \frac{\vec{B}_{<}(R)}{\mu_o} \right] = \frac{1}{\mu_o} \, \widehat{n} \cdot \left[\vec{B}_{>}(R) - \vec{B}_{<}(R) \right] = 0 \tag{6.70}$$

$$\widehat{s} \cdot \left[\vec{H}_{>}(R) - \vec{H}_{<}(R) \right] = \widehat{s} \cdot \left[\vec{K}_f \times \widehat{n} \right] = 0 \tag{6.71}$$

Thus, not only is there a perfect correspondence between fields, potentials, and $r \to \infty$ boundary conditions in the two problems, there is also a correspondence between matching conditions at $r = R$. Thus, we can just apply the solution to the electrostatic problem with the substitutions $\epsilon_o \, \vec{E} \to \vec{H}$, $\epsilon_o \, V \to V_M$, $\vec{P} \to \vec{M}$, and $\vec{D} \to \vec{B}/\mu_o$.

Applying this correpondence to Equations 4.70 and 4.71 gives us

$$V_M(r < R) = -\frac{3\,\mu_o}{2\,\mu_o + \mu}\,H_0\,z = -\frac{3\,\mu_o}{2\,\mu_o + \mu}\,\frac{B_0}{\mu_o}\,z \tag{6.72}$$

$$V_M(r > R) = -H_0\,z + \frac{\vec{m}\cdot\widehat{r}}{4\,\pi\,r^2} = -\frac{B_0}{\mu_o}\,z + \frac{\vec{m}\cdot\widehat{r}}{4\,\pi\,r^2} \tag{6.73}$$

$$\vec{m} \equiv \frac{4\,\pi}{3}\,R^3\,\vec{M}(r < R) = \frac{4\,\pi}{3}\,R^3\,H_0\,\frac{3\,(\mu - \mu_o)}{2\,\mu_o + \mu}\,\widehat{z} = \frac{4\,\pi}{3}\,R^3\,\frac{B_0}{\mu_o}\,\frac{3\,(\mu - \mu_o)}{2\,\mu_o + \mu}\,\widehat{z} \tag{6.74}$$

From the above, we calculate the fields and the magnetostatic surface charge density ($\rho_M = 0$ because $\vec{M}$ is uniform):

$$\vec{H}(r < R) = \frac{3\,\mu_o}{2\,\mu_o + \mu}\,\frac{\vec{B}_0}{\mu_o} = \frac{\vec{B}_0}{\mu_o} - \frac{\vec{M}(r < R)}{3} \tag{6.75}$$

$$\vec{M}(r < R) = 3\,\frac{\mu - \mu_o}{2\,\mu_o + \mu}\,\frac{\vec{B}_0}{\mu_o} \qquad\qquad \sigma_M = 3\,\frac{\mu - \mu_o}{2\,\mu_o + \mu}\,\frac{B_0}{\mu_o}\,\cos\theta \tag{6.76}$$

$$\begin{aligned}\vec{B}(r < R) &= \mu_o\left[\vec{H}(r < R) + \vec{M}(r < R)\right] = \mu_o\left[\frac{\vec{B}_0}{\mu_o} - \frac{\vec{M}(r < R)}{3} + \vec{M}(r < R)\right] \\ &= \vec{B}_0 + \frac{2}{3}\,\mu_o\,\vec{M}(r < R) = \left(\frac{3\,\mu}{2\,\mu_o + \mu}\right)\vec{B}_0\end{aligned} \tag{6.77}$$

Explicitly, we find that:

- Like $\vec{E}$, $\vec{H}$ is uniform inside the sphere and points in the direction of the uniform field. For $\chi_m > 0$, like for $\chi_e > 0$, it is smaller in magnitude than the uniform field at infinity.
- The magnetization density is in the direction of the uniform field for $\chi_m > 0$ as it was for $\vec{P}$ and $\chi_e > 0$.
- The magnetostatic surface charge density has a $\cos\theta$ dependence and is positive at the north pole for $\chi_m > 0$, as it was for the electrostatic surface charge density and $\chi_e > 0$.
- $\vec{B}$ is enhanced relative to the uniform field for $\chi_m > 0$. We did not calculate $\vec{D}$ in the electrostatic case, but we would have found that it, too, was enhanced relative to the uniform field.

We again see the fact that $\vec{H}$ corresponds to $\vec{E}$ and $\vec{B}$ to $\vec{D}$. In the electrostatic case, we noted how the field of the polarization counters the uniform field so that the total field inside the sphere is smaller in magnitude than the uniform field. That is true here too, but for $\vec{H}$, not for $\vec{B}$. $\vec{B}$ itself is enhanced inside the sphere! This difference in the behavior of the "true" fields arises directly from the above somewhat unexpected correspondence of $\vec{H}$ rather than $\vec{B}$ to $\vec{E}$.

There is a shortcut method that is much faster, so good to know from the point of view of technique. It makes the *ansatz* that the sphere magnetizes uniformly so then the total field is the superposition of a uniform field and a uniformly magnetized sphere (Equation 6.16). This assumption is made initially without relating $\vec{M}$ and $\vec{H}$. It then uses the relation $\vec{M} = \chi_m \vec{H}$ (equivalently, $\vec{B} = \mu \vec{H}$) to relate the two and solve for the fields.

The ansatz based on superposition gives

$$\vec{B}(r < R) = \vec{B}_{uniform} + \vec{B}_{sphere} = \vec{B}_0 + \frac{2}{3}\mu_o \vec{M} \tag{6.78}$$

$$\begin{aligned}\vec{H}(r < R) &= \vec{H}_{uniform} + \vec{H}_{sphere} = \vec{H}_{uniform} + \left(\frac{\vec{B}_{sphere}}{\mu_o} - \vec{M}_{sphere}\right) \\ &= \frac{\vec{B}_0}{\mu_o} - \frac{1}{3}\vec{M}\end{aligned} \tag{6.79}$$

Then we apply $\vec{B}(r < R) = \mu \vec{H}(r < R)$ to relate the above two equations and solve for $\vec{M}$. One finds one gets the same result. One can then calculate the field at $r \geq R$ from superposition. Admittedly, this technique is somewhat backhanded; when trying to understand things for the first time, reapplying the scalar potential to the full problem is more straightforward.

Lecture 19:

Magnetostatics in Matter III:
Boundary Value Problems in Magnetostatics (cont.)
Nonlinear Magnetic Materials

Date Revised: 2024/03/07 06:30
Date Given: 2024/03/07

Example 6.6: Magnetically Permeable Spherical Shell

Consider a spherical shell of inner radius a and outer radius b consisting of a highly permeable ($\mu/\mu_o \gg 1$) material placed in a uniform external field $\vec{B}_0$. We shall see that this shell shields its inner volume from the external field by a factor μ/μ_o. This technique is of great importance for magnetically sensitive experiments and equipment.

There are no free currents, so we may use the magnetostatic scalar potential technique. Furthemore, $\vec{\nabla} \cdot \vec{H} = 0$ in each region since μ is constant in each region. So the scalar potential V_M satisfies Laplace's Equation, allowing us to apply our techniques for the solution of Laplace's Equation from electrostatics.

In particular, given the azimuthal symmetry, we may assume the solution in each of the three regions is of the form given in Equation 3.115:

$$V_M(r < a, \theta) \equiv V_1(r, \theta) = \sum_{\ell=0}^{\infty} A_\ell \, r^\ell P_\ell(\cos\theta) \tag{6.80}$$

$$V_M(a < r < b, \theta) \equiv V_2(r, \theta) = \sum_{\ell=0}^{\infty} \left(C_\ell \, r^\ell + \frac{D_\ell}{r^{\ell+1}} \right) P_\ell(\cos\theta) \tag{6.81}$$

$$V_M(r > b, \theta) \equiv V_3(r, \theta) = -H_0 \, r \, \cos\theta + \sum_{\ell=0}^{\infty} \frac{E_\ell}{r^{\ell+1}} P_\ell(\cos\theta) \tag{6.82}$$

where we have already applied the requirements that V_M be finite as $r \to 0$ and that it yield the uniform field as $r \to \infty$ with $H_0 = B_0/\mu_o$. We have also assumed that V_M has no constant offset as $r \to \infty$.

There are no free currents, so our matching conditions are (as for the magnetically permeable sphere, Equations 6.70 and 6.71) that the normal component of $\vec{B}$ and the tangential component of $\vec{H}$ be continuous. Using $\vec{H} = -\vec{\nabla} V_M$, we thus have the four conditions

$$\mu_o \left.\frac{\partial V_1}{\partial r}\right|_a = \mu \left.\frac{\partial V_2}{\partial r}\right|_a \qquad \mu \left.\frac{\partial V_2}{\partial r}\right|_b = \mu_o \left.\frac{\partial V_3}{\partial r}\right|_b \tag{6.83}$$

$$\left.\frac{\partial V_1}{\partial \theta}\right|_a = \left.\frac{\partial V_2}{\partial \theta}\right|_a \qquad \left.\frac{\partial V_2}{\partial \theta}\right|_b = \left.\frac{\partial V_3}{\partial \theta}\right|_b \tag{6.84}$$

Note that we do not impose continuity on V_M. In the electrostatic case, we imposed continuity of V and the boundary condition on the normal derivative, ignoring continuity of the tangential derivative. In electrostatics, continuity of V comes from constructing it as the line integral of the electric field, which we in turn were motivated to write down in order to calculate the work done by the electric field on a point charge. Since $\vec{H}$ does not do such work, writing down the line integral is not physically motivated, though it is mathematically reasonable to do so because $\vec{H} = -\vec{\nabla} V_M$. So, here, we instead use continuity of the radial and tangential derivatives. This is an arbitrary choice driven by our physical intuition. We will see below that continuity of V_M would yield information redundant with tangential derivative continuity.

Before we dive into a lot of calculation, let's see what we can figure out without doing much work. The radial derivative equations only connect terms on the two sides of the equations with the same ℓ because they do not modify the orthonormal $P_\ell(\cos\theta)$. What about the angular derivative equations? Recall Equation 3.156:

$$P_\ell^m(x) = (-1)^m(1-x^2)^{m/2}\frac{d^m}{dx^m}P_\ell(x) \tag{6.85}$$

Let's write $\frac{\partial P_\ell(\cos\theta)}{\partial\theta}$ using this:

$$\frac{\partial P_\ell(\cos\theta)}{\partial\theta} = \frac{dP_\ell(\cos\theta)}{d\cos\theta}\frac{d\cos\theta}{d\theta} = \frac{P_\ell^1(\cos\theta)}{(-1)^1(1-\cos^2\theta)^{1/2}}(-\sin\theta) \tag{6.86}$$

$$= P_\ell^1(\cos\theta) \tag{6.87}$$

where we note that, since $0 < \theta < \pi$, there is no sign ambiguity and thus $\sin\theta = (1-\cos^2\theta)^{1/2}$. The $P_\ell^1(\cos\theta)$ are also orthonormal polynomials (the P_ℓ^m over all ℓ at fixed m form an orthonormal set in order for the $Y_{\ell m}$ to form an orthonormal set), so the same point we made above about the equations connecting terms at the same ℓ holds for these equations also. Note however that, for $\ell = 0$, the $\partial/\partial\theta$ matching condition yields zero.

Note also that, for $\ell \geq 1$, these equations are the same as one would have obtained by requiring continuity of V_M since $\partial/\partial\theta$ doesn't modify the radial factor of each term.

Taking the necessary derivatives for the radial derivative equations and then equating the two sides of all six equations (four for $\ell > 0$, only two for $\ell = 0$) term-by-term gives us:

$$\ell > 0: \qquad \mu_o \, \ell \, A_\ell \, a^{\ell-1} = \mu \, \ell \, C_\ell \, a^{\ell-1} - \mu \, (\ell+1) \, \frac{D_\ell}{a^{\ell+2}} \tag{6.88}$$

$$\mu \, \ell \, C_\ell \, b^{\ell-1} - \mu \, (\ell+1) \, \frac{D_\ell}{b^{\ell+2}} = -\mu_o \, H_0 \, \delta_{\ell 1} - \mu_o \, (\ell+1) \, \frac{E_\ell}{b^{\ell+2}} \tag{6.89}$$

$$A_\ell \, a^\ell = C_\ell \, a^\ell + \frac{D_\ell}{a^{\ell+1}} \tag{6.90}$$

$$C_\ell \, b^\ell + \frac{D_\ell}{b^{\ell+1}} = -H_0 \, b \, \delta_{\ell 1} + \frac{E_\ell}{b^{\ell+1}} \tag{6.91}$$

$$\ell = 0: \qquad 0 = -\mu \, \frac{D_0}{a^2} \qquad\qquad -\mu \, \frac{D_0}{b^2} = -\mu_o \, \frac{E_0}{b^2} \tag{6.92}$$

We explicitly write out the $\ell = 0$ equations because they yield qualitatively different conditions than the $\ell > 0$ terms.

For $\ell > 1$, solving for C_ℓ and D_ℓ results in both vanishing, so then A_ℓ and E_ℓ vanish for $\ell > 1$.

For $\ell = 0$, the radial derivative matching equations imply $D_0 = E_0 = 0$. We expect $E_0 = 0$ because it would yield a magnetic monopole potential for $r > b$, which we know is physically disallowed.

There are no equations that explicitly determine A_0 and C_0, which correspond to offsets of V_M for $r < a$ and $a < r < b$. We actually don't need to find them, since they do not affect $\vec{H}$ when the gradient is taken. (Recall, there is no issue of this potential being related to work or a potential energy, so we do not need to worry about discontinuities due to offsets.) But we can specify them by applying a restricted version of continuity of V_M, which is that we require V_M have the same offset in all regions. The lack of an offset for $r > b$ then implies $A_0 = 0$ and $C_0 = 0$.

For $\ell = 1$, we can do a lot of algebra to find explicit formulae for all the coefficients (you can find these in Jackson §5.12). These formulae are not particularly illuminating, but they become more intuitive when we take the limit $\mu/\mu_o \gg 1$. Inserting those coefficients into the solutions, we obtain

$$V_1(r,\theta) \overset{\frac{\mu}{\mu_o}\gg 1}{=} A_1\, r\, \cos\theta = -\frac{9}{2\,\frac{\mu}{\mu_o}\left(1-\frac{a^3}{b^3}\right)}\, H_0\, r\, \cos\theta \tag{6.93}$$

$$V_2(r,\theta) \overset{\frac{\mu}{\mu_o}\gg 1}{=} \left(C_1\, r + \frac{D_1}{r^2}\right)\cos\theta = -\frac{3}{\frac{\mu}{\mu_o}\left(1-\frac{a^3}{b^3}\right)}\, H_0 \left(r + \frac{1}{2}\,\frac{a^3}{r^2}\right)\cos\theta \tag{6.94}$$

$$V_3(r,\theta) \overset{\frac{\mu}{\mu_o}\gg 1}{=} \left(-H_0\, r + \frac{E_1}{r^2}\right)\cos\theta = H_0 \left(-r + \frac{b^3}{r^2}\left(1 - \frac{3\left(1+\frac{1}{2}\,\frac{a^3}{b^3}\right)}{\frac{\mu}{\mu_o}\left(1-\frac{a^3}{b^3}\right)}\right)\right)\cos\theta \tag{6.95}$$

Note that we include the term of order μ_o/μ in the $r > b$ solution so we can see that the matching condition on the tangential derivative at $r = b$ (equivalent to matching of V_M itself) is explicitly satisfied even in the limit $\mu/\mu_o \gg 1$.

Here are the resulting fields in the three regions:

$$\vec{H}_1(r,\theta) \overset{\frac{\mu}{\mu_o}\gg 1}{=} -\frac{9}{2}\,\frac{H_0\,\widehat{z}}{\frac{\mu}{\mu_o}\left(1-\frac{a^3}{b^3}\right)} \qquad \vec{B}_1(r,\theta)=\mu_o\,\vec{H}_1(r,\theta) \quad (6.96)$$

$$\vec{H}_2(r,\theta) \overset{\frac{\mu}{\mu_o}\gg 1}{=} \frac{3\,H_0\,\widehat{z}}{\frac{\mu}{\mu_o}\left(1-\frac{a^3}{b^3}\right)} + \frac{3\,(\vec{m}_a\cdot\widehat{r})\,\widehat{r}-\vec{m}_a}{4\,\pi\,r^3} \qquad \vec{B}_2(r,\theta)=\mu\,\vec{H}_2(r,\theta) \quad (6.97)$$

$$\vec{H}_3(r,\theta) \overset{\frac{\mu}{\mu_o}\gg 1}{=} H_0\,\widehat{z} + \frac{3\,(\vec{m}_b\cdot\widehat{r})\,\widehat{r}-\vec{m}_b}{4\,\pi\,r^3} \qquad \vec{B}_3(r,\theta)=\mu_o\,\vec{H}_3(r,\theta) \quad (6.98)$$

$$\text{with} \quad \vec{m}_a = -\frac{9}{2}\,\frac{H_0}{\frac{\mu}{\mu_o}\left(1-\frac{a^3}{b^3}\right)}\left(\frac{4\,\pi}{3}\,a^3\right)\widehat{z} \quad (6.99)$$

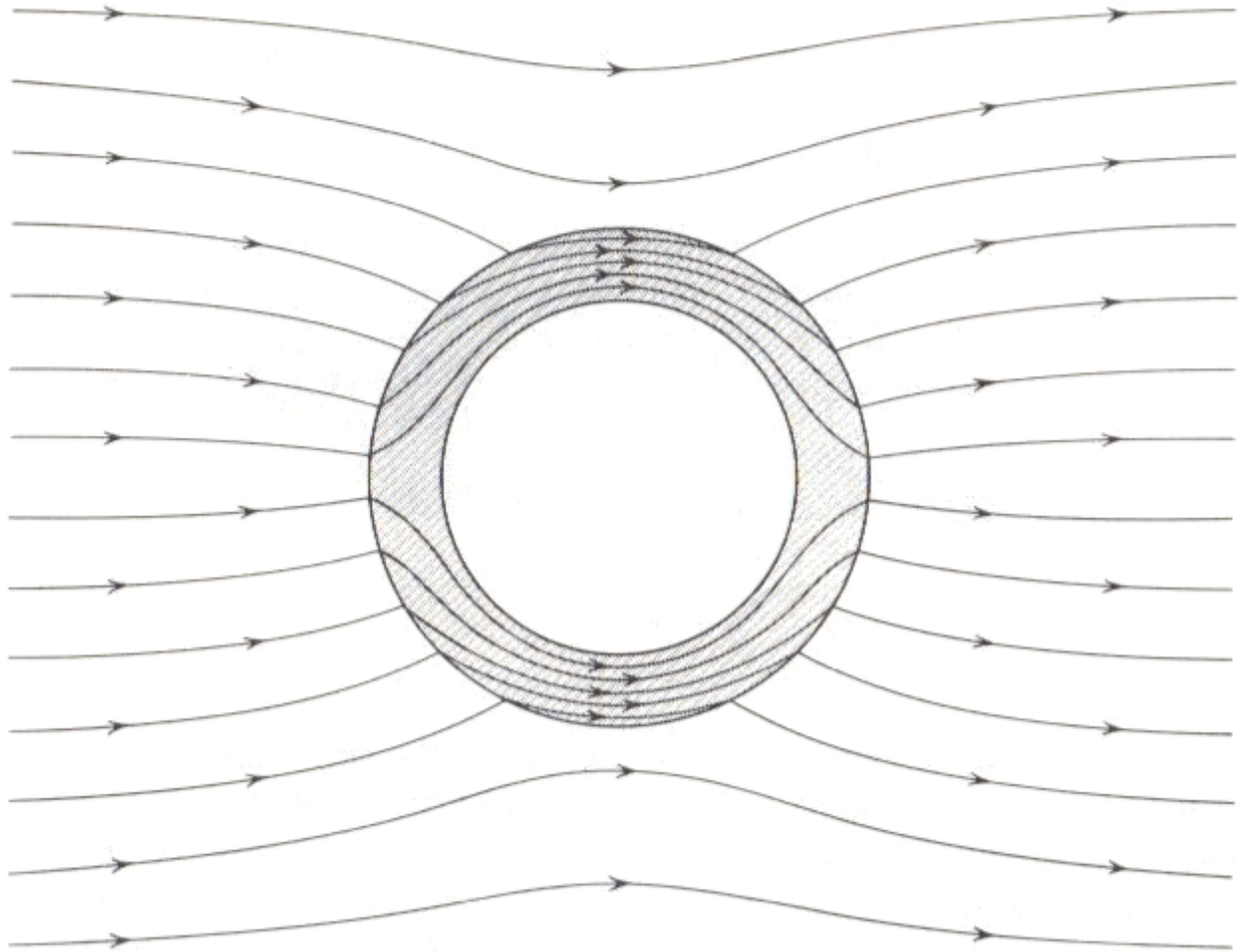
$$\vec{m}_b = 3\left[1-\frac{3\left(1+\frac{1}{2}\,\frac{a^3}{b^3}\right)}{\frac{\mu}{\mu_o}\left(1-\frac{a^3}{b^3}\right)}\right] H_0\left(\frac{4\,\pi}{3}\,b^3\right)\widehat{z} \quad (6.100)$$

It is not obvious but it is true that $\vec{m}_b$ incorporates $\vec{m}_a$, which is why there is no explicit contribution from $\vec{m}_a$ to the field at $r > b$. We will see this more clearly below.

The following features can be pointed out:

- Inside $r < a$, we have a uniform field weakened by a factor of μ/μ_o (for both B and H).
- In the permeable material, we have a uniform H field as well as a dipole field, but both are of order $(\mu_o/\mu)\,H_0$ (*i.e.*, attenuated) with the dipole moment pointed to $-\widehat{z}$. The dipole field cancels the uniform field at the poles at $r = a$ and adds to it at the equator.
- In the permeable material, the B field receives a factor of μ, so the B field receives uniform field and dipole contributions of order B_0 in the permeable material, though the vanishing at the poles at $r = a$ remains.
- One caveat to the above two statements is due to the $(1 - a^3/b^3)$ factor in the denominator of both terms (explicitly in the first term, hiding in $\vec{m}_a$ in the second term). If the shell is quite thin, then a/b is close to unity and this factor is much smaller than unity, resulting in an enhancement in both H_2 and B_2 by the geometric factor $(1 - a^3/b^3)^{-1}$. This factor accounts for the fact that magnetic field lines cannot be broken, and so the vast majority of the field lines that would have threaded through the $r < b$ region (a fraction $1 - \mu_o/\mu$ of them) now must flow entirely through the $a < r < b$ region: the factor is the ratio of the volume of the sphere of radius b to the volume of the shell.
- Finally, the field outside is the uniform field (for H and B) plus that of a dipole in the $+\widehat{z}$ direction. The dominant part of the dipole field cancels the uniform field at the equator at $r = b$, leaving a small residual field of order μ_o/μ smaller. At the poles, the dipole field adds to the uniform field, increasing the fields to $3\,H_0$ and $3\,B_0$ there.

Here is a picture from Jackson of $\vec{B}$. Note the concentration of field lines in the permeable material and their absence in the empty central region.



Let's now consider the analogy to electrostatic shielding. Electrostatic shielding is easily provided by conductors, and perfect conductors ($\epsilon/\epsilon_o \to \infty$) provide perfect electrostatic shielding. They do this by setting up surface charge that perfectly cancels the externally applied field *both in the conductor and in the cavity.*

The net effect for magnetostatic shielding effect is very similar — zero field in the cavity as $\mu/\mu_o \to \infty$ — but the physical cause is quite different, resulting in a *nonzero (and potentially high)* $\vec{B}$ field in the permeable material rather than zero field as in the electrostatic case. The fundamental reason is that magnetic field lines do not terminate: there are no magnetic monopoles. (The one exception is the field of the magnetic moment of a fundamental particle.) Therefore, the shielding effect is obtained by redirecting the field lines away from the cavity and into the permeable material, not by terminating the field lines. The high magnetic permeability of the materials causes magnetic dipoles to orient themselves such that they generate surface currents that cancel the externally applied field in the cavity (both $\vec{H}$ and $\vec{B}$) and enhance $\vec{B}$ in the material by the volume factor needed to accommodate all the field lines that would have gone through the cavity. In the limit of the equivalent of an electric conductor, with $\mu/\mu_o \to \infty$, the shielding would be perfect, as for an electric conductor.

All that said, we get a perfect analogy to an electric conductor if we instead look at the auxiliary field $\vec{H}$ and the magnetostatic surface charge density σ_M. $\vec{H}$ does get very small in the permeable material and in fact vanishes in the $\mu/\mu_o \to \infty$ limit. If we calculate the magnetostatic surface charge density σ_M from the discontinuity in $\vec{H}$ or $\vec{M}$, one would see that σ_M would look very much like σ_b for the case of a spherical shell of high dielectric susceptibility ($\epsilon/\epsilon_o \gg 1$) and, in the limit $\mu/\mu_o \to \infty$, σ_M would mimic the surface charge density of the electrical conductor limit (which is the same as $\epsilon/\epsilon_o \to \infty$). In particular, σ_M would only be nonzero at $r = b$: $\vec{H}$ would vanish for $r < a$ and $a < r < b$, so there could be no discontinuity in the radial derivative of $\vec{H}$ at $r = a$, which would occur if $\sigma_M \neq 0$ there. That is, in the magnetostatic scalar potential picture, the $\vec{H}$ field lines *do terminate* on the magnetostatic surface charge at $r = b$, yielding vanishing $\vec{H}$ for $r < b$. Of course, $\vec{H}$ is not a physical field, so the analogy is only mathematical, not physical.

Either way one does it, this calculation has important practical implications: such highly permeable materials are in widespread use for magnetic shielding from, especially, Earth's field in magnetically sensitive experiments and equipment such as SQUIDs (very sensitive magnetometers) and photomultiplier tubes (where the electrons' paths can be substantially bent and thus the gain modified by magnetic fields).

While we have benefited from our boundary-value problem techniques to get directly to the fields without having to calculate the bound surface currents, it would be nice to see how the bound surface currents give the observed fields. Recall Equation 5.84, which gives the bound surface current density from the change in the tangential component of the magnetic field:

$$\vec{K}(\vec{r}) = \frac{1}{\mu_o}\hat{n}(\vec{r}) \times \left[\vec{B}_>(\vec{r}) - \vec{B}_<(\vec{r})\right] \tag{6.101}$$

where $\hat{n}$ points from the $<$ region to the $>$ region. In our case, $\vec{K} = \vec{K}_b$ because there are no free currents. Since $\hat{n} = \hat{r}$ for our spherical surfaces and $\vec{B}$ only has components in the $\hat{r}$ and $\hat{\theta}$ directions, this reduces to

$$K_b(r)\,\hat{\phi} = \frac{1}{\mu_o}\left[B_{>,\theta} - B_{<,\theta}\right]\hat{\phi} = \frac{1}{\mu_o}\left[-\frac{\mu_>}{r}\left.\frac{\partial V_M}{\partial\theta}\right|_{r_>} + \frac{\mu_<}{r}\left.\frac{\partial V_M}{\partial\theta}\right|_{r_<}\right]\hat{\phi} \tag{6.102}$$

$$= \frac{\mu_< - \mu_>}{\mu_o}\frac{1}{r}\left.\frac{\partial V_M}{\partial\theta}\right|_r \hat{\phi} \tag{6.103}$$

where $<$ and $>$ indicate the two sides of the particular boundary at r and we use the fact that the tangential component of $\vec{H}$, which is given by $-(1/r)\,\partial V_M/\partial\theta$ here, is continuous and thus has the same value on both sides of the interface at r. So it is straightforward to calculate the surface currents given V_M.

We also know how to calculate $\vec{B}$ given surface currents derived from a uniform magnetization: we did it in our first calculation of the magnetic field of the permanently magnetized sphere (Equation 6.16) and saw (valid only for $\vec{K}_b \propto \widehat{\phi}\sin\theta$!)

$$\vec{B}_{\vec{M}}(r<R) = \frac{2}{3}\,\mu_o\,\vec{M} = \frac{2}{3}\,\mu_o\,\frac{\vec{K}_b\cdot\widehat{\phi}}{\sin\theta}\,\widehat{z}$$

$$\vec{B}_{\vec{M}}(r>R) = \frac{\mu_o}{4\,\pi}\,\frac{3\,(\vec{m}\cdot\widehat{r})\,\widehat{r}-\vec{m}}{r^3} \qquad \vec{m} = \frac{4}{3}\,\pi\,R^3\,\vec{M} = \frac{4}{3}\,\pi\,R^3\,\frac{\vec{K}_b\cdot\widehat{\phi}}{\sin\theta}\,\widehat{z}$$

where the relation between $\vec{M}$ and $\vec{K}_b$ comes from the definition of the bound surface current, $\vec{K}_b = \vec{M}\times\widehat{n} = M\,\widehat{z}\times\widehat{r} = \widehat{\phi}\,M\,\sin\theta$. Here, the $\vec{H}$ and $\vec{B}$ fields are uniform for $r<a$, and the fields are the sum of a uniform field and the field of a dipole for $a<r<b$. It thus looks like we have the superposition of a uniform magnetization in one direction for $r<a$ and a uniform magnetization in the opposite direction for $a<r<b$, with the two magnetizations having a fractional difference of magnitude of μ_o/μ. (The fact that we have permeable materials present is irrelevant for the calculation of $\vec{B}$: once one has all the bound currents, one can calculate $\vec{B}$ directly from them.) So, we expect that, in this case, we can just add the field of the above form due to the bound currents to the uniform applied field to get the total field in the three regions.

That is, we expect (again, valid only for $\vec{K}_a, \vec{K}_b \propto \widehat{\phi}\sin\theta$!)

$$\vec{B}_1(r,\theta) = \vec{B}_0 + \frac{2}{3}\,\mu_o\,\frac{\vec{K}_b(a)\cdot\widehat{\phi} + \vec{K}_b(b)\cdot\widehat{\phi}}{\sin\theta}\,\widehat{z} \tag{6.104}$$

$$\vec{B}_2(r,\theta) = \vec{B}_0 + \frac{2}{3}\,\mu_o\,\frac{\vec{K}_b(b)\cdot\widehat{\phi}}{\sin\theta}\,\widehat{z} + \frac{\mu_o}{4\,\pi}\,\frac{3\,(\vec{m}_a\cdot\widehat{r})\,\widehat{r} - \vec{m}_a}{r^3} \tag{6.105}$$

$$\vec{B}_3(r,\theta) = \vec{B}_0 + \frac{\mu_o}{4\,\pi}\,\frac{3\,(\vec{m}_b\cdot\widehat{r})\,\widehat{r} - \vec{m}_b}{r^3} \tag{6.106}$$

$$\vec{m}_a = \frac{4}{3}\,\pi\,a^3\,\frac{\vec{K}_b(a)\cdot\widehat{\phi}}{\sin\theta}\,\widehat{z} \qquad \vec{m}_b = \vec{m}_a + \frac{4}{3}\,\pi\,b^3\,\frac{\vec{K}_b(b)\cdot\widehat{\phi}}{\sin\theta}\,\widehat{z} \tag{6.107}$$

and then we can obtain $\vec{H}$ from the usual relation $\vec{H}(\vec{r}) = \vec{B}(\vec{r})/\mu(\vec{r})$. Note that we now see explicitly that $\vec{m}_b$ incorporates $\vec{m}_a$ as we had stated without proof above.

The explicit results for the bound surface currents are

$$\vec{K}_b(a,\theta) = \frac{\mu_< - \mu_>}{\mu_o}\frac{1}{a}\left.\frac{\partial V_M}{\partial \theta}\right|_a \widehat{\phi} = -\widehat{\phi}\left(\frac{\mu}{\mu_o}-1\right)\frac{9}{2\,\frac{\mu}{\mu_o}\left(1-\frac{a^3}{b^3}\right)}\,\frac{B_0}{\mu_o}\,\sin\theta$$

$$\overset{\mathcal{O}(\mu_o/\mu)^0}{\approx} -\widehat{\phi}\,\frac{9}{2\left(1-\frac{a^3}{b^3}\right)}\,\frac{B_0}{\mu_o}\,\sin\theta \tag{6.108}$$

$$\vec{K}_b(b,\theta) = \frac{\mu_< - \mu_>}{\mu_o}\frac{1}{b}\left.\frac{\partial V_M}{\partial \theta}\right|_b \widehat{\phi} = \widehat{\phi}\left(\frac{\mu}{\mu_o}-1\right)\frac{3\left(1+\frac{1}{2}\,\frac{a^3}{b^3}\right)}{\frac{\mu}{\mu_o}\left(1-\frac{a^3}{b^3}\right)}\,\frac{B_0}{\mu_o}\,\sin\theta$$

$$\overset{\mathcal{O}(\mu_o/\mu)^0}{\approx} \widehat{\phi}\,\frac{3\left(1+\frac{1}{2}\,\frac{a^3}{b^3}\right)}{\left(1-\frac{a^3}{b^3}\right)}\,\frac{B_0}{\mu_o}\,\sin\theta \tag{6.109}$$

We will explain the choice for the order of the approximation below.

There is an important subtletly in trying to do this calculation of surface currents using the approximate forms for the fields we have written down (valid for $\mu_o/\mu \ll 1$). We expect the magnetic field for $r < a$ to be of order $(\mu_o/\mu)\, B_0$. But $\vec{B}_0$ is in the expression for $\vec{B}_1$, so that implies the second term in that expression due to the surface currents will carry one term of order B_0 to cancel B_0 and then a second term of order $(\mu_o/\mu)\, B_0$ to give the residual field. As we explained above, our expressions for the contribution of the surface current to the field are of the following form for $r \lesssim b$:

$$B_K \sim \mu_o K \sim \pm\mu_o \frac{\mu - \mu_o}{\mu_o} \frac{1}{r} \frac{\partial V_M}{\partial \theta} \sim \mu_o \left(\frac{\mu}{\mu_o} - 1 \right) H_\theta \tag{6.110}$$

$$\sim \mu_o \left(\mathcal{O}\left(\frac{\mu}{\mu_o} \right)^1 + \mathcal{O}\left(\frac{\mu_o}{\mu} \right)^0 \right) \mathcal{O}\left(\frac{\mu_o}{\mu} \right)^1 H_0 \tag{6.111}$$

$$\sim \left[\mathcal{O}\left(\frac{\mu_o}{\mu} \right)^0 + \mathcal{O}\left(\frac{\mu_o}{\mu} \right)^1 \right] B_0 \tag{6.112}$$

(In the second line, we used $H \sim \mathcal{O}(\mu_o/\mu)^1 H_0$, which one can see from the expressions for $\vec{H}_1$ and $\vec{H}_2$. It is not so obvious that this is true for $\vec{H}_3$ at $r \sim b$, but it must be true because H_θ is continuous. It turns out to be true because the dipole field cancels the applied field to first order in H_0 (*i.e.*, zeroth order in μ_o/μ) at the equator, leaving a residual field of order $\mathcal{O}(\mu_o/\mu)^1 H_0$. The cancellation does not happen at the poles, but $H_\theta = 0$ at the poles.)

We now see the problem. The term that is $\mathcal{O}(\mu_o/\mu)^0$ will cancel the $\vec{B}_0$ term. So then the $\mathcal{O}(\mu_o/\mu)^1 B_0$ term is all that is left and is our full field, as expected. But we have not done the approximation self-consistently. We would have obtained a term of the same order by including terms $\mathcal{O}(\mu_o/\mu)^2$ in the expression for H because they would yield $\mathcal{O}(\mu_o/\mu)^1$ terms when multiplied by the $\mathcal{O}(\mu/\mu_o)^1$ term from the $(\mu/\mu_o - 1)$ prefactor. Without including that term, we will get the incorrect coefficient for the residual field.

We could have included that higher order term, but then we would run into the same problem at the next order: our calculation of the field using the surface currents would be correct to $\mathcal{O}(\mu_o/\mu)^1$, but our expression for the fields would have terms of order $\mathcal{O}(\mu_o/\mu)^2$ that we would not be able to fully reproduce. Given that it would be algebraically challenging to do this even to $\mathcal{O}(\mu_o/\mu)^1$ correctly, we punt on trying to calculate the residual field.

However, we can self-consistently check our results (Equations 6.96-6.100) to $\mathcal{O}(\mu_o/\mu)^0$, so let's do that because it will show us that the zeroth order field does vanish at $r < a$ and it will tell us interesting things for other regions.

First, for $r < a$, we have, *to zeroth order in* μ_o/μ,

$$\widehat{z} \cdot \vec{B}_1(r,\theta) \overset{\mathcal{O}(\mu_o/\mu)^0}{\approx} B_0 + \frac{2}{3}\frac{B_0}{1-\frac{a^3}{b^3}}\left[-\frac{9}{2} + 3\left(1+\frac{1}{2}\frac{a^3}{b^3}\right)\right] = 0 \tag{6.113}$$

$$\widehat{z} \cdot \vec{H}_1(r,\theta) = \frac{\vec{B}_1(r,\theta)}{\mu_o} \overset{\mathcal{O}(\mu_o/\mu)^0}{\approx} 0 \tag{6.114}$$

As expected, both the magnetic and auxiliary fields vanish to zeroth order in μ_o/μ inside the cavity.

For $a < r < b$, we have

$$\widehat{z} \cdot \vec{B}_2(r,\theta) \overset{\mathcal{O}(\mu_o/\mu)^0}{\approx} B_0 + \frac{2}{3} \frac{B_0}{1 - \frac{a^3}{b^3}} (3) \left(1 + \frac{1}{2} \frac{a^3}{b^3}\right) + \frac{\mu_o}{4\pi} \frac{3(\vec{m}_a \cdot \widehat{r}) \cos\theta - m_a}{r^3}$$

$$= \frac{3 B_0}{1 - \frac{a^3}{b^3}} + \frac{\mu_o}{4\pi} m_a \frac{3\cos^2\theta - 1}{r^3} \tag{6.115}$$

$$\widehat{z} \cdot \vec{H}_2(r,\theta) = \frac{\vec{B}_2(r,\theta)}{\mu} \overset{\mathcal{O}(\mu_o/\mu)^0}{\approx} 0 \tag{6.116}$$

$$\text{with} \quad \vec{m}_a \overset{\mathcal{O}(\mu_o/\mu)^0}{\approx} -\frac{4\pi}{3} a^3 \left(\frac{9}{2}\right) \frac{1}{1 - \frac{a^3}{b^3}} \frac{B_0}{\mu_o} \widehat{z} \tag{6.117}$$

The total magnetic field in the permeable material is of order B_0 because both terms shown are of order B_0. In the limit $a \ll b$, one recovers $3 B_0$ as we expect from the case of the permeable sphere (Equation 6.77 with $\mu_o/\mu \to 0$). The auxiliary field vanishes in the permeable material to order $(\mu_o/\mu)^0$ because one must divide the entire expression by μ to get H from B, which combines with the μ_o in the expression for B to give a prefactor of μ_o/μ that vanishes at the level of approximation we are considering. The vanishing of $\vec{H}$ in the permeable material is exact in the limit $\mu/\mu_o \to \infty$ because the above approximations become exact.

Finally, let's look at $r > b$, for which we obtain

$$\widehat{z} \cdot \vec{B}_3(r,\theta) \overset{\mathcal{O}(\mu_o/\mu)^0}{\approx} B_0 + \frac{\mu_o}{4\pi} \frac{3\left(\vec{m}_b \cdot \widehat{r}\right)\cos\theta - m_b}{r^3} \tag{6.118}$$

$$\text{with} \quad \vec{m}_b \overset{\mathcal{O}(\mu_o/\mu)^0}{\approx} \frac{4\pi}{3} \frac{1}{1-\frac{a^3}{b^3}} \frac{B_0}{\mu_o} \left[-\frac{9}{2} a^3 + b^3 (3)\left(1 + \frac{1}{2}\frac{a^3}{b^3}\right)\right] \widehat{z}$$

$$= \quad 3\left(\frac{4\pi}{3} b^3\right) \frac{B_0}{\mu_o} \widehat{z} \tag{6.119}$$

$$\widehat{z} \cdot \vec{H}_3(r,\theta) \quad = \quad \frac{\vec{B}_3(r,\theta)}{\mu_o} \tag{6.120}$$

One can see that the expressions for $\vec{B}_0$ and $\vec{m}_b$ match *to zeroth order in* μ_o/μ the results we obtained via the boundary value problem technique, Equations 6.96-6.100. The expression for H has the same form with B_0 replaced by H_0 and it also matches the expressions we obtained earlier, again to zeroth order in μ_o/μ.

So, in the end, we see that, to the level of approximation for which we can self-consistently do calculations, the fields we calculate from the surface currents match the fields that we used to calculate those surface currents.

Nonlinear Materials and Ferromagnetism

There are materials whose magnetic response is *nonlinear*. In such materials, in addition to the tendency of magnetic dipoles due to unpaired electrons to align with the applied magnetic field, these dipoles interact with each other in such a way as to prefer aligning with each other, too. This extra preference for magnetization causes the magnetization to depend nonlinearly on $\vec{H}$.

Beyond nonlinearity, there is the phenomenon of *ferromagnetism*, in which there are additional interactions that cause the magnetization to be preserved even after the applied field is reduced.

Both phenomena are caused by unpaired electrons as paramagnetism is; one might have guessed this by the fact that all three phenomena involve the alignment of magnetic dipoles with the applied field. The additional dipole-dipole interaction that causes nonlinearity is due to the *exchange effect* in quantum mechanics. We will explain this in the following.

The Exchange Effect in a Single Atom: Hund's Rules

As you know, electrons in atoms occupy *shells* corresponding to different energies for the electron-nucleus Coulomb interaction. Only a certain number of states are allowed for each shell (n^2 for shell n), and electrons can be put in a shell with spin "up" or spin "down" (multiplying by 2 the number of allowed states).

When a shell is partially filled, the electrons prefer to be *unpaired*, meaning that they have different orbital wavefunctions (probability distributions) and *the same spin direction (*i.e., *aligned spins)* rather than the same orbital wavefunction and different spin directions. This behavior, where electrons prefer to be in different orbitals but have the same spin, is frequently termed *Hund's Rules*.

The reason for this preference against having the same orbital wavefunction is that the electrostatic repulsive energy of two electrons in the same orbital state is high: in quantum mechanics, that energy is determined by the integral of the product of their wavefunctions weighted by $1/|\vec{r} - \vec{r}'|$ where $\vec{r}$ and $\vec{r}'$ are their positions, so the less similar their wavefunctions are, the lower the (positive) electrostatic repulsive energy is.

Next, considering the Pauli exclusion principle, the overall state (spatial wavefunction and spin) must be antisymmetric under exchange of the two electrons. One could achieve this by taking the spatial wavefunction to be antisymmetric and the spin state to be symmetric under exchange or vice versa. It turns out that, when one calculates the Coulomb repulsion energy integral, there is a second term that arises due to the extra terms created by requiring the overall state to be symmetric or antisymmetric under exchange. Moreover, because of the symmetry constraints on the overall state, this *exchange* term is negative when the spatial state is antisymmetric and positive when it is symmetric. Thus, the exchange term ensures that the antisymmetry of the overall state is preferred to be in the spatial wavefunction, not the spin state.

With the above, it would still be possible for the electrons to either have the same spin projection or for them to have opposite spin projections and the spin state to be the symmetric combination of the two possible anti-aligned states. It turns out that spin-orbit coupling causes the latter state to be higher energy, so the case in which the two electrons are aligned is preferred. Thus, we are able to explain Hund's Rules.

The Exchange Effect Among Atoms: Nonlinearity

In addition, though, one needs a mechanism for unpaired electrons in nearby atoms to align with each other; alignment of the unpaired electrons in a single atom is not enough. A similar exchange interaction is required, of which there are many types that depend on the details of the material and how the electrons in nearby atoms interact. The key requirement for such exchange effects to occur, though, is delocalization — the electron wavefunctions must be large enough that they spread to nearby atoms — so that there can be exchange interactions between electrons in adjacent atoms. This explains why nonlinearity occurs only in atoms with d- and f-shell electrons — the electrons in these orbitals are more weakly bound than s- and p-shell electrons, providing the necessary delocalization.

The exchange interaction leads to a nonlinear magnetic permeability, where these interactions cause the magnetic dipoles to prefer to align with each other macroscopically when they have been nudged into alignment by an applied field. This would yield a relationship of the form $\vec{B} = \vec{F}(\vec{H})$, which cannot be summarized by a simple constant of proportionality, but the relation is at least well-defined.

Interactions with the Crystal: Ferromagnetism

Ferromagnets have domains, which are regions of coaligned magnetic dipoles, caused by the aforementioned nonlinearity: it is energetically favorable for the magnetic moments of unpaired electrons in nearby atoms to align. By default, these domains are macroscopic in size (fractions of a mm, perhaps), but they do not align with each other because alignment would create a large magnetic field outside the material, which is not a low-energy state (which we will see when we talk about magnetic energy). When a large field is applied, though, the energy cost of not aligning with the magnetic field ($\vec{m} \cdot \vec{B}_{applied}$) is larger than the energy savings of not having a large field energy ($|\vec{B}_{material}|$), and so the domains align with the applied field.

We then must consider the phenomenon of *saturation*, whereby, at large fields, one gets to the point where *all* the unpaired electrons' dipole moments are aligned with the field and there are no more magnetic dipoles left to align. The magnetization density stops increasing and saturates. The applied field $\vec{H}$ may continue to be increased, but $\vec{B} = \mu_o(\vec{H} + \vec{M}_{sat})$, where $\vec{M}_{sat}$ is the saturated magnetization density. Thus, $\vec{B}$ increases due to the first term only. (At lower fields, $\vec{M}$ increases with $\vec{H}$, leading to a large amplification of $\vec{H}$ to yield $\vec{B}$.)

After bringing a nonlinear paramagnetic material into saturation, what happens when one turns off the applied field? The large exchange interaction energy makes it favorable for the moments to remain aligned with the direction of the applied magnetic field that has been removed. It is not that full alignment is the lowest-energy state, but that there is an energy barrier between the fully aligned state and the lower-energy state with randomly aligned domains. In fact, to reduce $\vec{M}$ and $\vec{B}$ to zero requires a significant $\vec{H}$ in the direction opposite to $\vec{B}$. After $\vec{B}$ goes through zero, it can then begin to align with $\vec{H}$ again and one can reach saturation in the other direction. And so on.

This phenomenon of the magnetization (and thus $\vec{B}$) being dependent on past history is called *hysteresis*: not only is $\vec{B}$ a nonlinear function of $\vec{H}$, but, in addition, $\vec{B}$ depends on the history of $\vec{H}$. Hysteresis curves are shown in Griffiths Figures 6.28 and 6.29.

The exchange phenomenon explains why ferromagnetics becomes less magnetized if they are dropped. The mechanical shock of dropping provides enough vibrational energy to exceed the barrier between the fully aligned state and the random domain state, allowing the domains to randomize in direction again.

We note that ferromagnets have a *Curie* or *transition temperature*, T_c. This temperature corresponds to roughly the exchange energy of nearby dipoles. When the temperature is larger than T_c, the thermal energy available overcomes the exchange energy, causing magnetic ordering to go away. If a saturated ferromagnetic is raised above T_c, the ordering will dissipate. Then, when recooled in zero applied field, randomly oriented domains will appear but there will be no overall ordering of the magnetic dipoles. Cooling in a high enough applied field, by contrast, will result in magnetic ordering and a permanent $\vec{M}$.

That is, if you heat the refrigerator magnet you have dropped many times past its T_c and then cool it slowly in a field large enough to obtain $\vec{M}_{sat}$, it will work well again!

There is not much more we can say about ferromagnetism without considering specific cases.

Section 7
Electrodynamics

Lecture 20:

Electrodynamics I:

Currents and Ohm's Law
Electromotive Forces

Date Revised: 2024/04/02 05:20
Date Given: 2024/04/02

Currents and Ohm's Law

Ohm's Law and Joule Heating: Differential Version

We state the very nonobvious point that the current due to an ensemble of flowing charges is proportional to the *force on a single charge* $\vec{f}$ acting on them:

$$\vec{J} \propto \vec{f} \tag{7.1}$$

Since current is proportional to velocity, and force is proportional to acceleration, why is this true? In an ideal conductor, it would not be true, we would expect current to be proportional to the integral of the force over time. But in all real conductors, there are two important effects that change this picture:

- The first is the random thermal motion of the charge carriers. The forces we can apply yield velocities that are small perturbations to this random thermal motion. So the mean *speed* of the carriers is dominated by the thermal speed $v_{thermal}$. This thermal motion is even larger than would be obtained from thermal equipartition, $mv_{thermal}^2/2 = 3k_B T/2$, because Fermi exclusion causes the electrons active in electrical conduction to have quite high energies (the Fermi energy). In Cu, for example, they have $T_F = 80,000$ K and $v_{thermal} = v_F = 1600$ km/s in Cu.
- The second is scattering, which is the cause of the randomness of the thermal motion. The charge carriers scatter off of impurities, defects, and thermal vibrations present in the material. This scattering is elastic in general, resulting in no loss of energy but in a redirection of velocity.

In the presence of such effects, our picture should not be of a charge carrier smoothly accelerating under the influence of an external force, but rather of a carrier with a large randomly directed velocity, scattering frequently, and with acceleration by the force between scatters resulting in a net motion in the direction of the electric force. The scattering randomly redirects the velocity, so the velocity due to the externally applied force is, on average, reset to zero after each collision. If the thermal speed is $v_{thermal}$ and the typical distance traveled between scatters is λ, then the time available for the externally applied force to accelerate a carrier between scatters is

$$t = \frac{\lambda}{v_{thermal}} \tag{7.2}$$

The *average* velocity acquired from the applied force during this time is

$$\vec{v}_{ave} = \frac{1}{2}\,\vec{a}\,t = \frac{1}{2}\,\frac{\vec{f}}{m}\,\frac{\lambda}{v_{thermal}} \tag{7.3}$$

This velocity is the average overall velocity because of the zeroing of the instantaneous velocity after each collision.

If we then use $\vec{J} = n\,q\,\vec{v}_{ave}$ where n is the number density of charge carriers and q is the charge per carrier, and we use $\vec{f} = q\,\vec{E}$, we then can write

$$\vec{J} = \left(\frac{n\,q^2\,\lambda}{2\,m\,v_{thermal}}\right)\vec{E} \qquad \Longrightarrow \qquad \boxed{\vec{J} = \sigma\,\vec{E} \qquad \sigma = \frac{n\,q^2\,\lambda}{2\,m\,v_{thermal}}} \tag{7.4}$$

Thus, we see our earlier expression is justified. This is *Ohm's Law*.

There is power dissipated in this process — the work done on the charge carriers by the electric field is lost to random motion when they scatter. The infinitesimal amount of energy lost per unit time dP in an infinitesimal volume $d\tau$ is equal to the work done by the electric field on the charge carriers:

$$dP = \textit{number density} \cdot \textit{velocity} \cdot \frac{\textit{force}}{\textit{carrier}}\,d\tau = n\,\vec{v}_{ave}\cdot\vec{f}\,d\tau = n\,\frac{\vec{J}}{n\,q}\cdot q\,\vec{E}\,d\tau = \vec{J}\cdot\vec{E}\,d\tau \tag{7.5}$$

This is known as *Joule Dissipation* or *Joule Heating*.

We note that the possibility of $\vec{E} \neq 0$ does not contradict our earlier discussions of conductors in electrostatics; here, we have non-stationary charges, where in that case we considered the final static situation after any currents had flowed.

Integral Version of Ohm's Law and Joule Heating

We integrate the above to obtain a more familiar version of Ohm's Law. We start with:

$$\mathrm{I} = \int_{\mathcal{S}} da\, \hat{n} \cdot \vec{J} = \int_{\mathcal{S}} da\, \sigma\, \hat{n} \cdot \vec{E} \tag{7.6}$$

Let's assume the cross-sectional area of the conductor is constant and the conductor is uniform. This lets us do the area integral trivially, yielding $\mathrm{I} = \sigma\, A\, \hat{n} \cdot \vec{E}$. If we then do a line integral directed along the wire, such that $d\vec{\ell} \propto \hat{n}$, we have

$$\mathrm{I}\ell = \int d\ell\, \mathrm{I} = \sigma\, A \int d\ell\, \hat{n} \cdot \vec{E} = \sigma A \int d\vec{\ell} \cdot \vec{E} = \sigma A\, V \tag{7.7}$$

$$\implies \quad V = \mathrm{I}\, R \qquad \text{with} \quad R = \frac{\ell}{A}\,\frac{1}{\sigma} \equiv \frac{\ell}{A}\,\rho \quad \text{and} \quad \rho = \frac{1}{\sigma} \tag{7.8}$$

which is the familiar version of Ohm's law in terms of current, voltage, and resistance. This is the *integral version of Ohm's Law* while $\vec{J} = \sigma\, \vec{E}$ is the differential (or local) version. We also define the resistivity ρ as the reciprocal of the conductivity σ. We can also integrate the Joule heating expression to get the usual integral expression for Joule heating:

$$P = \int_{\mathcal{V}} dP = \int_{\mathcal{V}} d\tau\, \vec{J} \cdot \vec{E} = \int_{\mathcal{S}} da \int d\ell\, \frac{\mathrm{I}}{A}\, \hat{n} \cdot \vec{E} = \mathrm{I}\, V = \mathrm{I}^2 R = \frac{V^2}{R} \tag{7.9}$$

Steady-State Assumption and Uniform Conductivity $\Longrightarrow$ Zero Charge Density

Do we need to worry about charge accumulation in conductors? Let's calculate the divergence of $\vec{E}$ to find the charge density, assuming uniform conductivity:

$$\vec{\nabla} \cdot \vec{E} = \frac{1}{\sigma} \vec{\nabla} \cdot \vec{J} = 0 \tag{7.10}$$

where the first step was possible by Ohm's Law and the assumed uniformity of the conductivity and the second step by the steady-state assumption on macroscopic scales. So the answer is no, as long as the conductivity is uniform and the system is steady-state, no charge density accumulates. *Note that this is not a circular argument: the steady-state assumption corresponded to* $\partial\rho/\partial t = 0$*, not* $\rho = 0$*. Now, with the combination of* $\partial\rho/\partial t = 0$ *and Ohm's Law, we conclude* $\rho = 0$*.*

Later, we will see how it is possible for charge to accumulate when we consider non-steady-state systems (in particular, with sinusoidal currents).

Note that our microscopic picture is not consistent with the steady-state assumption, but, averaged over time, our macroscopic picture is. On the microscopic scale, there are nonzero positive and negative charge density fluctuations away from the mean density associated with the nonuniform motion of the carriers (even on scales over which we can treat the carrier density as continuous), and then, of course, on the most microscopic scales, the smooth charge density consists of individual electrons that have some finite spatial extent.

Uniformity of Electric Field in a Uniform Wire

We implicitly assumed in proving the integral version of Ohm's Law above that the uniformity of the conductor implied that the field and thus the current were uniform over the cross-sectional area. We can prove this. We did not explicitly require that the electric field also be uniform with position along the wire, but we can prove that, too.

We define a uniform conductor to be one with uniform conductivity and uniform cross-sectional area.

We proved above that the charge density vanishes in a uniform conductor with steady currents. Therefore, the conductor satisfies Laplace's Equation. Dirichlet boundary conditions are set at the two ends of the conductor by the potential difference ΔV. We assume these equipotentials are (by connections to a battery) transverse to the wire axis at $z = 0$ and $z = \ell$. On the outer surface of the wire, $\vec{J} \cdot \widehat{n} = 0$ because no current flows out of the wire, which implies that $\vec{E} \cdot \widehat{n} = 0$, which provides a boundary condition on the normal gradient of the potential (a Neumann boundary condition). (Equivalently, this implies the charge density vanishes at the surface.)

We guess a solution that satisfies these boundary conditions,

$$V(\vec{r}) = \frac{\Delta V}{\ell} z \qquad \Longrightarrow \qquad \vec{E} = -\vec{\nabla} V = -\frac{\Delta V}{\ell} \widehat{z} \tag{7.11}$$

Note that we do not need the sinusoidal solutions from separation of variables here — we only need the linear solution (which we ignored in our discussion of separation of variables in linear coordinates). This will be of relevance for the homework, too!

This linear solution satisfies the boundary conditions — equipotential surfaces at $z = 0$ and $z = \ell$ and vanishing normal derivative at the outer surface (whose normal is always perpendicular to $\widehat{z}$) — and therefore it must be the solution.

Therefore, it is valid to assume that the field is uniform over the cross-sectional area of the wire *and* along the length of the wire if the wire is of fixed cross-sectional area, the conductivity is uniform, and the currents are steady-state. The latter two conditions told us Laplace's Equation is satisfied, while the former one provided the z-translation symmetry needed to guess the solution.

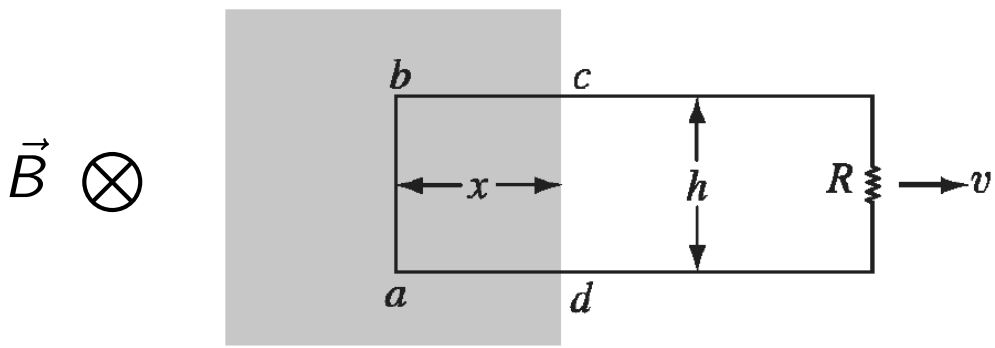
What happens to this argument if the wire changes in some way along its length; *e.g.*, the conductivity changes, or the wire diameter changes?

Motional Electromotive Force

We deviate from Griffiths somewhat in the introduction of electromotive forces; his §7.1.2 just seems confusing.

Moving Rectangular Loop

Consider a rectangular loop with a resistor in it with part of the loop's area intersecting a region of uniform magnetic field perpendicular to the loop *into the page*, as shown in the figure.



Let's consider the force on a charge carrier in the portion of the wire that intersects the field. If $\vec{v} = v\,\hat{x}$ and $\vec{B} = -B\,\hat{z}$ (into the page), then these charges feel a Lorentz force

$$\vec{F}_{mag} = (q\,v\,\hat{x}) \times (-B\,\hat{z}) = q\,v\,B\,\hat{y} \tag{7.12}$$

Since this force is aligned with the vertical portion of the wire, the carriers in that section can move. Assuming for the moment the charge carriers are positive (the argument can be reversed if they are negative), they would start to collect at b at the top end of the vertical portion and a deficit would appear at a at the bottom end. The local electrostatic repulsion between like charges would cause the charge carriers to start flowing through the rest of the circuit and would prevent this clumping of carriers. In this way, a current is generated around the loop *without the influence of a large-scale electric field in the circuit*. If the loop is pulled at constant speed, one satisfies the steady-state assumption, with no charge buildup.

There is work done on a charge carrier during its movement up the vertical portion of the wire

$$W_{ab} = \int_a^b d\vec{\ell} \cdot \vec{F}_{mag} = q\, v\, B\, h \tag{7.13}$$

We will see below that this work is not done by the Lorentz force as suggested above (recall, the Lorentz force can do no work because $\vec{F}_{mag} \perp \vec{v}$), but it is nevertheless done. The energy gained by the charge carriers via this work is dissipated as Joule heating in the resistor because the carriers quickly reach some steady-state velocity and a steady-state current flows.

We define the work done per unit charge on the charges as they move from a to b as the *motional electromotive force* or *motional emf*:

$$\mathcal{E} = \frac{W_{ab}}{q} = v\, B\, h \tag{7.14}$$

Let's think about how we can interpret $\mathcal{E}$. Let I be the current that flows. I is q times the number of charges that flow per unit time past a given point. Therefore $\mathrm{I}\,\mathcal{E}$ is the power being supplied to the ensemble of charges, the work done on them per unit time. By conservation of energy, it is also the power being dissipated in Joule heating in R. But we know that latter quantity is also $\mathrm{I}^2 R$. Equating the two, we see

$$\mathcal{E} = \mathrm{I}\,R \tag{7.15}$$

That is, $\mathcal{E}$ plays the role of voltage in Ohm's Law for the resistor. $\mathcal{E}$ has the right units for this purpose. In fact, if one attaches a voltmeter across the resistor R, it will report a voltage $V = \mathcal{E}$: a voltmeter works essentially by measuring the current in a very large resistor $R' \gg R$ placed in parallel with R, and the current that will flow through R' is identical to what would flow if a battery $\mathcal{E}$ were placed across R with R' in parallel. So, what appeared to just be a work done on a unit charge now can be interpreted as equivalent to a voltage! But be sure to remember that the current is generated by movement of the circuit in a magnetic field; it is not due to an electric field! We will return to the distinction between $\mathcal{E}$ and a voltage later when we consider electromagnetic induction and Faraday's Law.

Let's now think about what force is doing the work. As we discussed some time ago, the Lorentz force does no work because $\vec{F} \perp \vec{v}$. However, a force must pull the loop. There is a force counteracting this force that the pulling force must match to keep the loop at constant speed: the Lorentz force due to the velocity the carriers have acquired in the $\widehat{y}$ direction, which we will denote by $\vec{u} = u\,\widehat{y}$. This force is

$$-\vec{F}_{pull} = \vec{F}'_{mag} = q\,u\,\widehat{y} \times -B\,\widehat{z} = -q\,u\,B\,\widehat{x} \tag{7.16}$$

The total velocity of the charge carriers is

$$\vec{w} = \vec{v} + \vec{u} = v\,\widehat{x} + u\,\widehat{y} \tag{7.17}$$

The pulling force must cancel $\vec{F}'_{mag}$, so the work done per unit time by the pulling force is

$$\frac{dW_{pull}}{dt} = \vec{F}_{pull} \cdot \vec{w} = q\,u\,B\,\widehat{x} \cdot (v\,\widehat{x} + u\,\widehat{y}) = q\,u\,B\,v \tag{7.18}$$

Note that the charge carriers move on a diagonal line relative to the lab frame as they move from point a to point b on the wire, with this line partly in the direction of $\vec{F}_{pull}$. It takes the charge carriers a time $t = h/u$ to move on this trajectory since their $\widehat{y}$ direction speed is u. Therefore, the work done by $\vec{F}_{pull}$ during the movement of a charge from a to b is:

$$W_{pull} = \frac{dW_{pull}}{dt}\,\frac{h}{u} = q\,B\,v\,h \tag{7.19}$$

$$\Longrightarrow \qquad \frac{W_{pull}}{q} = B\,v\,h = \mathcal{E} \tag{7.20}$$

That is, the work done by the pulling force, per unit charge, matches the motional emf. The pulling force provides the energy that is eventually dissipated as heat as the carriers flow through the resistor.

Mechanically, how does this work? A magnetic field does no work, so it should only change the direction of the velocity of the charge carriers. So, initially, when the pulling force begins to act and the carriers start to move in the x direction and feel a Lorentz force in the y direction, their x velocity starts to be transformed into y velocity. But the loop is being pulled at constant speed $v\,\widehat{x}$, so the walls of the wire exert a force so the carriers' x velocity remains equal to $v\,\widehat{x}$ as the magnetic force acts. Similarly, as the carriers acquire a velocity in the y direction, they feel $\vec{F}'_{mag}$ in the $-\widehat{x}$ direction, and the walls of the wire must exert a force to keep them moving at $v\,\widehat{x}$ in the x direction. By Newton's third law, the charge carriers exert a reaction force on the walls of the wire, which would slow down the loop if there were not a force pulling it. Thus, we see it is the force pulling the loop that ultimately provides the work to drive the current.

And note: All this motion is accomplished without a large-scale electric field. Of course, it relies on the microscopic Coulomb repulsion between like charge carriers and the Coulomb binding to the wire that keeps the charge carriers from flying out of the wire.

Returning to the emf itself, we can rewrite it in a useful form. We define the *magnetic flux* to be the integral of $\vec{B}$ dotted into the normal to a surface over the surface:

$$\Phi = \int_{\mathcal{S}} da\, \hat{n} \cdot \vec{B}(\vec{r}) \tag{7.21}$$

Using the definition of x in the figure, we have in this case

$$\Phi = B\, h\, x \tag{7.22}$$

The time derivative is

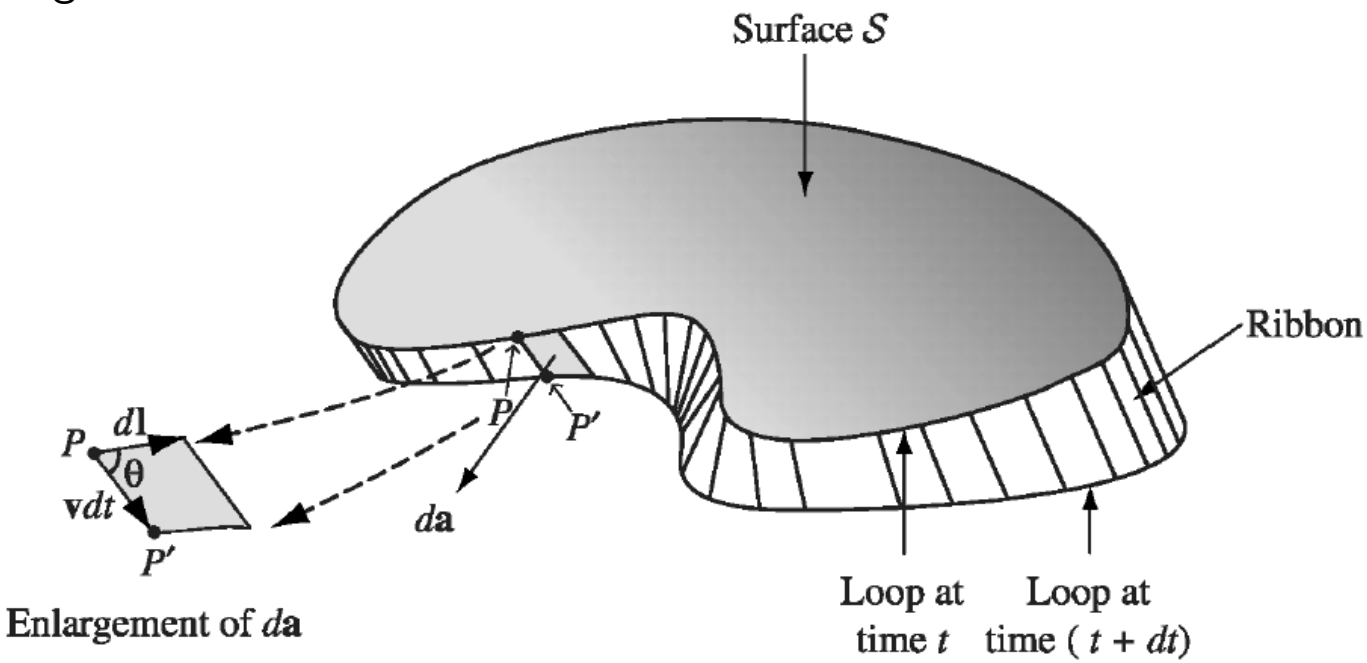
$$\frac{d\Phi}{dt} = B\, h\, \frac{dx}{dt} = -B\, h\, v \tag{7.23}$$

(x decreases with time for $v > 0$) which is just the negative of the motional emf. That is, we have

$$\mathcal{E} = -\frac{d\Phi}{dt} \tag{7.24}$$

Moving Arbitrary Loop

Let's prove rigorously that this rule holds more generally for any shape of loop with any type of motion through an arbitrary magnetic field. Consider the motion of a closed loop of arbitrary shape over a time dt. The loop is defined by a contour $\mathcal{C}(t)$ that depends on t. Each point on the loop has a velocity $\vec{v}$ that may depend on the position on the loop. Regardless, each piece of the loop moves by the vector $\vec{v}\,dt$ during this time where $\vec{v}$ is position-dependent. The charges in that piece of the loop acquire a velocity $\vec{u}$ along the direction of the loop due to the action of the Lorentz force during that time.



We can see, through some work, that the motion $\vec{v}$ also describes the change in the area of and flux through the loop. The flux changes by

$$d\Phi = \int_{\mathcal{S}(\mathcal{C}(t+dt))} da\, \hat{n}_{\mathcal{C}} \cdot \vec{B} - \int_{\mathcal{S}(\mathcal{C}(t))} da\, \hat{n}_{\mathcal{C}} \cdot \vec{B} \tag{7.25}$$

We subscript $\hat{n}$ with $\mathcal{C}$ to distinguish it from a different $\hat{n}$ we define below. Let's rewrite this expression in a more usable form. Consider a closed surface that consists of the surfaces defined by $\mathcal{C}(t)$ and $\mathcal{C}(t+dt)$ as well as the ribbon-like surface connecting the two contours. (If the two contours were circular loops, the ribbon-like surface would be the wall of the cylinder formed by the two contours.) $\vec{\nabla} \cdot \vec{B} = 0$ tells us the surface integral of $\hat{n}_{\mathcal{S}} \cdot \vec{B}$ (where $\hat{n}_{\mathcal{S}}$ is the outward surface normal, identical to $\hat{n}_{\mathcal{C}}$ for only some parts of the surface) through this surface vanishes. That surface integral is related to the above integrals by

$$0 = \oint_{\text{closed } \mathcal{S}} da\, \hat{n}_{\mathcal{S}} \cdot \vec{B} = -\int_{\mathcal{S}(\mathcal{C}(t+dt))} da\, \hat{n}_{\mathcal{C}} \cdot \vec{B} + \int_{\mathcal{S}(\mathcal{C}(t))} da\, \hat{n}_{\mathcal{C}} \cdot \vec{B} + \int_{\text{ribbon}} da\, \hat{n}_{\mathcal{S}} \cdot \vec{B} \tag{7.26}$$

where we have used the fact that $\hat{n}_{\mathcal{S}} = -\hat{n}_{\mathcal{C}}$ on the $\mathcal{S}(\mathcal{C}(t+dt))$ surface and $\hat{n}_{\mathcal{S}} = \hat{n}_{\mathcal{C}}$ on the $\mathcal{S}(\mathcal{C}(t))$ surface. (The direction of $\hat{n}_{\mathcal{C}}$ is set by the direction of $d\vec{\ell}$ in the figure and the right-hand rule.) The negative sign is present in the former because the orientation of $\hat{n}_{\mathcal{C}}$ that maintains its direction as the contour moves has $\hat{n}_{\mathcal{C}}(t+dt)$ on this surface pointing into the enclosed volume rather than outward.

The first two terms give $-d\Phi$, so

$$d\Phi = \int_{ribbon} da\, \hat{n}_{\mathcal{S}} \cdot \vec{B} \tag{7.27}$$

The area element on the ribbon (with outwardly directed normal as in the above integral) is given by

$$\hat{n}_{\mathcal{S}}\, da = d\vec{r} \times d\vec{\ell} \tag{7.28}$$

where: $d\vec{\ell}$ is the line element along $\mathcal{C}(t)$ with orientation set by consistency with $\hat{n}_{\mathcal{C}}$ for $\mathcal{S}(\mathcal{C}(t))$ and the right-hand rule; and $d\vec{r}$ is the change in the vector position of that line element between t and $t + dt$. The difference between these two positions is related to $\vec{v}$, $d\vec{r} = \vec{v}\, dt$, so:

$$\hat{n}_{\mathcal{S}}\, da = \vec{v}\, dt \times d\vec{\ell} \tag{7.29}$$

$$\text{Therefore,} \quad d\Phi = \oint_{\mathcal{C}(t)} \left(\vec{v}\, dt \times d\vec{\ell}\right) \cdot \vec{B} \tag{7.30}$$

We turned an area integral into a line integral, but it still calculates magnetic flux.

Since $\vec{u} \parallel d\vec{\ell}$, we can add $\vec{u}$ to $\vec{v}$ to obtain $\vec{w}$ without affecting the integral:

$$d\Phi = \oint_{\mathcal{C}(t)} \left(\vec{w}\, dt \times d\vec{\ell} \right) \cdot \vec{B} \tag{7.31}$$

Using the cyclic property of the triple vector product, reversing the resulting cross product $\vec{B} \times \vec{w}$, and moving dt to the left side, we obtain

$$\frac{d\Phi}{dt} = - \oint_{\mathcal{C}(t)} d\vec{\ell} \cdot \left(\vec{w} \times \vec{B} \right) \tag{7.32}$$

The quantity $\vec{w} \times \vec{B}$ is just the Lorentz force per unit charge:

$$\frac{d\Phi}{dt} = - \oint_{\mathcal{C}(t)} d\vec{\ell} \cdot \frac{\vec{F}_{mag}}{q} \tag{7.33}$$

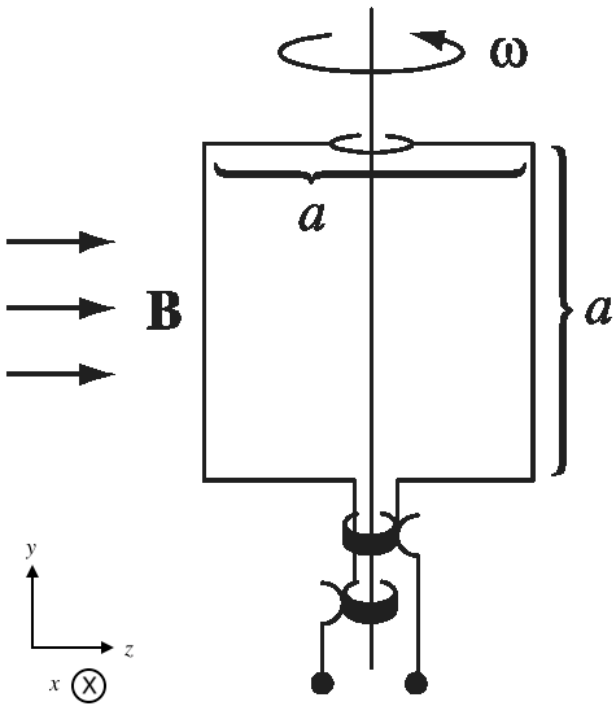
The integral of the Lorentz force per unit charge integrated around the loop is the generalization for arbitrary loops of our earlier expression for the motional emf for the rectangular loop (earlier, we integrated over only the section of length h from a to b of the rectangular loop for which $\vec{F}_{mag}$ was nonzero), so

$$\boxed{\mathcal{E} = \oint_{\mathcal{C}(t)} d\vec{\ell} \cdot \frac{\vec{F}_{mag}}{q} = -\frac{d\Phi}{dt} \qquad \text{moving circuit}} \tag{7.34}$$

The motional emf, as defined by the line integral of the Lorentz force per unit charge around the loop, is given by the negative of the rate of change of the magnetic flux through the loop. The signs of the line integral and the flux are set by requiring that the orientation of the line integral (via $d\vec{\ell}$) be consistent via the right-hand rule with the orientation of the surface normal $\widehat{n}_{\mathcal{C}}$ used for the flux calculation.

Example 7.1: Alternating Current Generator (Griffiths 7.10)

The classic and pervasive use of the above relationship is the alternating current generator. Consider a square loop placed in a uniform magnetic field and rotated about a midline at constant angular speed ω. That is, the rotation is such that, at one point of the motion, the magnetic field is normal to the loop while, one fourth of the period before or after this time, the magnetic field is in the plane of the loop. What is the motional emf around the loop generated by this motion?



The magnetic field is constant, so the flux is just given by B times the area of the loop projected onto the direction of $\vec{B}$:

$$\Phi(t) = a^2\,\vec{B}\cdot\hat{n}(t) = A\,B\,\cos\omega t \qquad (7.35)$$

where we have chosen $\hat{n} \parallel \vec{B}$ at $t = 0$ and written $a^2 = A$. Thus, the motional emf is

$$\mathcal{E}(t) = -\frac{d\Phi}{dt} = A\,B\,\omega\,\sin\omega t \qquad (7.36)$$

This is of course how 60-Hz AC voltage is generated.

It is instructive to think again about the Lorentz force experienced by the charge carriers in the loop and see how it generates the motional emf. Let the magnetic field be $\vec{B} = B\,\hat{z}$ and let the axis of rotation be $+\hat{y}$. Suppose the loop is just moving past having $\hat{n} = -\hat{x}$, as shown in the figure. Then the carriers all have a velocity parallel to $\pm\hat{x}$ due to the motion of the loop (this is $\vec{v}$). (They also have motion in the $\hat{z}$ direction, but this is parallel to $\vec{B}$ and thus no Lorentz force is generated.) The carriers in the sections of the loop parallel to $\hat{z}$ (perpendicular to the axis of rotation, parallel to the field) cannot move in response to this force because they feel a force in the $\hat{y}$ direction, transverse to the section of wire they are in. Those in the parts of the loop parallel to $\pm\hat{y}$ (parallel to axis of rotation) also feel a force along $\hat{y}$, and they can move along $\hat{y}$. As the loop turns away from this orientation, the arm at $+\hat{z}\,a/2$ has velocity in the $+\hat{x}$ direction and vice versa for the arm at $-\hat{z}\,a/2$. Positive charge carriers in these arms feel forces in the $-\hat{y}$ and $+\hat{y}$ directions, respectively. This forces a current to flow in direction defined by the $-\hat{n} = +\hat{x}$ orientation by right-hand rule, generating a field through the loop in the $-\hat{n} = +\hat{x}$ direction.

As the loop passes through this orientation, the flux is zero and is changing from negative ($\hat{n}\cdot\vec{B} < 0$) to positive ($\hat{n}\cdot\vec{B} > 0$). One can see that the driven current is in the direction needed for its field to counter the change in magnetic flux. This is a manifestation of *Lenz's Law*, which we will return to later.

If one taps the loop as is typical for such a generator, as shown in the figure, the tap connected to the $+\widehat{z}\,a/2$ arm will have positive voltage and the tap connected to the $-\widehat{z}\,a/2$ arm will have negative voltage because they need to drive a current in an external circuit that carries current in the direction consistent with that argued above, from the $+\widehat{z}\,a/2$ arm to the $-\widehat{z}\,a/2$ arm.

Note the polarity of the above statement: we decide the sign of the voltage at the taps not by what is needed to drive the current in the loop (which is driven by the Lorentz force, not by this voltage) but rather by the sign needed to drive the current in the external load (the resistor) so that current exits the loop, goes through the load, and returns to the loop, where it is needed to conserve charge.

Lecture 21:

Electrodynamics II:

Electromagnetic Induction
Faraday's Law
Inductance

Date Revised: 2024/04/09 07:00
Revised lecture break
Date Given: 2024/04/04

Electromagnetic Induction

Faraday's Law

We are going to consider three different physical situations:

- *Moving loops*: As we considered above, the magnetic field is stationary but the loop is moving.
- *Moving magnetic fields*: The loop is held fixed but the magnetic field is changing because the currents sourcing the field are being translated.
- *Changing magnetic fields:* Both the loop and the sources of the field are stationary, but the currents sourcing the field are changing.

We just proved using the Lorentz Force Law that the first situation results in a *motional emf*: a force that causes the flow of a current around the loop, given by Equation 7.34:

$$\mathcal{E} = \oint_{\mathcal{C}(t)} d\vec{\ell} \cdot \frac{\vec{F}_{mag}}{q} = -\frac{d\Phi}{dt} \qquad \text{moving circuit} \tag{7.37}$$

Faraday's Law consists of the *empirical observation* that the same rule applies for the second and third situations. The subtlety is this: this law could not have been derived using the Lorentz Force applied to the situation described above of a fixed loop and a moving and/or changing magnetic field: there is no magnetic force if the charge carriers are not being forced to move in the magnetic field by the loop being pulled. A natural and important corollary is that the emf that appears for a moving or changing magnetic field is not due to a magnetic force. Rather, since the loop is at rest in the second and third situations, the force that appears arises from a true *electric field*. Mathematically, we write Faraday's Law as

$$\mathcal{E} = \oint_{\mathcal{C}(t)} d\vec{\ell} \cdot \frac{\vec{F}_{elec}}{q} = -\frac{d\Phi}{dt} \qquad \text{moving or changing magnetic field} \qquad (7.38)$$

We see that it is identical in form to the Lorentz Force law applied to a moving loop with the replacement of $\vec{F}_{mag}$ by $\vec{F}_{elec}$.

Combining the two forms, and defining $\vec{E}_{ind} = \vec{F}_{elec}/q$ where $_{ind}$ indicates that the electric field here is not an electrostatic one due to Coulomb's Law but rather an "induced" field due to the changing magnetic flux, we then may write a common law that applies in any situation:

$$\boxed{\mathcal{E} = \oint_{\mathcal{C}(t)} d\vec{\ell} \cdot \left[\vec{E}_{ind} + \frac{\vec{F}_{mag}}{q} \right] = -\frac{d\Phi}{dt} = -\frac{d}{dt} \int_{\mathcal{S}(\mathcal{C}(t))} da\, \hat{n}(\vec{r}, t) \cdot \vec{B}(\vec{r}, t)} \quad (7.39)$$

If there is any ambiguity in the sign, one should apply *Lenz's Law*: the emf has a sign such that the polarity of the current it would drive produces a magnetic field that counters the change in magnetic field. We will prove Lenz's Law explicitly later.

Quasistatic Assumption

Note that we have implicitly assumed in our derivations that the current everywhere in the loop responds instantaneously to the total emf on the left side, that there is no time delay between a buildup of charge at one point in the circuit and the driving of a current around the loop. We made the same assumption in deriving Ohm's Law. This is the "quasistatic assumption," that all fields and currents everywhere change instantaneously and that information is propagated infinitely quickly. Formally, this assumption consists of saying that, given a typical physical scale for a system ℓ and a typical timescale for variation t, we have

$$t \gg \ell/c \tag{7.40}$$

where c is the *speed of light* that will be defined later.

We will release this assumption when we discuss electromagnetic waves and radiation.

Motional EMF, Faraday's Law, Galilean Relativity, and Galilean Field Transformations

When first proposed, Faraday's Law was an empirical observation. However, it could have been justified using the principle of Galilean relativity: physics is the same in all inertial reference frames, those moving at constant velocity.

Consider the problem of the magnetic field moving at fixed velocity. One could go to the rest frame of the magnetic field and consider the loop to be moving at fixed velocity as in our moving loop cases. The magnetic force implied by the motional emf appears. In Galilean relativity, forces are invariant upon change of inertial (fixed velocity) frame. This would imply that the magnetic force in the field-fixed frame is still present in the loop-fixed frame, but now we interpret it as an electric force because the loop is not moving.

In the case of changing magnetic fields, we simply have to invoke the expectation that the loop has no way of knowing whether it experiences a changing field because the current sourcing the field is moving or because it is changing: it only knows about the field that results, not the source of the field.

This Galilean relativity argument was, however, not recognized until after Faraday's observation.

We can make use of this argument to understand how electric and magnetic fields mix with each other under such Galilean (nonrelativistic) transformations. Let's assume we have written down our law, Equation 7.39, in both the rest frame of the loop and in the lab frame in which the loop is moving. The fields and position vectors in the loop rest frame are given $'$ symbols, the ones in the lab frame have no primes. The total emf can be determined explicitly using a voltmeter to measure the voltage across the resistor in the loop, and it is a scalar that is independent of frame (the reading on the voltmeter doesn't change if you see the voltmeter moving with the loop!). So we can equate the lab and rest frame expressions through $\mathcal{E}$:

$$\oint_{\mathcal{C}'} d\vec{\ell}' \cdot \vec{E}'_{ind} = \oint_{\mathcal{C}(t)} d\vec{\ell} \cdot \left[\vec{E}_{ind} + \frac{\vec{F}_{mag}}{q} \right] \tag{7.41}$$

($\mathcal{C}' = \mathcal{C}(t = 0)$ can be assumed by appropriate choice of when the lab and loop rest frame coordinate systems coincide). Now, let's use our expression for the magnetic force term from our derivation of Equation 7.34, dropping the $\vec{u}$ contribution that we had added in:

$$\oint_{\mathcal{C}'} d\vec{\ell}' \cdot \vec{E}'_{ind} = \oint_{\mathcal{C}(t)} d\vec{\ell} \cdot \left[\vec{E}_{ind} + \vec{v} \times \vec{B} \right] \tag{7.42}$$

Since the circuit is arbitrary, we may thus conclude

$$\vec{E}'_{ind} = \vec{E}_{ind} + \vec{v} \times \vec{B} \tag{7.43}$$

The equation can be taken to be completely general because adding a standard electrostatic field to both sides would leave the statement true while accounting for such electrostatic fields:

$$\boxed{\vec{E}\,' = \vec{E} + \vec{v} \times \vec{B} \qquad \vec{E} = \vec{E}_{Coul} + \vec{E}_{ind}} \tag{7.44}$$

Therefore, this is a rule for how electric fields transform from one frame to another under Galilean relativity, regardless of the source of the field. Electric fields are not the same in a fixed and a moving frame if magnetic fields are present, even before special relativity is considered! Special relativity then *only* adds correction coefficients to the above equation.

It is important to note that the expectation that the electrostatic fields do not depend on frame has been an assertion so far, based on the assumption that Coulomb's Law is unaffected by whether the charges are moving or not. We will return to this point later in connection to Maxwell's Equations, as it will lead to a symmetrization of the above equation between $\vec{E}$ and $\vec{B}$.

Galilean relativity is consistent with the quasistatic assumption. We need only consider special relativity when the nonzero travel time of light becomes important because special relativity says the speed of light is the same in all frames.

Example 7.2: A Stationary Alternating Current Generator

Recall the previous example of an AC generator that used a rotating square loop in a constant magnetic field. Instead, hold the loop fixed but assume that the magnetic field is being varied sinusoidally, $\vec{B}(t) = \vec{B}_0 \cos \omega t$. Then the flux is

$$\Phi(t) = A\,\vec{B}(t) \cdot \widehat{n} = A\,B_0\,\cos\omega t \tag{7.45}$$

Therefore, the emf generated is

$$\mathcal{E}(t) = -\frac{d\Phi}{dt} = A\,B_0\,\omega\,\sin\omega t \tag{7.46}$$

just as before.

Note, again, the polarity of the emf! As before, the emf's polarity is such that it causes current to flow in an external resistor attached to the two ends of the circuit in a direction consistent with the current that flows in the loop.

Something one has to be careful about is incorrectly believing that, because of the emf's sign, it should also drive a current in the zero-resistance *loop* in the direction implied by the emf. That erroneous belief arises because one is assuming the electric field is conservative, that the integral of $\vec{E}$ around a loop vanishes. No: that sign of emf would drive current in the wrong direction! For the current flowing in the loop, the emf measures the work per unit charge done by $\vec{E}_{ind} + \vec{F}_{mag}/q$ as they push the current around the loop, but they are not pushing the charges down an electrostatic potential! The *effect* of having this current flow is that the same current flows through the resistor, creating an apparent potential drop across the resistor that we can measure with a voltmeter. But the voltmeter is just measuring the current flowing through a known resistance, which, by Ohm's Law, is proportional to the line integral of the electric field through the resistor. The voltmeter's ability to measure something that looks like a voltage does not imply that an electrostatic potential can be defined everywhere in the loop and resistor!

In thinking about what causes the current to flow, it is better to visualize the electric field: one recognizes that the changing magnetic field generates an electric field that pushes current in the direction it needs to flow to counter the change in magnetic field. *This electric field has nonzero loop integral around the circuit!* Therefore, the existence of the emf $\mathcal{E}$ at the ends of the circuit does not imply the same emf is experienced by the current flowing in the loop itelf; the nonzero loop integral of the electric field invalidates the rule that the total voltage drop around a loop must vanish, which is the source of the misconception that $\mathcal{E}$, appearing at the ends of the circuit, is also the driver of the current in the loop.

Differential Version of Faraday's Law

Consider the special case of an arbitrary closed contour $\mathcal{C}$ fixed in space. Equation 7.39 tells us

$$\oint_{\mathcal{C}} d\vec{\ell} \cdot \vec{E} = -\frac{d}{dt} \int_{\mathcal{S}(\mathcal{C})} da\, \hat{n}(\vec{r}) \cdot \vec{B}(\vec{r}, t) \tag{7.47}$$

Let's use Stokes' Theorem on the left side, and, since the contour is time-independent, we can move the time derivative inside the integral on the right side. We turn it into a partial derivative to make it clear that we do not need to worry about any possible time-dependence of $\vec{r}$ (of which there is none here). This yields

$$\oint_{\mathcal{S}(\mathcal{C})} da\, \hat{n}(\vec{r}) \cdot \left[\vec{\nabla} \times \vec{E}(\vec{r})\right] = -\int_{\mathcal{S}(\mathcal{C})} da\, \hat{n}(\vec{r}) \cdot \frac{\partial \vec{B}(\vec{r}, t)}{\partial t} \tag{7.48}$$

Since the loop is arbitrary, the integrands must be equal:

$$\boxed{\vec{\nabla} \times \vec{E}(\vec{r}) = -\frac{\partial \vec{B}(\vec{r}, t)}{\partial t}} \tag{7.49}$$

This differential version of Faraday's Law is the generalization of $\vec{\nabla} \times \vec{E} = 0$ for time-dependent situations. We now explicitly see what was said in the previous example: a changing $\vec{B}$ creates a nonconservative electric field!

Biot-Savart and Ampere's Law for the Induced Electric Field in the Absence of Charges

If we consider the special case of no charge density, then we have

$$\vec{\nabla} \cdot \vec{E} = 0 \qquad \vec{\nabla} \times \vec{E} = -\frac{\partial \vec{B}}{\partial t} \tag{7.50}$$

This is mathematically identical to the equations of magnetostatics,

$$\vec{\nabla} \cdot \vec{B} = 0 \qquad \vec{\nabla} \times \vec{B} = \mu_o \vec{J} \tag{7.51}$$

In magnetostatics, we saw that the above two equations, combined with the assumption $\vec{\nabla} \cdot \vec{A} = 0$, yielded Poisson's Equation for $\vec{A}$ with $\mu_o \vec{J}$ as the source (Equation 5.56). By correspondence, we can thus state

$$\boxed{\vec{E} = \vec{\nabla} \times \vec{A}_E \qquad \nabla^2 \vec{A}_E = \frac{\partial \vec{B}}{\partial t} \qquad \vec{\nabla} \cdot \vec{A}_E = 0} \tag{7.52}$$

This is of course very interesting: we see that $\vec{E}$ receives a contribution from a vector potential that satisfies Poisson's Equation with $\partial \vec{B}/\partial t$ as the source!

Now, if we assume appropriate boundary conditions — fields falling off at infinity, no other surfaces on which the vector potential or field are specified — then we know from Equation 5.56 that the solution to the Poisson's Equation for $\vec{A}_E$ is

$$\boxed{\vec{A}_E(\vec{r}) = -\frac{1}{4\,\pi}\int_{\mathcal{V}} d\tau' \frac{\frac{\partial \vec{B}(\vec{r}')}{\partial t}}{|\vec{r}-\vec{r}'|}} \tag{7.53}$$

Finally, we may take the curl of the above expression to recover the analogue of the Biot-Savart Law. We did this backwards in the case of magnetostatics: we started with the empirical Biot-Savart Law and derived that the field could be written as the curl of the form of the vector potential corresponding to the above. Nevertheless, that proof could be reversed, so we may conclude that the analogous Biot-Savart Law holds (compare to Equation 5.32)

$$\boxed{\vec{E}(\vec{r}) = -\frac{1}{4\,\pi}\int_{\mathcal{V}} d\tau' \frac{\frac{\partial \vec{B}(\vec{r}')}{\partial t}\times(\vec{r}-\vec{r}')}{|\vec{r}-\vec{r}'|^3} = -\frac{1}{4\,\pi}\frac{\partial}{\partial t}\int_{\mathcal{V}} d\tau' \frac{\vec{B}(\vec{r}')\times(\vec{r}-\vec{r}')}{|\vec{r}-\vec{r}'|^3}} \tag{7.54}$$

where we pulled the time derivative outside the integral under the assumption that the volume itself is time-independent.

We also note that, because $\vec{E}$ satisfies the analogue of Ampere's Law, one can apply standard Ampere's Law techniques for finding $\vec{E}$ when $\frac{\partial \vec{B}}{\partial t}$ is given.

Caution: We have made the *quasistatic* assumption, that all time derivatives are small enough that the propagation time for disturbances in the magnetic fields is much less than the timescales on which the field vary. This is what allows us to use the magnetostatic formulae in time-varying situations. If the time derivatives become large, then one needs the full formalism of electromagnetic waves, which we will develop later.

Poisson's Equation for Induced Electric Field, Proof of Lenz's Law

Further pursuing the analogy to magnetostatics, let's see what we get if take the curl of the curl of the induced electric field:

$$\vec{\nabla} \times \left(\vec{\nabla} \times \vec{E}_{ind} \right) = \vec{\nabla} \times \left(-\frac{\partial \vec{B}}{\partial t} \right) = -\frac{\partial}{\partial t} \left(\vec{\nabla} \times \vec{B} \right) = -\mu_o \frac{\partial \vec{J}}{\partial t} \tag{7.55}$$

We may rewrite the left side using the vector identity for the curl of the curl as we did when deriving Poisson's Equation for $\vec{A}$ in terms of $\vec{J}$ (Equation 5.56):

$$\vec{\nabla} \left(\vec{\nabla} \cdot \vec{E}_{ind} \right) - \nabla^2 \vec{E}_{ind} = -\mu_o \frac{\partial \vec{J}}{\partial t} \tag{7.56}$$

If we again assume no charge density (valid since we are considering only the induced electric field $\vec{E}_{ind}$) and that the currents are localized so the fields fall off appropriately at infinity, we have a Poisson's Equation for $\vec{E}_{ind}$, whose solution we know:

$$\boxed{\nabla^2 \vec{E}_{ind} = \mu_o \frac{\partial \vec{J}}{\partial t} \quad \overset{\text{localized currents}}{\Longleftrightarrow} \quad \vec{E}_{ind}(\vec{r}) = -\frac{\mu_o}{4\pi} \int_{\mathcal{V}} d\tau' \, \frac{\frac{\partial \vec{J}(\vec{r}\,')}{\partial t}}{|\vec{r} - \vec{r}\,'|}} \tag{7.57}$$

Because of the vector alignment of $\vec{E}_{ind}$ and $-\partial \vec{J}/\partial t$, we thus have *Lenz's Law*: the induced electric field is in the direction needed to drive a current to counter the change in the current that is causing the changing magnetic field.

Furthermore, using Equation 5.56, the relation between $\vec{A}$ and $\vec{J}$ obtained by solving the Poisson Equation $\nabla^2 \vec{A} = -\mu_o \vec{J}$, we have

$$\vec{E}_{ind}(\vec{r}) = -\frac{\partial}{\partial t} \frac{\mu_o}{4\,\pi} \int_{\mathcal{V}} d\tau' \, \frac{\vec{J}(\vec{r}\,')}{|\vec{r} - \vec{r}\,'|} = -\frac{\partial \vec{A}}{\partial t} \tag{7.58}$$

which is a rather remarkable statement — this is an almost trivial sourcing equation for $\vec{E}_{ind}$, certainly much simpler than the sourcing equation involving $\partial \vec{J}/\partial t$ or the Biot-Savart Law for $\vec{E}_{ind}$ sourced by $\partial \vec{B}/\partial t$.

We will return to this contribution to the electric field when we reconsider our scalar and vector potentials after writing down Maxwell's Equations and in light of the fact that Faraday's Law implies the electric field is not curl-free and thus cannot be derived from a scalar potential.

Example 7.3: Induced Electric Field for Coaxial Conductors (Griffiths 7.16)

An alternating current $\mathrm{I} = \mathrm{I}_0 \cos \omega t$ flows down a long straight wire of negligible radius and returns along a coaxial conducting tube of radius a and negligible thickness. Both conductors are assumed to be perfect (infinite conductivity). We want to find the induced electric field as a function of the transverse radius s in cylindrical coordinates.

For reasons that we will be able to explain later when we discuss EM waves in the presence of conductors, the currents flow in sheets at the surfaces of the conductors because they have infinite conductivity.

In the region between the wire and the outer conductor, the field of the wire is the usual $\vec{B}(s,t) = \widehat{\phi}\,\mu_o\,\mathrm{I}(t)/2\,\pi\,s$. The magnetic field of the return-current cylinder is zero inside (consider an Amperian loop in the xy-plane with radius $s < a$: none of the return current flows through the surface enclosed by that loop). Outside the return-current sheet, its magnetic field is that of a wire carrying the total return current, which has the same magnitude but opposite sign of the field of the inner wire. Thus, the total magnetic field is the inner conductor's magnetic field between the conductors and is zero outside the outer conductor.

The system has azimuthal and z-translation symmetry, so the induced electric field must have the form $\vec{E} = E_s(s)\,\widehat{s} + E_\phi(s)\,\widehat{\phi} + E_z(s)\,\widehat{z}$.

If we think about what kind of Amperian loop has a nonzero flux of $\partial\vec{B}/\partial t$ (not $\vec{J}$!), it is a loop in the sz plane with normal in the $\widehat{\phi}$ direction. Let's first consider a loop of this kind with one z leg at infinity and the other at $s > a$. The contributions to the loop integral of the electric field along the two radial legs cancel, and the contribution from the leg at infinity vanishes assuming the fields fall off as $s \to \infty$, so this loop only gets a contribution from the z leg at finite radius, which picks out $E_z(s > a)$. The enclosed flux of $\partial\vec{B}/\partial t$ vanishes, so we can conclude $E_z(s > a) = 0$.

Now, repeat with one z leg at s between 0 and a and one z leg outside the outer conductor. The radial legs cancel and the z leg outside the outer conductor contributes nothing. When calculating the enclosed flux of $\partial\vec{B}/\partial t$, a similar thing holds: there is no magnetic field outside a, so the area integral only goes from s to a. If the loop's z dimension is ℓ, we have

$$E_z(s < a)\,\ell = -\int_0^\ell dz \int_s^a ds' \,\frac{\partial B_\phi(s', t)}{\partial t} = -\frac{\mu_o}{2\,\pi}\,\frac{\partial \mathrm{I}}{\partial t}\,\ell \int_s^a \frac{ds'}{s'} \tag{7.59}$$

$$= \frac{\mu_o}{2\,\pi}\,\omega\,\mathrm{I}_0\,\ell\,\sin\omega t\,\ln\frac{a}{s} \tag{7.60}$$

$$\implies E_z(s < a) = \frac{\mu_o}{2\,\pi}\,\omega\,\mathrm{I}_0\,\sin\omega t\,\ln\frac{a}{s} \tag{7.61}$$

Note the sign: taking the loop normal to be $\widehat{\phi}$ implies that the z leg with the nonzero contribution yields a positive contribution. Then the usual minus sign enters, which is cancelled by the sign of the derivative of $\cos\omega t$.

We can see $E_\phi(s)$ vanishes by using a loop in the $s\phi$ plane that has radial legs (ϕ constant) and azimuthal legs (s constant). One azimuthal leg can be taken to infinity so it yields no contribution, and the radial legs' contributions cancel, leaving only the contribution from the azimuthal leg at finite radius. But, unlike the E_z case, this loop has no magnetic flux and thus no $\partial\vec{B}/\partial t$ through it, so $E_\phi(s) = 0$.

Finally, consider E_s, which we can show to vanish by both a conceptual and a mechanical argument. As we argued above, E_s can be a function of s only and must be independent of z. Suppose E_s points outward along $\widehat{s}$ at a particular s and consider $E_s(s, z = 0)$. If we rotate the system about this direction by 180°, then the current changes direction. But $E_s(s, z = 0)$ cannot change direction (sign) — it is tied to the current distribution. Yet the reversal of the direction of the current changes the sign of $\vec{B}$ and thus $\partial\vec{B}/\partial t$. Then, by the Biot-Savart Law for $\vec{E}$, $\vec{E}$ should change sign. We have a contradiction unless $E_s(s, z = 0) = 0$. Because of z-translation symmetry, the same must hold at any z.

More mechanically, consider the Biot-Savart integral for $\vec{E}$. Given that $\vec{B}$ and $\partial\vec{B}/\partial t$ are both proportional to $\widehat{\phi}$, the vector $\vec{r} - \vec{r}\,'$ must have a piece proportional to $\widehat{z}$ to yield a contribution to the $\widehat{s}$ component of $\vec{E}$. But $\vec{B}$ is independent of z, while the $\widehat{z}$ component of $\vec{r} - \vec{r}\,'$ is odd about $z = z'$. So the integrand is odd about $z = z'$, causing the integral to vanish.

Thus, $$\vec{E}(s < a, t) = \widehat{z}\,\frac{\mu_o}{2\,\pi}\,\omega\,\mathrm{I}_0\,\sin\omega t\,\ln\frac{a}{s} \qquad \vec{E}(s > a, t) = 0 \qquad (7.62)$$

There are many other approaches to this problem.

- We could have derived the same $\vec{E}$ by first determining $\vec{A}$ and taking its time derivative. We calculated $\vec{A}$ of a wire in Ph106b Problem Set 7:

$$\vec{A}(s) = -\frac{\mu_o \mathrm{I}}{2\pi} \ln \frac{s}{a} \tag{7.63}$$

  The vector potential vanishes for $s > a$ because the vector potential of the return current cancels the above vector potential of the wire. If we take $\mathrm{I} = \mathrm{I}_0 \cos \omega t$ and calculate $\vec{E} = -\partial \vec{A}/\partial t$, we get the same result as we obtained already.

- We could have also used the same Poisson's Equation solution form for the Poisson's Equation in which $\partial \vec{J}/\partial t$ sources $\vec{E}$.

- Lastly, instead of using the Ph106b PS7 solution for $\vec{A}$, we could have obtained $\vec{A}$ from $\vec{B}$ using the same types of loops we used to obtain $\vec{E}$ from $\partial \vec{B}/\partial t$.

No matter how one finds $\vec{E}$, one can see the sign makes sense. This $\vec{E}$ tries to drive a current parallel or antiparallel to the current already flowing in the wire. When the current is decreasing, the electric field is increasing to try to drive a current in the same direction in which current is being lost by the decreasing current. It tries to generates a magnetic field that would compensate for the magnetic field that is being removed by the decreasing central conductor current. And vice versa for an increasing current.

Note how, when we can calculate the induced electric field directly, there is no ambiguity about the direction the driven current would flow, unlike for $\mathcal{E}$.

Mutual Inductance

Except for the example we just did, We have so far considered magnetic fields and fluxes in the abstract, without any concern about where they come from. But they are generated by currents, so it is natural to want to connect the Faraday's Law emf to changing currents. We do that through mutual inductance.

Consider two circuits $\mathcal{C}_1$ and $\mathcal{C}_2$. Suppose a current I_1 is flowing in $\mathcal{C}_1$. The magnetic flux at $\mathcal{C}_2$ is

$$\Phi_{21} = \int_{\mathcal{S}(\mathcal{C}_2)} da_2\, \hat{n}_2 \cdot \vec{B}_1(\vec{r}_2) = \int_{\mathcal{S}(\mathcal{C}_2)} da_2\, \hat{n} \cdot \left[\vec{\nabla} \times \vec{A}_1(\vec{r}_2)\right] = \oint_{\mathcal{C}_2} d\vec{\ell}_2 \cdot \vec{A}_1(\vec{r}_2) \quad (7.64)$$

where we used the fact that $\vec{B}$ is derived from a vector potential followed by Stokes' Theorem. Now, let's use the relation between $\vec{A}_1$ and the current in $\mathcal{C}_1$ using the usual solution of the Poisson's Equation for $\vec{A}_1$ (assuming appropriate boundary conditions):

$$\Phi_{21} = \frac{\mu_o}{4\,\pi} \oint_{\mathcal{C}_2} d\vec{\ell}_2 \cdot \oint_{\mathcal{C}_1} \frac{\mathrm{I}_1 d\vec{\ell}_1}{|\vec{r}_2 - \vec{r}_1|} = \frac{\mu_o}{4\,\pi} \mathrm{I}_1 \oint_{\mathcal{C}_2} \oint_{\mathcal{C}_1} \frac{d\vec{\ell}_2 \cdot d\vec{\ell}_1}{|\vec{r}_2 - \vec{r}_1|} \quad (7.65)$$

We rewrite this as follows:

$$\boxed{\Phi_{21} = M_{21}\,\mathrm{I}_1 \qquad M_{21} = \frac{\mu_o}{4\,\pi}\oint_{\mathcal{C}_2}\oint_{\mathcal{C}_1}\frac{d\vec{\ell}_2\cdot d\vec{\ell}_1}{|\vec{r}_2-\vec{r}_1|}} \qquad \text{Neumann Formula} \qquad (7.66)$$

where M_{21} is the mutual inductance between $\mathcal{C}_1$ and $\mathcal{C}_2$ and has units of Henries (volt-second/amp). Two important characteristics:

- $M_{21} = M_{12}$ because the definition is symmetric.
- M_{21} is a completely geometric quantity: it does not care about the amount of current flowing, just on the relative positions of the two contours. It is like the capacitance matrix in this respect.

We may now take the time derivative to calculate the emf at $\mathcal{C}_2$ due to a change in I_1:

$$\boxed{\mathcal{E}_2 = -\frac{d\Phi_{21}}{dt} = -M_{21}\,\frac{d\mathrm{I}_1}{dt}} \qquad (7.67)$$

If unclear, the sign should be chosen to satisfy Lenz's Law.

Self-Inductance

The above derivation works even when $\mathcal{C}_1$ and $\mathcal{C}_2$ are identical: a current loop induces an emf on itself. In practice, calculating the integral can be difficult because of the singularity at $\vec{r}_1 = \vec{r}_2$, but one can be assured that self-inductance exists and is not infinite. The symbol used is L and the corresponding equations are

$$\boxed{\Phi = L\,\mathrm{I} \qquad L = \frac{\mu_o}{4\,\pi}\oint_{\mathcal{C}}\oint_{\mathcal{C}}\frac{d\vec{\ell}_2 \cdot d\vec{\ell}_1}{|\vec{r}_2 - \vec{r}_1|} \qquad \mathcal{E} = -L\,\frac{d\mathrm{I}}{dt}} \tag{7.68}$$

In both the cases of mutual inductance and self-inductance, one rarely does the integral directly. Instead, one tries to find the field using Ampere's Law, then calculate the flux, and finally get M or L from Φ/I. This eliminates the need to deal directly with the singularity in the above integral.

Generalization to Volume Currents

We may generalize the above to volume currents by using the usual relation between the vector potential and the volume current density (assuming appropriate boundary conditions):

$$\vec{A}(\vec{r}) = \frac{1}{4\,\pi}\int_{\mathcal{V}} d\tau'\,\frac{\vec{J}(\vec{r}\,')}{|\vec{r}-\vec{r}\,'|} \tag{7.69}$$

$$\Longrightarrow \qquad M_{ij} = \frac{\mu_o}{4\,\pi}\frac{1}{\mathrm{I}_i\,\mathrm{I}_j}\int_{\mathcal{V}_i} d\tau \int_{\mathcal{V}_j} d\tau'\,\frac{\vec{J}(\vec{r})\cdot\vec{J}(\vec{r}\,')}{|\vec{r}-\vec{r}\,'|} \tag{7.70}$$

$$L = \frac{\mu_o}{4\,\pi}\frac{1}{\mathrm{I}^2}\int_{\mathcal{V}} d\tau \int_{\mathcal{V}} d\tau'\,\frac{\vec{J}(\vec{r})\cdot\vec{J}(\vec{r}\,')}{|\vec{r}-\vec{r}\,'|} \tag{7.71}$$

where $\mathcal{V}_i$ is the volume of the ith inductor. We notice that the currents do not drop out as cleanly, but, assuming linear behavior of the current flow (the current does not flow differently as the overall magnitude of the current is changed), we expect $\vec{J} \propto \mathrm{I}$ and indeed, once the functional dependence of the current density on position has been established, the inductances are purely geometrical quantities as for the case of line currents.

This is not a rigorous proof. We will return to this when we discuss magnetic energy.

Example 7.4: Self and Mutual Inductances of Solenoids

Let's first calculate the self-inductance of a solenoid of radius a. Recall that the field of a solenoid is only nonzero inside it and has value

$$\vec{B} = \mu_o \, n \, \mathrm{I} \, \widehat{z} \tag{7.72}$$

where n is the number of turns per unit length, the solenoid axis is along $\widehat{z}$, and the current flows along the $\widehat{\phi}$ direction. The magnetic flux threading the solenoid and the self-inductance are therefore

$$\Phi = n \, \ell \, \pi \, a^2 \, B = \mu_o \, n^2 \, \ell \, \pi \, a^2 \, \mathrm{I} \qquad \Longrightarrow \qquad L = \mu_o \, n^2 \, \ell \, \pi \, a^2 \tag{7.73}$$

If we have two interpenetrating solenoids with turn densities n_1 and n_2, radii $a_1 < a_2$, and lengths $\ell_1 < \ell_2$, then the flux into solenoid 1 of the field from solenoid 2 and the mutual inductance are

$$\Phi_{12} = n_1 \, \ell_1 \, \pi \, a_1^2 \, B_2 = \mu_o \, n_1 \, n_2 \, \ell_1 \, \pi \, a_1^2 \, \mathrm{I}_2 \qquad \Longrightarrow \qquad M = \mu_o \, n_1 \, n_2 \, \ell_1 \, \pi \, a_1^2 \tag{7.74}$$

It is interesting and useful to note that we may also calculate Φ_{21} using M given the symmetry of M. This is very convenient, as calculating the contribution to Φ_{21} from the portion of solenoid 1's field past its ends would be nontrivial. It is also interesting to see that the mutual inductance is not manifestly symmetric under index exchange $1 \leftrightarrow 2$. This reflects the asymmetry of the setup between solenoids 1 and 2.

Lecture 22:

Electrodynamics III:

Magnetic Energy of Currents and Fields
Magnetic Energy in the Presence of Magnetizable Materials
Magnetic Forces

Date Revised: 2024/04/10 17:00
Revised lecture break
Date Given: 2024/04/09

Magnetic Energy and Forces

Magnetic Energy in Terms of Currents

Let's consider the work that has be done to drive current against the emf in an inductive object (*e.g.*, a simple loop or a solenoid). The emf is sometimes called the "back emf" because it is the line integral of a force that tries to drive a current that is intended to counter the changing field due to the current one is varying and so the current one is varying must be driven against the emf.

That is, when a varying current is driven through an inductive object, it has to be driven against a force per unit charge whose line integral (note that we did not say potential!) along the current's path is $\mathcal{E}$. (The force is due to an induced electric field for this case of a stationary loop that is experiencing a $d\Phi/dt$ due to its own current varying.) The force that must be exerted, and the work that must be done, is above and beyond the force needed to overcome the inertia of the charge carriers (*i.e.*, the Newton's Law force $F = m\,a$).

The rate at which this work is being done is given by the same expression we derived before for the work done by the pulling force in the case that the field was fixed but the loop was moving: it is the work done per unit charge by the battery to push the current against the back emf, $-\mathcal{E}$, times the charge flowing past a given point per unit time, I:

$$\frac{dW}{dt} = Power = -\mathrm{I}\,\mathcal{E} = L\,\mathrm{I}\,\frac{d\mathrm{I}}{dt} \tag{7.75}$$

We can integrate this over time to get the total work done and the *magnetic energy* created:

$$\boxed{W = \frac{1}{2} L \, \mathrm{I}^2} \tag{7.76}$$

It is natural to ask why, when we considered the situation with the loop being pulled through a magnetic field, we did not worry about this magnetic energy: we said that the work done by the pulling force was completely dissipated in the Joule heating of the resistor. Or, put another way, why did we not need to include a resistor in the calculation here? When we include the resistor, some of the work done by the pulling force *as the loop was accelerated from rest to* $\vec{v}$ goes into this magnetic energy, the $L\,\mathrm{I}^2/2$ energy. Once at fixed velocity, however, the current and thus this energy stay constant. The pulling force continues to do work, however. Since we specifically made the steady-state assumption — that the loop had been moving at fixed $\vec{v}$ for all time and would stay moving for all time — this transient process of creating $L\,\mathrm{I}^2/2$ was not relevant, and our conservation of energy argument was valid; we just neglected noting the path that the energy took through the magnetic energy in the steady state. Now, without the resistor, we are focused entirely on the transient portion of the process, hence the importance of the magnetic energy.

Let's generalize the above result. Consider a system of N inductive elements with inductance matrix M_{ij} ($M_{ii} \equiv L_i$, $M_{ij} = M_{ji}$). We turn on the currents in the order $i = 1, 2, \cdots, N$. We first have to maintain the current I_i against the emf on inductor i felt due to its changing current. Once it has reached its final value, we also have to maintain it as the currents in the inductors $j > i$ are increased from 0 to their final values (note: $j > i$, not $j < i$ as we had in electrostatics). The required power is:

$$\frac{dW}{dt} = \sum_{i=1}^{N} \frac{dW_i}{dt} = \sum_{i=1}^{N} (-\mathrm{I}_i \, \mathcal{E}_i) = \sum_{i=1}^{N} \left[\mathrm{I}_i M_{ii} \frac{d\mathrm{I}_i}{dt} + \mathrm{I}_i \sum_{j>i}^{N} M_{ij} \, \frac{d\mathrm{I}_j}{dt} \right]$$

$$\implies \quad W = \sum_{i=1}^{N} \left[\frac{1}{2} \, M_{ii} \, \mathrm{I}_i^2 + \mathrm{I}_i \sum_{j>i}^{N} M_{ij} \, \mathrm{I}_j \right] = \frac{1}{2} \sum_{i,j=1}^{N} M_{ij} \mathrm{I}_i \mathrm{I}_j \tag{7.77}$$

where we have symmetrized the sum over j by including a factor of $1/2$, and then we combined the cross-terms with the self-terms. If we rewrite all our relations using matrix notation, with $\underline{\mathrm{I}}$ being a column vector of currents, $\underline{\Phi}$ being a column vector of fluxes, and $\underline{\underline{M}}$ being the matrix of mutual inductances, we have

$$\boxed{\underline{\Phi} = \underline{\underline{M}} \, \underline{\mathrm{I}} \qquad W = \frac{1}{2} \, \underline{\Phi}^T \underline{\mathrm{I}} = \frac{1}{2} \, \underline{\mathrm{I}}^T \underline{\underline{M}} \, \underline{\mathrm{I}}} \tag{7.78}$$

Note: we could have calculated the above somewhat differently, considering the work done to maintain the loops $j < i$ at their final current values, plus the work done in loop i itself, while loop i is begin ramped to its final value. The result would be the same.

Magnetic Energy in Terms of Magnetic Field

Let's manipulate our circuit equations above to try to get the energy in terms of the magnetic field. First, we can rewrite the circuit expressions using the vector potential:

$$L\,\mathrm{I} = \Phi = \int_{\mathcal{S}(\mathcal{C})} da\,\hat{n}\cdot\vec{B} = \oint_{\mathcal{C}} d\vec{\ell}\cdot\vec{A} \tag{7.79}$$

$$\implies \quad W = \frac{1}{2}\,L\,\mathrm{I}^2 = \frac{\mathrm{I}}{2}\oint_{\mathcal{C}} d\vec{\ell}\cdot\vec{A} = \frac{1}{2}\oint_{\mathcal{C}} d\ell\,\vec{\mathrm{I}}\cdot\vec{A} \tag{7.80}$$

We can obviously generalize this for volume currents to

$$\boxed{W = \frac{1}{2}\int_{\mathcal{V}} d\tau\,\vec{J}\cdot\vec{A}} \tag{7.81}$$

Aside: The above equation now justifies Equation 7.71: if one uses Equation 5.56 to write $\vec{A}$ in terms of $\vec{J}$ and then calculates $L = 2\,W/I^2$, one recovers Equation 7.71. By considering two separate volume current distributions, one can recover Equation 7.70 also.

We can use Ampere's Law to obtain

$$W = \frac{1}{2\,\mu_o} \int_{\mathcal{V}} d\tau\, \vec{A} \cdot \left(\vec{\nabla} \times \vec{B} \right) \tag{7.82}$$

We use the product rule for the divergence of a cross-product, $\vec{\nabla} \cdot (\vec{a} \times \vec{b}) = \vec{b} \cdot (\vec{\nabla} \times \vec{a}) - \vec{a} \cdot (\vec{\nabla} \times \vec{b})$ to rewrite this as

$$W = \frac{1}{2\,\mu_o} \int_{\mathcal{V}} d\tau \left[\vec{B} \cdot \left(\vec{\nabla} \times \vec{A} \right) - \vec{\nabla} \cdot \left(\vec{A} \times \vec{B} \right) \right] \tag{7.83}$$

$$= \frac{1}{2\,\mu_o} \int_{\mathcal{V}} d\tau\, |\vec{B}|^2 - \frac{1}{2\,\mu_o} \oint_{\mathcal{S}(\mathcal{V})} da\, \hat{n} \cdot \left(\vec{A} \times \vec{B} \right) \tag{7.84}$$

Now, the original volume integral was over only the region containing the current, but the volume integral could be extended to a larger region since there would be no additional contribution. So we do the usual thing and expand the volume to include all of space and take the bounding surface to infinity.

We assume that $\vec{A} \times \vec{B}$ falls off more quickly than $1/r^2$ (true for finite current distributions) so that the surface term goes to zero, or that the particulars of the configuration ensure the integral vanishes even if the current distribution is not finite and we expect a finite energy. Therefore,

$$\boxed{W = \frac{1}{2\,\mu_o} \int d\tau\, |\vec{B}|^2} \tag{7.85}$$

Thus we see that the magnetic energy is just given by the integral of the square of the field. We now see that the magnetic energy created as the currents are ramped from zero to their final values is *stored in the field*. (We specifically avoided saying earlier where it was stored!)

It is interesting to think about how it is possible to store energy in a magnetic field given that the field can do no work. One has to think of this as the work done to drive against the induced electric field as the field was increased from zero to its final value. As usual, this work is supplied by a battery, not the magnetic field.

On the point about the surface term in the case of configurations for which the fields do not fall of at infinity:

- For an infinite solenoid, the surface term only includes the endcaps of the solenoid, since $\vec{B}$ vanishes outside the solenoid. The contributions of the two endcaps vanish because $\vec{A} \times \vec{B}$ points along $\widehat{s}$ in cylindrical coordinates, but the endcap's normal is along $\widehat{z}$. While the surface term vanishes, the energy is still infinite because the volume integral is over an infinite volume with a constant energy density.
- For an infinite wire, even when calculated per unit length, all the terms are logarithmically infinite. This is because the current and the fields do not die off quickly enough at infinity. The calculation fails even if one does the calculation using $\vec{J} \cdot \vec{A}$, and even for a finite diameter wire (if one accounts for the fact that $\vec{J}$ becomes a surface current in the perfect conductor case).

Example 7.5: Magnetic Energy in a Solenoid

A solenoid of radius a with n turns per unit length and current I has a field $B = \mu_o\, n\, \mathrm{I}$. Therefore, the magnetic energy in such a solenoid of length ℓ is

$$W = \pi\, a^2\, \ell\, \frac{1}{2\,\mu_o}\, B^2 = \frac{1}{2}\, \mu_o\, n^2 \mathrm{I}^2 \pi\, a^2\, \ell \tag{7.86}$$

Note that we can extract from this the self-inductance using $W = L\,\mathrm{I}^2/2$, yielding $L = \mu_o\, n^2\, \pi\, a^2\, \ell$ as we obtained by calculating the flux. To put some numbers on this, the LHC CMS experiment (`http://home.web.cern.ch/about/experiments/cms`) has a solenoid with a field of 4 T with radius $a = 3$ m and length 13 m. The stored energy is therefore about 2.5 gigajoules, an enormous number.

Example 7.6: Magnetic Energy in a Coaxial Cable

This is Griffiths Example 7.13. For a coaxial cable of length ℓ with inner and outer conductor radii a and b, the energy and resulting self-inductance are

$$W = \frac{\mu_o}{4\,\pi}\, \mathrm{I}^2\, \ell\, \ln\frac{b}{a} \qquad\qquad L = \frac{\mu_o}{2\,\pi}\, \ell\, \ln\frac{b}{a} \tag{7.87}$$

An Alternate Logical Path

We followed Griffiths in first developing concepts for inductance and energy using current loops and then generalizing both to volume current distributions in a fairly obvious way. That said, there is a more rigorous way to do all of this by first considering the work done in maintaining current densities in the presence of the electric field generated by changes in those current densities. This then leads to the idea of $\vec{J} \cdot \vec{A}$ being the energy density in the magnetic field. Then one can define inductances by writing the field energy in terms of the total currents that normalize the current distributions. It can then be shown that these inductances relate currents to fluxes and thus rates of change of currents to emfs. This alternative logical path is followed in Jackson §5.16–5.17.

Magnetic Energy of an Assembly of Free Currents in the Presence of Magnetizable Materials

This analysis is done similarly to the electrostatic case and follows Jackson §5.16. Recall that we discussed the distinction between the total energy needed to assemble the final configuration, including the construction of the bound dipoles, and the energy needed to bring the free charges in assuming the bound dipoles already exist and neglecting the potential energy of creating them. In this case, we assume the bound magnetic dipoles are created and maintained by someone else — someone else has built them and raised their currents to their full values for us and also maintains those currents in the presence of back emf generated when the free currents change — and we need only consider the work that has to be done to turn on some free currents in the presence of these bound magnetic dipoles.

This separation is not academic: all naturally occurring magnetic materials rely on the magnetic dipole moments of fundamental particles. Those magnetic dipoles are unchangeable, and thus the energy stored in them is effectively a constant offset that we have no experimental access to. It therefore makes sense to ignore it in calculations of magnetic energy.

Consider the differential of work the battery must do to maintain the *free currents* $\vec{J}_f$ during a change in the magnetic field $\delta\vec{B}$. (We need not specify whether this change is due to a change in $\vec{J}_f$ and/or the location of the magnetizable materials — all that matters is $\delta\vec{B}$.) Equation 7.80 implies that the change in energy would be (holding the geometry fixed)

$$\delta W = -\mathrm{I}\,\mathcal{E}\,\delta t = \mathrm{I}\,\frac{d\Phi}{dt}\,\delta t = \mathrm{I}\,\delta\Phi = \mathrm{I}\,\delta\left[\int_{\mathcal{S}(\mathcal{C})} da\,\widehat{n}\cdot\vec{B}\right] = \mathrm{I}\,\delta\left[\oint_{\mathcal{C}} d\vec{\ell}\cdot\vec{A}\right] = \mathrm{I}\oint_{\mathcal{C}} d\vec{\ell}\cdot\delta\vec{A} \tag{7.88}$$

for which the volume generalization would be

$$\delta W = \int_{\mathcal{V}} d\tau\,\vec{J}_f\cdot\delta\vec{A} = \int_{\mathcal{V}} d\tau\,\left(\vec{\nabla}\times\vec{H}\right)\cdot\delta\vec{A} \tag{7.89}$$

Apply the same algebra and the same discarding of the surface term as in free space:

$$\delta W = \int_{\mathcal{V}} d\tau\,\vec{H}\cdot\left(\vec{\nabla}\times\delta\vec{A}\right) = \int_{\mathcal{V}} d\tau\,\vec{H}\cdot\delta\vec{B} \tag{7.90}$$

For nonlinear materials, we would need to apply the specific $\vec{B}(\vec{H})$ function go further. For linear materials, we use $\delta\vec{B} = \mu\delta\vec{H}$ to do the integral and obtain the expected analogue to the free-space result:

$$\boxed{W = \frac{1}{2\,\mu}\int_{\mathcal{V}} d\tau\,|\vec{B}|^2 = \frac{\mu}{2}\int_{\mathcal{V}} d\tau\,|\vec{H}|^2 = \frac{1}{2}\int_{\mathcal{V}} d\tau\,\vec{H}\cdot\vec{B}} \tag{7.91}$$

Magnetic Energy of a Magnetizable Material in an External Field

We can do this derivation along the lines of what we did for polarizable materials, following Jackson §5.16. Let's assume that we have a configuration of currents that generates fields $\vec{B}_1$ and $\vec{H}_1$ in a volume containing a permeable material μ_1. Now, bring in a material of permeability μ_2 such that it occupies a volume $\mathcal{V}_2$ contained in $\mathcal{V}$ while holding the free source currents fixed. The fields (everywhere) change to $\vec{B}_2$ and $\vec{H}_2$.

The energy difference we want to calculate is

$$U_2 - U_1 = \frac{1}{2}\int d\tau \left[\vec{B}_2 \cdot \vec{H}_2 - \vec{B}_1 \cdot \vec{H}_1\right] \tag{7.92}$$

We can apply similar manipulations as we did for the electrostatic case. First, we rewrite the above as

$$U_2 - U_1 = \frac{1}{2}\int d\tau \left[\vec{B}_2 \cdot \vec{H}_1 - \vec{B}_1 \cdot \vec{H}_2\right] + \frac{1}{2}\int d\tau \left[\vec{B}_1 + \vec{B}_2\right] \cdot \left[\vec{H}_2 - \vec{H}_1\right] \tag{7.93}$$

Since $\vec{\nabla} \cdot \left[\vec{B}_1 + \vec{B}_2\right] = 0$, it can be derived from a vector potential $\vec{A}$, allowing us to rewrite the second term as

$$\frac{1}{2}\int d\tau \left[\vec{H}_2 - \vec{H}_1\right] \cdot \left(\vec{\nabla} \times \vec{A}\right) \tag{7.94}$$

We use again the vector identity $\vec{\nabla} \cdot (\vec{a} \times \vec{b}) = \vec{b} \cdot (\vec{\nabla} \times \vec{a}) - \vec{a} \cdot (\vec{\nabla} \times \vec{b})$ to integrate by parts, and we turn the divergence into a surface term that we can discard because $\vec{H}_2 - \vec{H}_1$ should vanish as we go far from the permeable material, yielding for the second term

$$\frac{1}{2}\int d\tau \, \vec{A} \cdot \vec{\nabla} \times \left(\vec{H}_2 - \vec{H}_1\right) \tag{7.95}$$

The curl in the integrand vanishes because $\vec{H}_2$ and $\vec{H}_1$ are sourced by the same free currents.

We are thus left with the first term from the equation we started with

$$U_2 - U_1 = \frac{1}{2}\int d\tau \left[\vec{B}_2 \cdot \vec{H}_1 - \vec{B}_1 \cdot \vec{H}_2\right] \tag{7.96}$$

Applying linearity, $\vec{B} = \mu\,\vec{H}$, we then obtain

$$U_2 - U_1 = \frac{1}{2}\int d\tau\,(\mu_2 - \mu_1)\,\vec{H}_2 \cdot \vec{H}_1 \tag{7.97}$$

Finally, we recognize $\mu_2 - \mu_1 = 0$ except in $\mathcal{V}_2$, so

$$U_2 - U_1 = \frac{1}{2}\int_{\mathcal{V}_2} d\tau\,(\mu_2 - \mu_1)\,\vec{H}_2 \cdot \vec{H}_1 = \frac{1}{2}\int_{\mathcal{V}_2} d\tau \left(\frac{1}{\mu_1} - \frac{1}{\mu_2}\right)\vec{B}_2 \cdot \vec{B}_1 \tag{7.98}$$

This is the analogue of Equation 4.86 aside from a sign flip, which mechanically is due to the fact that $\vec{B} = \mu\,\vec{H}$ (rather than $\vec{H} = \mu\,\vec{B}$). If we take $\mu_1 = \mu_0$ and $\mu_2 = \mu$, we can use $\vec{M}_2 = (\mu_2/\mu_0 - 1)\vec{H}_2 = (\mu/\mu_0 - 1)\vec{H}_2$ to rewrite this as

$$W = U_2 - U_1 = \frac{1}{2}\int_{\mathcal{V}_2} d\tau\,\vec{M} \cdot \vec{B}_1 \qquad \Longleftrightarrow \qquad w = \frac{1}{2}\,\vec{M} \cdot \vec{B}_1 \tag{7.99}$$

where now we replace $\vec{M}_2$ by $\vec{M}$ since $\vec{M}_1 = \vec{0}$ if $\mu_1 = \mu_o$. So $\vec{M}$ is the magnetization density of the volume occupied by μ and $\vec{B}_1$ is the *magnetic field in the absence of the permeable material.* There is a sign flip relative to the electrostatic case (Equation 4.87) that, mechanically, came from the sign flip in Equation 7.98.

How do we understand this sign flip conceptually? Trying to track the sign through the derivation is not illuminating. But it can be understood by comparing to our calculation of the energy of magnetic dipole in an external field, Equation 5.153, where we assumed that the magnetic dipole moment and field were given and held fixed without our having to account for how this was done. In that case, the potential energy of the configuration was $U = -\vec{m} \cdot \vec{B}$. (The factor of $1/2$ here comes from the linear relationship between $\vec{m}$ and $\vec{B}$ and the integration from zero field to $\vec{B}$, which is not important for this discussion). We see that we have a sign flip relative to that situation. It is sensible, then, to attribute the sign flip to the fact that, in deriving the expression $w = |\vec{B}|^2/2\,\mu$ that was the starting point for this derivation, we accounted for the work done by the batteries to maintain the free currents as the permeable material was brought in. No such work was required in the previously considered case of a fixed dipole moment $\vec{m}$ and fixed field $\vec{B}$.

Note that, importantly, we do not account for how the magnetization density $\vec{M}$ is maintained. This is to be distinguished from not considering how $\vec{M}$ is created, which we argued was just an unchangeable offset. We may ignore this additional consideration here because, again, the magnetization density is, in naturally occurring systems, due to fundamental magnetic dipoles that require no batteries to maintain their magnetic moments.

When we compare to the electrostatic analogy, Equation 4.87, we recognize a sign flip, too. The rationale is the same: in the electrostatic case, we do not have to do any work to maintain the free charges sourcing the applied field $\vec{E}$ at their nominal positions, while here we do have to do work with a battery to maintain currents at the nominal values and positions due to the back emf from the changing $\vec{M}$.

Magnetic Forces from Magnetic Energy with Fluxes Fixed

To evaluate magnetic forces, we need to consider what happens if we have an infinitesimal generalized displacement of one of our inductors. Because it is more straightforward, let's first consider the fixed fluxes case, which is analogous to holding charges fixed in electrostatics. If $d\Phi/dt = 0$, then there are no emfs and there is no need for a battery to do work to drive currents against those emfs. So we only need to consider $dW_{field}|_{\Phi}$. We can directly calculate the generalized force from the energy holding the fluxes fixed:

$$F_\xi\Big|_\Phi = -\left(\frac{\partial W_{field}}{\partial \xi}\right)_\Phi = -\frac{1}{2}\sum_{i,j=1}^{N} \Phi_i\,\Phi_j\,\frac{\partial M_{ij}^{-1}}{\partial \xi}\Bigg|_\Phi = -\frac{1}{2}\,\underline{\Phi}^T\left[\frac{\partial}{\partial \xi}\underline{\underline{M}}^{-1}\right]\underline{\Phi}\Bigg|_\Phi \quad (7.100)$$

which is the analogue of Equation 4.90.

It's not clear at a microscopic level (*i.e.*, what has to happen to the currents) how one maintains fixed fluxes as inductors are moved around. But, certainly, one is assured that, if one sets up a system of inductors with currents and then disconnects them from their batteries, any movement of the loops must keep the fluxes fixed and change the currents accordingly since there are no batteries to work against the emfs and maintain the currents. This issue will be revisited in homework and is discussed in Griffiths Section 8.3 (4th edition).

Magnetic Forces from Magnetic Energy with Currents Fixed

We follow Jackson §5.16. Let's approach this case like we did in electrostatics, first fixing the fluxes (charges), allowing the currents (voltages) to change, and then returning the currents (voltages) to their original values. The contribution to the change in energy from the fluxes-fixed portion is

$$dW_{field}\Big|_{\Phi} = \frac{1}{2}\sum_{i,j=1}^{N} \Phi_i\,\Phi_j\,d\left[\underline{\underline{M}}^{-1}\right]_{ij} \tag{7.101}$$

This causes changes in the currents (at fixed flux)

$$d\mathrm{I}_i\Big|_{\Phi} = \sum_{j=1}^{N} d\left[\underline{\underline{M}}^{-1}\right]_{ij}\Phi_j \tag{7.102}$$

If we add back current to return to a fixed-current situation, then changes in fluxes result:

$$d\Phi_k\Big|_{\mathrm{I}} = \sum_{i=1}^{N} M_{ki}\,(-d\mathrm{I}_i)_{\Phi} = -\sum_{i,j=1}^{N} M_{ki}\,d\left[\underline{\underline{M}}^{-1}\right]_{ij}\Phi_j \tag{7.103}$$

The above infinitesimal flux changes cause emfs. The currents I_i must be maintained by batteries in the presence of these emfs. The work done by the batteries over the infinitesimal time dt needed to make the flux changes is

$$dW^{bat}_{field}\Big|_{\mathrm{I}} = dt\,\frac{dW^{bat}_{field}}{dt}\bigg|_{\mathrm{I}} = dt\,\sum_{k=1}^{n}(-\mathrm{I}_k\,\mathcal{E}_k) = dt\,\sum_{k=1}^{n}\mathrm{I}_k\,\frac{d\Phi_k}{dt}\bigg|_{\mathrm{I}} \tag{7.104}$$

$$= \sum_{k=1}^{n}\mathrm{I}_k\,d\Phi_k\Big|_{\mathrm{I}} = -\sum_{i,j,k=1}^{n}\mathrm{I}_k\,M_{ki}\,d\left[\underline{\underline{M}}^{-1}\right]_{ij}\Phi_j \tag{7.105}$$

$$= -\sum_{i,j=1}^{n}\Phi_i\,\Phi_j\,d\left[\underline{\underline{M}}^{-1}\right]_{ij} = -2\,dW_{field}\Big|_{\Phi} \tag{7.106}$$

Note that we did not need to worry about the work done by the battery to make the canceling change in current $d\mathrm{I}_k$ because this current change would be multiplied against $dt\,\mathcal{E}_k = -d\Phi_k$, which is already infinitesimal. We need only consider the above term consisting of the nominal currents I_k multiplied against $dt\,\mathcal{E}_k$. We had the same situation in electrostatics, where we did not consider the $dV_k\,dQ_k$ terms, only the $V_k\,dQ_k$ terms.

The total change in the field energy is then obtained by adding the two contributions to the field energy: the field energy change at fixed flux followed by the energy added to the field by the batteries as they return the currents to their initial values:

$$dW_{field}\Big|_{\mathrm{I}} = dW_{field}\Big|_{\Phi} + dW^{bat}_{field}\Big|_{\mathrm{I}} = dW_{field}\Big|_{\Phi} - 2\,dW_{field}\Big|_{\Phi} = -dW_{field}\Big|_{\Phi} \qquad (7.107)$$

We thus find a perfect analogy to the electrostatic case, where we found $dW_{field}|_V = -dW_{field}|_Q$. We may thus use the same guidance: the force cannot depend on whether the situation used is fixed flux or fixed current, and so the forces calculated at fixed flux and fixed current must be the same. Thus, we must conclude

$$F_\xi\big|_{\mathrm{I}} = \left(\frac{\partial W_{field}}{\partial \xi}\right)_{\mathrm{I}} = -\left(\frac{\partial W_{field}}{\partial \xi}\right)_{\Phi} = F_\xi\big|_{\Phi} \qquad (7.108)$$

That is, just like in the electrostatic case, when the battery is involved and we consider the energy of the entire system, we see we must take the positive gradient of the field energy at fixed current, rather than considering only the energy of the field and taking the negative gradient of the field energy at fixed current. The reason these two gradients are different, with a sign between them, is because the derivative is calculationally different depending on whether I or Φ is held fixed.

We can see this works mathematically by trying it:

$$\left(\frac{\partial W_{field}}{\partial \xi}\right)_{\mathrm{I}} = \frac{\partial}{\partial \xi}\left[\frac{1}{2}\sum_{i,j=1}^{N} \mathrm{I}_i\,\mathrm{I}_j\,M_{ij}\right]_{\mathrm{I}} = \frac{1}{2}\sum_{i,j=1}^{N} \mathrm{I}_i\,\mathrm{I}_j\,\frac{\partial M_{ij}}{\partial \xi}$$
$$= \frac{1}{2}\,\underline{\mathrm{I}}^{T}\left[\frac{\partial}{\partial \xi}\underline{\underline{M}}\right]\underline{\mathrm{I}} \tag{7.109}$$

Since $\partial \underline{\underline{M}}^{-1}/\partial\xi = -\underline{\underline{M}}^{-1}[\partial \underline{\underline{M}}/\partial\xi]\underline{\underline{M}}^{-1}$ (one can see this in the same way we proved the analogous relationship for $\underline{\underline{C}}$), this form yields Equation 7.100 for $F_\xi\big|_\Phi$. Thus,

$$F_\xi\big|_{\mathrm{I}} = \left(\frac{\partial W_{field}}{\partial \xi}\right)_{\mathrm{I}} = -\left(\frac{\partial W_{field}}{\partial \xi}\right)_{\Phi} = F_\xi\big|_\Phi \tag{7.110}$$

Lecture 23:

Electrodynamics IV:
Displacement Current
Maxwell's Equations

Conservation Laws:
Conservation of Charge and Energy

Electromagnetic Waves I:
Electromagnetic Waves in Vacuum:

Date Revised: 2024/04/10 17:00
Date Given: 2024/04/11

Maxwell's Equations

The Inconsistency in our Equations

Let's write the full set of equations we have come to:

$$\vec{\nabla} \cdot \vec{E} = \frac{\rho}{\epsilon_o} \qquad \vec{\nabla} \cdot \vec{B} = 0 \tag{7.111}$$

$$\vec{\nabla} \times \vec{E} = -\frac{\partial \vec{B}}{\partial t} \qquad \vec{\nabla} \times \vec{B} = \mu_o \vec{J} \tag{7.112}$$

Now, we know that the divergence of a curl vanishes: it's a vector identity. We should check that it holds! For the electric field, we have

$$\vec{\nabla} \cdot \vec{\nabla} \times \vec{E} = \vec{\nabla} \cdot -\frac{\partial \vec{B}}{\partial t} = -\frac{\partial}{\partial t} \vec{\nabla} \cdot \vec{B} = 0 \tag{7.113}$$

If we repeat with $\vec{B}$, we obtain

$$\vec{\nabla} \cdot \vec{\nabla} \times \vec{B} = \mu_o \vec{\nabla} \cdot \vec{J} = -\mu_o \frac{\partial \rho}{\partial t} \neq 0 \quad \text{in general} \tag{7.114}$$

There is a more physical way to see this by applying Ampere's Law to a circuit containing a parallel-plate capacitor. Construct an Ampere's Law loop around the wire carrying the current. Ampere's Law is satisfied because there is magnetic field in an azimuthal direction around the wire (giving a nonzero line integral of $\vec{B}$) and there is current passing through the disk-like surface whose boundary is the contour.

Now pick another surface that passes between the capacitor plates. This is an equally valid surface; nothing about our proof of Ampere's Law from the Biot-Savart Law assumed a particular choice of surface for the Ampere's Law surface integral. But this surface has no current intersecting it because it passes through the capacitor!

The reason this problem happens and we never noticed it before is because we have a non-steady-state case here: charge piles up on the capacitor plates giving $\partial\rho/\partial t \neq 0$; we had assumed all along during magnetostatics and during our discussion of induction that $\partial\rho/\partial t = 0$.

While charge cannot jump across the capacitor plates so that there can be a current to keep Ampere's Law satisfied, we do recognize that, as charge enters one plate of the capacitor, an equal amount of charge leaves the other plate, ensuring that $dQ/dt = 0$ for the capacitor as a whole. This is suggestive of the idea that perhaps there is some sort of current flowing across the capacitor gap, just not the physical movement of charges we are used to. This new current will be called the *displacement current*.

The Displacement Current

In order to solve the above problem, we need something that will cancel

$$\mu_o \vec{\nabla} \cdot \vec{J} = -\mu_o \frac{\partial \rho}{\partial t} = -\mu_o \frac{\partial}{\partial t}\left(\epsilon_o \vec{\nabla} \cdot \vec{E}\right) = -\mu_o \vec{\nabla} \cdot \left(\epsilon_o \frac{\partial \vec{E}}{\partial t}\right) \tag{7.115}$$

Let's just add the necessary term to Ampere's Law:

$$\vec{\nabla} \times \vec{B} = \mu_o \vec{J} + \mu_o \epsilon_o \frac{\partial \vec{E}}{\partial t} \tag{7.116}$$

Physically, what we have done is defined a second current density so that the divergence of the *total* current density $\vec{J} + \epsilon_o \partial \vec{E}/\partial t$ vanishes. This vanishing is equivalent to the vanishing of the flux of the total current through any surface, which is what is needed to solve the problem we pointed out: now the surface integral of the enclosed current does not depend on the surface chosen.

Was it ok to do this? Does it violate any of our previous conclusions? The only equation we have modified is the $\vec{\nabla} \times \vec{B}$ equation, so we only need to consider our study of magnetostatics, where we applied this equation. The addition preserves the usual behavior of $\vec{\nabla} \times \vec{B}$ for magnetostatics because $\partial \vec{E}/\partial t = 0$ in magnetostatics. Why? Two things can result in time dependence of $\vec{E}$. The first is time dependence in ρ. But in magnetostatics, we make the steady-state assumption, explicitly requiring no buildup of charge and hence $\partial \rho/\partial t = 0$. The second is time dependence of $\vec{B}$, which can yield time dependence of $\vec{E}$ via Faraday's Law. But magnetostatics assumes $\vec{B}$ is constant in time, so there is no worry there.

The added term is called the *displacement current density*,

$$\boxed{\vec{J}_d \equiv \epsilon_o \frac{\partial \vec{E}}{\partial t}} \tag{7.117}$$

This is the "current" we foresaw we needed. It is not a physical current carried by charges, but it represents the fact that, when charge builds up in some point in a circuit because of a gap that prevents physical current from flowing, it causes a changing electric field that then pushes charge away from that point, causing current to flow. One needs a *changing* electric field because otherwise one would quickly reach a steady state in which no new charge would move and thus there would be no source for current. (Maintaining fixed charge on a capacitor does not require current to flow.) Effectively, the displacement current carries the current across physical gaps in the circuit. It is therefore justified, both on the basis of units and on physical intuition, to call it a current. One could even argue that the name is suitable: the "displacement" current causes the displacement of charges on the two sides of a gap across which true current cannot flow. (This argument is in disagreement with Griffiths' statement that the displacement current has nothing to do with current.)

More importantly, we also now see for the first time that a changing electric field sources a magnetic field. Unlike with Faraday's Law, however, there is no negative sign and the induced magnetic field does not act in such a way as to try to cancel the changing electric field.

By construction, $\vec{J}_d$ solves the problem with $\vec{\nabla} \cdot \vec{\nabla} \times \vec{B}$, and we already intuitively expect it will sove the problem with the integral version of Ampere's Law, but let's see that explicitly. The electric field in the capacitor is

$$\vec{E} = \frac{\sigma}{\epsilon_o}\,\widehat{n} = \frac{1}{\epsilon_o}\,\frac{Q}{A}\,\widehat{n} \tag{7.118}$$

where $\widehat{n}$ is the normal from the positive plate to the negative plate. Therefore, the displacement current is

$$\vec{J}_d = \epsilon_o\,\frac{\partial \vec{E}}{\partial t} = \frac{1}{A}\,\frac{dQ}{dt}\,\widehat{n} = \frac{\mathrm{I}}{A}\,\widehat{n} \tag{7.119}$$

The integral form of Ampere's Law with the displacement current is therefore

$$\oint_{\mathcal{C}} d\vec{\ell} \cdot \vec{B} = \mu_o\,\mathrm{I}_{encl} + \mu_o \int_{\mathcal{S}(\mathcal{C})} da\,\widehat{n} \cdot \vec{J}_d \tag{7.120}$$

If we choose the first surface we discussed earlier, the flat surface in the plane of the contour $\mathcal{C}$, we get the first term but the second term vanishes, yielding $\mu_o\,\mathrm{I}$. If we choose the second surface, the one between the capacitor plates, the first term vanishes but the second term gives $\mu_o\,\mathrm{I}$. Thus, the inconsistency seen earlier has been eliminated.

Example 7.7: Displacement Current for Coaxial Conductors (Griffiths 7.36)

This is a continuation of the example from earlier. We want to calculate the displacement current density and the total displacement current and to compare quantitatively I and I_d.

The displacement current density is $\vec{J}_d = \epsilon_o \, \partial \vec{E}/\partial t$:

$$\vec{J}_d(s < a) = \epsilon_o \, \frac{\partial}{\partial t} \, \widehat{z} \, \frac{\mu_o}{2\,\pi} \, \omega \, \mathrm{I}_0 \, \sin\omega t \, \ln\frac{a}{s} = \widehat{z} \, \mu_o \, \epsilon_o \, \frac{\omega^2 \, \mathrm{I}_0}{2\,\pi} \, \cos\omega t \, \ln\frac{a}{s} \tag{7.121}$$

Let's integrate over the (s, ϕ) plane to get the total displacement current:

$$\begin{aligned} \mathrm{I}_d &= \int_{\mathcal{S}(\mathcal{C})} \widehat{n} \cdot \vec{J}_D = \int_0^{2\,\pi} d\phi \, \mu_o \, \epsilon_o \, \frac{\omega^2 \, \mathrm{I}_0}{2\,\pi} \, \cos\omega t \int_0^a s \, ds \, \ln\frac{a}{s} \\ &= \mu_o \, \epsilon_o \, \omega^2 \, \mathrm{I}_0 \cos\omega t \left(a^2 \left[\frac{x^2}{2} \left(\ln x - \frac{1}{2} \right) \right] \Bigg|_{x=1}^{0} \right) \\ &= \mu_o \, \epsilon_o \, \omega^2 \, \mathrm{I}_0 \, \frac{a^2}{4} \, \cos\omega t \end{aligned} \tag{7.122}$$

where the indeterminate form $x^2 \ln x$ as $x \to 0$ must be evaluated by L'Hopital's rule (write it as $(\ln x)/(1/x^2)$) to be seen to vanish.

We did not include the displacement current in the calculation of the magnetic field in the system. Is that a problem?

Well, the problem is, in principle, even worse: we ought to include the displacement current in the calculation of $\vec{B}$, but then our calculation of $\vec{E}$ via Faraday's Law needs to also be corrected for the $\vec{B}$ due to the displacement current, yielding a correction to $\vec{E}$, which itself will yield a correction to $\vec{J}_d$, and so on. The proper way to handle this is to develop the formalism for electromagnetic waves, where we self-consistently solve all of Maxwell's Equations.

For now, it is instructive to look at the relative size of $\vec{J}_d$ and $\vec{J}$ so we can understand why these corrections are small and thus why our previous results, while not precisely correct, are an excellent approximation.

The ratio of the amplitudes of the displacement current and the true current, up to factors of order unity, is

$$\frac{\mathrm{I}_d(t)}{\mathrm{I}(t)} = \frac{\mu_o\,\epsilon_o\,\omega^2\,a^2\,\mathrm{I}_0\,\cos\omega t}{\mathrm{I}_0\,\cos\omega t} = \frac{\omega^2\,a^2}{c^2} = \frac{(a/c)^2}{1/\omega^2} \tag{7.123}$$

The numerator of the final expression is the square of the light travel time over the length scale of the problem, a. The denominator is, up to a factor $(2\,\pi)^2$, the square of the oscillation period. Thus, this quantity is a measure of how quasistatic the system is. We have mentioned before that, if $a/c \ll 1/\omega$ is not satisfied, then our quasistatic approximation is invalid. This calculation corroborates that point: if the oscillation period becomes comparable to the light travel time so the system is no longer quasistatic, then the displacement current will approach the real current in magnitude and our prior calculation of $\vec{B}$ ignoring the displacement current will be a bad approximation.

The ratio of the displacement current to the true current scales as ω^2, so one must go to high frequency to notice it. Quantitatively, if we ask how high in frequency one must go to obtain $\mathrm{I}_d/\mathrm{I} = 0.01$ if we take $a = 2$ mm as the dimension of the coaxial conductor, we obtain

$$\nu = \frac{\omega}{2\,\pi} = \frac{1}{2\,\pi}\,\frac{c}{a}\,\sqrt{\frac{\mathrm{I}_d}{\mathrm{I}}} = \frac{1}{2\,\pi}\,\frac{3\,\times\,10^{11}\ \mathrm{mm/s}}{2\ \mathrm{mm}}\,\sqrt{0.01} \approx 2\ \mathrm{GHz} \tag{7.124}$$

GHz oscillators were not available in Faraday's time, so the fact that he did not observe the effects of the displacement current is not surprising.

Maxwell's Equations in Vacuum

Putting it all together, we obtain *Maxwell's Equations*:

$$\boxed{\vec{\nabla}\cdot\vec{E}=\frac{\rho}{\epsilon_o} \qquad \vec{\nabla}\cdot\vec{B}=0 \qquad \vec{\nabla}\times\vec{E}=-\frac{\partial\vec{B}}{\partial t} \qquad \vec{\nabla}\times\vec{B}=\mu_o\,\vec{J}+\epsilon_o\,\mu_o\,\frac{\partial\vec{E}}{\partial t}} \tag{7.125}$$

These, combined with the force law and continuity:

$$\boxed{\vec{F}=q\left(\vec{E}+\vec{v}\times\vec{B}\right)} \qquad \boxed{\vec{\nabla}\cdot\vec{J}=-\frac{\partial\rho}{\partial t}} \tag{7.126}$$

summarize classical electrodynamics in vacuum. (The above explains why the $\vec{v}\times\vec{B}$ term is not needed explicitly in the differential version of Faraday's Law: it is a consequence of the force law, not of Faraday's Law.) We may rewrite the first set of equations in a way that emphasizes better the source terms:

$$\boxed{\vec{\nabla}\cdot\vec{E}=\frac{\rho}{\epsilon_o} \qquad \vec{\nabla}\cdot\vec{B}=0 \qquad \vec{\nabla}\times\vec{E}+\frac{\partial\vec{B}}{\partial t}=0 \qquad \vec{\nabla}\times\vec{B}-\epsilon_o\,\mu_o\,\frac{\partial\vec{E}}{\partial t}=\mu_o\,\vec{J}} \tag{7.127}$$

Maxwell's Equations in Matter

Just as we found it convenient to rewrite the individual equations of electrostatics and magnetostatics using only the free charges and currents, it makes sense to do the same for Maxwell's Equations. The new twists we must take into account are the time-dependence of $\vec{P}$ and $\vec{M}$. (We have already considered time dependence of ρ, $\vec{J}$, $\vec{E}$, and $\vec{B}$ for quasistatic situations, corresponding to all length scales in the system small compared to c/ν.)

How to treat $\vec{P}$ is motivated by the expression

$$\rho_b = -\vec{\nabla} \cdot \vec{P} \tag{7.128}$$

If $\vec{P}$ is time-varying, we expect there to be a current $\vec{J}_p$ associated with the resulting changes in ρ_b. In fact, the above expression suggests a good definition of $\vec{J}_p$:

$$\vec{J}_p = \frac{\partial \vec{P}}{\partial t} \qquad \Longleftrightarrow \qquad \vec{\nabla} \cdot \vec{J}_p = -\vec{\nabla} \cdot \frac{\partial \vec{P}}{\partial t} = -\frac{\partial}{\partial t} \vec{\nabla} \cdot \vec{P} = -\frac{\partial \rho_b}{\partial t} \tag{7.129}$$

That is, the definition on the left naturally gives the continuity relation between $\vec{J}_p$ and ρ_b one would like.

Intuitively, think of $\vec{J_p}$ as follows. Suppose one has a cylinder of polarizable material of length dz and cross-sectional area da and with polarization vector $\vec{P} = P\,\hat{z}$. The definition $\rho_b = -\vec{\nabla}\cdot\vec{P}$ implies that there is a bound surface charge $Q = \sigma\,da = \hat{n}\cdot\vec{P}\,da = \pm P\,da$ at each end. If, for example, we allow $\vec{P}$ to vary sinusoidally, $\vec{P} = \vec{P}_0 \sin\omega t$, which corresponds to the surface charge obeying $Q(t) = P_0\,da\sin\omega t$, then the current is

$$\mathrm{I}_p = \vec{J_p}\cdot\hat{n}\,da = P_0\,da\,\omega\cos\omega t = \frac{dQ}{dt} \tag{7.130}$$

as would be necessary to transfer charge back and forth between the two ends of the cylinder to yield the corresponding time-dependent surface charge. This current is, literally, the motion of the charges that make up the dipoles as they flip back and forth sinuisodally.

Do we have to worry about time dependence of $\vec{M}$? Recall that $\vec{M}$ yields a bound current density

$$\vec{J_b} = \vec{\nabla}\times\vec{M} \tag{7.131}$$

Time dependence of $\vec{M}$ yields time dependence of $\vec{J_b}$, which produces time dependence of $\vec{B}$ and $\vec{H}$. These time dependences are now fully accounted for by Maxwell's Equations.

Let's use all this to rewrite Maxwell's Equations in terms of free charges and currents. The charge and current densities have the following parts:

$$\rho = \rho_f + \rho_b = \rho_f - \vec{\nabla} \cdot \vec{P} \qquad \vec{J} = \vec{J}_f + \vec{J}_b + \vec{J}_p = \vec{J}_f + \vec{\nabla} \times \vec{M} + \frac{\partial \vec{P}}{\partial t} \tag{7.132}$$

Using Gauss's law, $\epsilon_o \vec{\nabla} \cdot \vec{E} = \rho_f - \vec{\nabla} \cdot \vec{P}$, and the definition of the displacement field, $\vec{D} = \epsilon_o \vec{E} + \vec{P}$, we obtain

$$\vec{\nabla} \cdot \vec{D} = \rho_f \tag{7.133}$$

Ampere's Law with the displacement current term is

$$\vec{\nabla} \times \vec{B} = \mu_o \left(\vec{J}_f + \vec{\nabla} \times \vec{M} + \frac{\partial \vec{P}}{\partial t} \right) + \epsilon_o \, \mu_o \, \frac{\partial \vec{E}}{\partial t} \tag{7.134}$$

We use $\vec{B} = \mu_o \left(\vec{H} + \vec{M} \right)$ as well as $\vec{D} = \epsilon_o \vec{E} + \vec{P}$ to obtain

$$\vec{\nabla} \times \vec{H} = \vec{J}_f + \frac{\partial \vec{D}}{\partial t} \tag{7.135}$$

Now it is clear why the last term is called the *displacement current* — it is the apparent current due to the time variation of the displacement vector $\vec{D}$!

Faraday's Law and $\vec{\nabla} \cdot \vec{B} = 0$ are not affected since they do not depend on the free and bound currents. Thus, Maxwell's Equations in matter are (again, putting all the fields on the left sides and the sources on the right):

$$\boxed{\vec{\nabla} \cdot \vec{D} = \rho_f} \qquad \boxed{\vec{\nabla} \cdot \vec{B} = 0} \tag{7.136}$$

$$\boxed{\vec{\nabla} \times \vec{E} + \frac{\partial \vec{B}}{\partial t} = 0} \qquad \boxed{\vec{\nabla} \times \vec{H} - \frac{\partial \vec{D}}{\partial t} = \vec{J}_f} \tag{7.137}$$

Note: the vacuum Maxwell's Equations (Equations 7.125 and 7.127) remain valid, but the above are more practically useful. Similarly, Equations 7.126, the Lorentz Force Law and the continuity equation, remain valid but it is more practically useful to have a version of the continuity equation involving only free currents (note that $\vec{\nabla} \cdot \vec{J}_b = \vec{\nabla} \cdot \vec{\nabla} \times \vec{M} = 0$ by mathematical identity):

$$\boxed{\vec{F} = q\left(\vec{E} + \vec{v} \times \vec{B}\right)} \qquad \boxed{\vec{\nabla} \cdot \vec{J}_f = -\frac{\partial \rho_f}{\partial t}} \tag{7.138}$$

These equations must be supplemented by specific *constitutive relations* between $\vec{E}$ and $\vec{D}$ and between $\vec{B}$ and $\vec{H}$ to completely specify the behavior (and, of course, boundary conditions must be provided). For linear media, these relations are:

$$\boxed{\textbf{linear media:} \qquad \vec{P} = \chi_e \epsilon_o \vec{E} \qquad \vec{M} = \chi_m \vec{H} \qquad \vec{D} = \epsilon \vec{E} \qquad \vec{B} = \mu \vec{H}} \tag{7.139}$$

Boundary Conditions for Maxwell's Equations

For both Maxwell's Equations in vacuum and in matter, we need to review how the fields change between regions that may be separated by surface charge and current densities and which may have different polarization and magnetization.

We recall from our prior calculations of this type that the discontinuity in the normal component of a field is determined by its divergence and the discontinuity in the tangential component by its curl. We also recall charge and current densities can become δ-function singular on a boundary but fields cannot. Thus:

- An integral of charge density over a volume containing a boundary reduces, as the height of the volume normal to the boundary is shrunk to zero, to the surface charge density integrated over the intersection of the volume with the boundary. The volume component of the charge density yields zero contribution.
- An integral of a current density through an area reduces, as the width of the area normal to the boundary shrinks to zero, to the surface current density passing through the area. The area component of the current density yields zero contribution.

Fields themselves never have singularities like this, so any integral of a field vanishes as the volume or area is shrunk to zero. Hence, the addition of the displacement current does not modify the boundary conditions we have calculated in the past! For Maxwell's Equations in matter, $\vec{J}_p$ has been introduced but it also cannot have δ-function singularity because it is based on $\vec{P}$, which we have defined to be a volume density of dipoles, and so its contribution to $\partial\vec{D}/\partial t$ is consistent with the above.

Individual electric or magnetic dipoles, or a line or sheet of them, which could in principle yield δ-function contributions to $\vec{P}$ or $\vec{M}$ and thus to $\vec{D}$ or $\vec{H}$, cannot be handled using our macroscopic picture of polarizable and magnetizable materials because the fields are δ-function like on the dipoles only (*e.g.*, Griffiths Problems 3.48 and 5.61) and our macroscopic picture assumes that it is valid to average over infinitesimally small volumes. They affect how the field changes at a boundary over an infinitesimally small area, and so we should not even consider their effect. If we want to consider their impact on the fields in such regions, they would need to be treated as free charges and currents and would again not affect the above statements about fields.

Section 8
Conservation Laws

Motivation and Analogy: Conservation of Charge

Back when we first discussed the Lorentz Force, we discussed conservation of charge and the continuity equation:

$$\vec{\nabla} \cdot \vec{J}(\vec{r}) = -\frac{\partial \rho(\vec{r})}{\partial t} \tag{8.1}$$

This is an interesting equation because it enforces *local* conservation of charge: not only is there no creation or destruction of charge over the whole universe, there is also no creation or destruction of charge at a given point. Charge cannot jump from one place to another without a current flowing to move that charge.

In electrodynamics, we want to ask the same question for energy and momentum because we want to understand whether the fields we have constructed have true physical meaning or are just mathematical constructs. Determining whether they carry energy and momentum is one way to answer that question, and such a consideration leads to the question of conservation of these quantities.

We will do all this in vacuum. It of course applies to polarizable and magnetizable materials, too, since our study of them is just a rewriting of our vacuum equations in a more convenient form.

Poynting's Theorem: Conservation of Energy

We have shown that the work required to set up distributions of charge or current is

$$W_e = \frac{\epsilon_o}{2}\int d\tau\,|\vec{E}|^2 \qquad W_m = \frac{1}{2\,\mu_o}\int d\tau\,|\vec{B}|^2 \tag{8.2}$$

Recall that this is the work needed to move new charge in from infinity due to the repulsion from the charge already there or the work that needs to be done to raise a current from zero to its final value against a back emf (induced electric field). It is thus natural to expect that the total energy density in the electric and magnetic fields is

$$u_{field} = \frac{1}{2}\left(\epsilon_o|\vec{E}|^2 + \frac{1}{\mu_o}|\vec{B}|^2\right) \tag{8.3}$$

We will show this is valid by considering the exchange of energy between the fields and charges/currents. We use the term *electromagnetic field* to reflect the fact that the fields influence each other and their energies are on the same footing.

Given a single particle of charge q acted on by the electromagnetic field, the work done on it as it moves by $d\vec{\ell}$ is

$$dW = \vec{F} \cdot d\vec{\ell} = q\left(\vec{E} + \vec{v} \times \vec{B}\right) dt = q\,\vec{E} \cdot \vec{v}\,dt \tag{8.4}$$

Now, if we consider a continuous distribution of charge and current, we may replace $q = \rho\,d\tau$ and $\rho\,\vec{v} = \vec{J}$, giving that the power is (as we saw from Ohm's Law)

$$\frac{dW}{dt} = \int d\tau\,\vec{E} \cdot \vec{J} \tag{8.5}$$

Let's manipulate the integrand using Ampere's Law:

$$\vec{E} \cdot \vec{J} = \frac{1}{\mu_o}\,\vec{E} \cdot \left(\vec{\nabla} \times \vec{B}\right) - \epsilon_o\,\vec{E} \cdot \frac{\partial \vec{E}}{\partial t} \tag{8.6}$$

One subtlety here: we started off talking about $\vec{J}$ being acted upon by an electromagnetic field, and now it seems like we are treating $\vec{J}$ as the source of that field. It is not the sole source of the field because, now with the displacement current term combined with Faraday's Law, there can be electric and magnetic fields that are sourced by each other's time variation rather than by physical currents. The above substitution is nevertheless valid because the second term subtracts off the displacement current term that is due to changing fields rather than physical current: one should not be able to do work on the displacement current!

Another subtlety is the issue of whether $\vec{J}$ can create fields that do work on itself. This is entirely possible, as we saw in the example of the time-varying current in the coaxial conductor: a time-varying current generated a time-varying magnetic field that generated a time-varying electric field aligned with the original current. If there were no battery driving the current, then the work being done by the field on the current should reduce the energy in the current in exactly the way that would be needed to conserve energy. Of course, if a battery is involved, then it can supply energy and we do not expect the energy of the currents and fields alone to be conserved.

Returning to our expression for $\vec{E}\cdot\vec{J}$, we can use the product rule $\vec{\nabla}\cdot(\vec{a}\times\vec{b}) = \vec{a}\cdot(\vec{\nabla}\times\vec{b}) - \vec{b}\cdot(\vec{\nabla}\times\vec{a})$ to rewrite it and then use Faraday's Law:

$$\vec{E}\cdot\vec{J} = \frac{1}{\mu_o}\,\vec{B}\cdot\left(\vec{\nabla}\times\vec{E}\right) - \frac{1}{\mu_o}\vec{\nabla}\cdot\left(\vec{E}\times\vec{B}\right) - \epsilon_o\,\vec{E}\cdot\frac{\partial\vec{E}}{\partial t} \tag{8.7}$$

$$= -\frac{1}{\mu_o}\,\vec{B}\cdot\frac{\partial\vec{B}}{\partial t} - \epsilon_o\,\vec{E}\cdot\frac{\partial\vec{E}}{\partial t} - \frac{1}{\mu_o}\vec{\nabla}\cdot\left(\vec{E}\times\vec{B}\right) \tag{8.8}$$

$$= -\frac{\partial}{\partial t}\frac{1}{2}\left(\epsilon_o|\vec{E}|^2 + \frac{1}{\mu_o}|\vec{B}|^2\right) - \frac{1}{\mu_o}\vec{\nabla}\cdot\left(\vec{E}\times\vec{B}\right) \tag{8.9}$$

Incorporating the above and applying the divergence theorem to the last term, we thus obtain *Poynting's Theorem*:

$$\boxed{\frac{dW}{dt} = -\left[\frac{d}{dt}\int_{\mathcal{V}} d\tau\,\frac{1}{2}\left(\epsilon_o|\vec{E}|^2 + \frac{1}{\mu_o}|\vec{B}|^2\right) + \oint_{\mathcal{S}(\mathcal{V})} da\,\widehat{n}\cdot\vec{S}\right]} \tag{8.10}$$

with the *Poynting vector* defined to be $\boxed{\vec{S} = \frac{1}{\mu_o}\left(\vec{E}\times\vec{B}\right)}$ (8.11)

Poynting's Theorem says that *the work per unit time done on the charges and currents in a volume $\mathcal{V}$ by electromagnetic forces is equal to negative of the sum of the change per unit time of the energy in the fields and the energy flowing outward through the surface of $\mathcal{V}$.* $\vec{S}$ has units of energy per unit time per unit area and is considered the *energy flux density*.

Another useful form is given by putting the field energy density term on the left side:

$$\frac{dW}{dt} + \frac{d}{dt}\int_{\mathcal{V}} d\tau\, \frac{1}{2}\left(\epsilon_o |\vec{E}|^2 + \frac{1}{\mu_o}|\vec{B}|^2\right) = -\frac{1}{\mu_o}\oint_{\mathcal{S}(\mathcal{V})} da\, \hat{n}\cdot\vec{S} \tag{8.12}$$

$$\boxed{\frac{d}{dt}\left(E_{mech} + E_{field}\right) = -\oint_{\mathcal{S}(\mathcal{V})} da\, \hat{n}\cdot\vec{S}} \tag{8.13}$$

The rate of change of the total energy in the volume is given by the flux of the Poynting vector through the boundary of the volume — this much more explicitly puts mechanical and field energy on the same footing and shows that both can be transported by the Poynting flux.

Note that this allays our fears about a current doing work on itself: while it may do so, energy remains conserved as long as one takes into account the field energy.

We may write both versions in a local form by recognizing that the volume being integrated over is arbitrary. If we define $u_{mech}(\vec{r})$ to be the density of mechanical energy ($W = \int_{\mathcal{V}} d\tau\, u_{mech}$) as a function of position and

$$u_{field} = \frac{1}{2}\left(\epsilon_o |\vec{E}(\vec{r})|^2 + \frac{1}{\mu_o}|\vec{B}(\vec{r})|^2\right) \tag{8.14}$$

to be the energy density of the electromagnetic field, then our two versions of Poynting's theorem yield the local relations (after converting the surface integral of $\vec{S}$ back to a volume integral using the divergence theorem):

$$\boxed{\frac{\partial u_{mech}(\vec{r})}{\partial t} = -\frac{\partial u_{field}(\vec{r})}{\partial t} - \vec{\nabla}\cdot\vec{S}(\vec{r}) \quad \Longleftrightarrow \quad \frac{\partial}{\partial t}\left[u_{mech}(\vec{r}) + u_{field}(\vec{r})\right] = -\vec{\nabla}\cdot\vec{S}(\vec{r})} \tag{8.15}$$

This is the kind of local conservation theorem we wanted, relating *the rate of change of a density (here the energy density) to the divergence of a current density (here the Poynting vector)*.

When there is no change in mechanical energy — *e.g.*, in empty space — then we can specialize the above to obtain the *continuity equation for the energy of the electromagnetic field*:

$$\boxed{u_{mech} = 0: \quad \frac{dE_{field}}{dt} = -\oint_{\mathcal{S}(\mathcal{V})} da\,\hat{n}\cdot\vec{S} \quad \Longleftrightarrow \quad \vec{\nabla}\cdot\vec{S}(\vec{r}) = -\frac{\partial u_{field}(\vec{r})}{\partial t}} \tag{8.16}$$

Example 8.1: Power Transported Down a Coaxial Cable (Griffiths Problem 8.1)

Consider a coaxial cable with central conductor of diameter a and outer conductor of radius b and zero thickness. A *static* current flows along $+\widehat{z}$ on the central conductor and back along $-\widehat{z}$ on the outer shell. We used a similar geometry in a previous example, but with a time-varying current in that case. The inner conductor is held at voltage V and the outer conductor at $V = 0$ (ground) at one end of the cylinder, and there is a resistive sheet of *sheet* conductivity $\sigma_{\square}$ (definition to be provided) capping the other end.

Because the inner conductor is assumed to have infinite conductivity, there can be no electric field inside and thus all the current must flow on the surface (consequence of Ohm's Law: $\vec{J} = 0$ because $\vec{E} = 0$). The calculation of the magnetic field is thus the same as the prior example in the same geometry with time-varying current. The magnetic field between the conductors is

$$\vec{B}(s) = \frac{\mu_o \, \mathrm{I}}{2\,\pi\,s}\,\widehat{\phi} \tag{8.17}$$

where I is the current due to V (value to be determined). In the prior example, we did not explicitly have a voltage on the inner conductor (effectively, the conductivity of the sheet at the end was infinite).

Here, since we have such a voltage, there is a line charge density on the inner conductor and a radial electric field. You are no doubt familiar with the Gauss' Law calculation of this configuration, which yields

$$\vec{E}(s) = \frac{\lambda}{2\,\pi\,\epsilon_o\,s}\,\widehat{s} \tag{8.18}$$

Let's find λ by matching to the applied voltage. The potential and field are

$$V(s) \propto \ln s \qquad V = V(a) - V(b) = \frac{\lambda}{2\,\pi\,\epsilon_o}\,\ln\frac{b}{a} \quad \Longrightarrow \quad \vec{E}(s) = \frac{V}{s\,\ln\frac{b}{a}}\,\widehat{s} \tag{8.19}$$

The Poynting vector is

$$\vec{S} = \frac{1}{\mu_o}\,\vec{E}\times\vec{B} = \frac{\mathrm{I}\,V}{2\,\pi\,s^2\ln\frac{b}{a}}\,\widehat{z} \tag{8.20}$$

The energy and energy current are between the conductors, not in them! The power flowing down the cable is found by integrating the Poynting vector over the cross-sectional area where the fields are:

$$P = \int_{\mathcal{S}} da\,\widehat{n}\cdot\vec{S}(s) = \int_a^b s\,ds\int_0^{2\,\pi} d\phi\,\frac{\mathrm{I}\,V}{2\,\pi\,s^2\ln\frac{b}{a}} = \frac{\mathrm{I}\,V}{\ln\frac{b}{a}}\int_a^b\frac{ds}{s} = \mathrm{I}\,V \tag{8.21}$$

Let's calculate the power dissipated in the resistive sheet at the end. The sheet does not disturb the potential because the sheet continues to satisfy Laplace's equation with the same boundary conditions in s: $V(a) = V$, $V(b) = 0$. Therefore, our electric field above is valid in the conducting sheet, and the surface current density and total current are

$$\vec{K}(s) = \sigma_{\square}\,\vec{E}(s) = \frac{\sigma_{\square}\,V}{s\,\ln\frac{b}{a}}\,\widehat{s} \tag{8.22}$$

where $\sigma_{\square}$ is the *conductivity per square*, which can be thought of as $\sigma_{\square} = \lim_{t\to 0}\sigma/t$ where σ is the usual conductivity and t is the thickness of the sheet. The total current is

$$\mathrm{I} = \int_0^{2\,\pi} s\,d\phi\,K(s) = 2\,\pi\,s\,K(s) = \frac{2\,\pi\,\sigma_{\square}\,V}{\ln\frac{b}{a}} \tag{8.23}$$

We do not need it, but it is interesting to note that the resistance is

$$R = \frac{V}{\mathrm{I}} = \frac{\ln \frac{b}{a}}{2\,\pi\,\sigma_{\square}} \tag{8.24}$$

The power dissipated in the resistor is

$$P = \int_{\mathcal{S}} da\,\vec{K}(s)\cdot\vec{E}(s) = \int_a^b s\,ds \int_0^{2\,\pi} d\phi\,\sigma_{\square}\left(\frac{V}{s\,\ln\frac{b}{a}}\right)^2 \tag{8.25}$$

$$= 2\,\pi\,\sigma_{\square}\left(\frac{V}{\ln\frac{b}{a}}\right)^2 \int_a^b \frac{ds}{s} = \frac{2\,\pi\,\sigma_{\square}V^2}{\ln\frac{b}{a}} = \mathrm{I}\,V = \mathrm{I}^2 R = \frac{V^2}{R} \tag{8.26}$$

as expected since this is the power coming down the central conductor and it cannot go beyond the resistive sheet since the fields go to zero out there (no current or charge density beyond the sheet).

Lecture 39:

Conservation Laws (Optional Material):
Stress Tensor/Linear Momentum
Torque Tensor/Angular Momentum

Relativistic Conservation Laws (Optional Material):
Maxwell Energy-Momentum Tensor and Conservation of Energy-Momentum
Relativistic Angular Momentum Tensor

Applications of Radiation: Diffraction (Optional Material)

Date Revised: 2023/06/07 07:00
Date Given: TBD

These three sections,
consisting of material colored green and orange,
is optional!

The Maxwell Stress Tensor: Conservation of Linear Momentum (Skip)

We showed in the previous section that fields carry energy and that one must account for that energy in order for conservation of energy to hold. The natural next question to ask is whether the electromagnetic fields carry momentum. The matter and fields are related by the fields exerting forces on the matter, so let's use these forces to connect the momentum of the matter and fields. The Lorentz Force Law is

$$\frac{d\vec{P}_{mech}}{dt} = \vec{F} = q\left(\vec{E} + \vec{v} \times \vec{B}\right) \tag{8.27}$$

Integrating this over a charge and current density gives

$$\frac{d\vec{P}_{mech}}{dt} = \int_{\mathcal{V}} d\tau \left(\rho\,\vec{E} + \vec{J} \times \vec{B}\right) \tag{8.28}$$

Using Maxwell's Equations, we can write this purely in terms of the fields:

$$\frac{d\vec{P}_{mech}}{dt} = \int_{\mathcal{V}} d\tau \left(\epsilon_o \left[\vec{\nabla} \cdot \vec{E}\right]\vec{E} + \left[\frac{1}{\mu_o}\vec{\nabla} \times \vec{B} - \epsilon_o \frac{\partial \vec{E}}{\partial t}\right] \times \vec{B}\right) \tag{8.29}$$

We recall the same subtleties as for energy: ρ and $\vec{J}$ now being taken as source of fields, and the last term subtracts off the displacement current since the magnetic field exerts no force on it.

After a remarkable amount of manipulation that we will not reproduce here—see Griffiths §8.2.2 or Jackson §6.8—one arrives at

$$\boxed{\frac{d\vec{P}_{mech}}{dt} = \int_{\mathcal{V}} d\tau \left[\vec{\nabla} \cdot \underline{\underline{\mathcal{T}}} - \epsilon_o \, \mu_o \, \frac{\partial \vec{S}}{\partial t} \right]} \tag{8.30}$$

where $\underline{\underline{\mathcal{T}}}$ is the *Maxwell Stress Tensor*

$$\boxed{\underline{\underline{\mathcal{T}}}(\vec{r}) = \sum_{i,j=1}^{3} T_{ij}(\vec{r}) \, \hat{r}_i \, \hat{r}_j \qquad T_{ij}(\vec{r}) = \epsilon_o \left[E_i \, E_j - \frac{1}{2} \, \delta_{ij} \, E^2 \right] + \frac{1}{\mu_o} \left[B_i \, B_j - \frac{1}{2} \, \delta_{ij} \, B^2 \right]} \tag{8.31}$$

(we do not show the fields' dependence on position for brevity) and where the vector dot products and divergence of $\underline{\underline{\mathcal{T}}}$ are given by

$$\vec{a} \cdot \underline{\underline{\mathcal{T}}} = \sum_{i=1}^{3} a_i \, T_{ij} \, \hat{r}_j \qquad \underline{\underline{\mathcal{T}}} \cdot \vec{a} = \sum_{j=1}^{3} \hat{r}_i \, T_{ij} \, a_j \qquad \vec{\nabla} \cdot \underline{\underline{\mathcal{T}}} = \sum_{i=1}^{3} \hat{r}_j \frac{\partial}{\partial r_i} T_{ij} \tag{8.32}$$

Note that T_{ij} is symmetric in its indices. We are not terribly concerned in this course with the transformation properties of scalars, vectors, and tensors under coordinate system rotations, so we will not comment further on what a tensor is. Recall we encountered the quadrupole moment tensor earlier.

Using the divergence theorem on the first term and moving the time derivative in the second term outside the integral, we obtain

$$\boxed{\frac{d\vec{P}_{mech}}{dt} = \oint_{\mathcal{S}(\mathcal{V})} da\, \hat{n}(\vec{r}) \cdot \underline{\underline{\mathcal{T}}}(\vec{r}) - \epsilon_o\, \mu_o \frac{d}{dt} \int_{\mathcal{V}} d\tau\, \vec{S}(\vec{r})} \tag{8.33}$$

This equation states that *the rate of change of the mechanical momentum in a volume $\mathcal{V}$ is equal to the integral over the surface of the volume of the stress tensor's flux through that surface minus the rate of change of the volume integral of the Poynting vector.*

Let us consider a situation in which the second term vanishes and we are left with the flux of $\underline{\underline{\mathcal{T}}}$ over the surface. This justifies the naming of $\underline{\underline{\mathcal{T}}}$: it gives the force per unit area due to the electromagnetic fields, or the *stress*. T_{ij} is the force per unit area acting in the ith direction on an area element who normal is in the jth direction. The diagonal elements are *pressures* and the off-diagonal forces are *shears*. More generally, the force per unit area in the $\hat{n}_1$ direction on an area element whose normal is in the $\hat{n}_2$ direction (not necessarily parallel or perpendicular to $\hat{n}_1$), or vice versa, is

$$\frac{F(\vec{r}, \hat{n}_1, \hat{n}_2)}{A} = \hat{n}_1 \cdot \underline{\underline{\mathcal{T}}}(\vec{r}) \cdot \hat{n}_2 \tag{8.34}$$

We may abstract out mechanical momentum and force densities $\vec{p}_{mech}(\vec{r})$ and $\vec{f}(\vec{r})$; *i.e.*, per unit volume expressions:

$$\boxed{\vec{p}_{mech}(\vec{r}) \equiv \rho_m(\vec{r})\vec{v}(\vec{r}) \qquad \vec{f}(\vec{r}) \equiv \vec{\nabla} \cdot \underline{\underline{\mathcal{T}}}(\vec{r}) - \epsilon_o\,\mu_o\,\frac{\partial \vec{S}(\vec{r})}{\partial t}} \tag{8.35}$$

where $\rho_m(\vec{r})$ is the *mass density*
$\vec{v}(\vec{r})$ is the *velocity field* of the mass density

We may conclude that these quantities are related locally because of the arbitrariness of the volume over which we are integrating:

$$\int_{\mathcal{V}} d\tau\,\frac{\partial \vec{p}_{mech}}{\partial t} = \frac{d}{dt}\int_{\mathcal{V}} d\tau\,\vec{p}_{mech} = \frac{d\vec{P}_{mech}}{dt} = \int_{\mathcal{V}} d\tau\,\vec{f} \tag{8.36}$$

$$\Longrightarrow \qquad \boxed{\frac{\partial \vec{p}_{mech}(\vec{r})}{\partial t} = \vec{f}(\vec{r}) = \vec{\nabla} \cdot \underline{\underline{\mathcal{T}}}(\vec{r}) - \epsilon_o\,\mu_o\,\frac{\partial \vec{S}(\vec{r})}{\partial t}} \tag{8.37}$$

This is the kind of conservation law we wanted to get to, a local one that relates *the rate of change of the local momentum density to the divergence of the local stress tensor and the rate of change of the Poynting vector*. It can also be viewed as a local force law, the generalization of Newton's Second Law.

As we did with the energy, and motivated by the appearance of a time derivative on the right side, we may rewrite the above as

$$\frac{d}{dt}\left(\vec{P}_{mech} + \epsilon_o\,\mu_o \int_{\mathcal{V}} d\tau\,\vec{S}\right) = \oint_{\mathcal{S}(\mathcal{V})} da\,\widehat{n}\cdot\underline{\underline{\mathcal{T}}} \tag{8.38}$$

We are thus motivated to define the *linear momentum density* and *linear momentum* of the electromagnetic field as

$$\boxed{\vec{p}_{field}(\vec{r}) \equiv \vec{g}(\vec{r}) \equiv \epsilon_o\,\mu_o\,\vec{S}(\vec{r}) = \epsilon_o\,\vec{E}(\vec{r})\times\vec{B}(\vec{r}) \qquad \vec{P}_{field} = \int_{\mathcal{V}} d\tau\,\vec{g}} \tag{8.39}$$

With that definition, we obtain

$$\boxed{\frac{d}{dt}\left(\vec{P}_{mech}+\vec{P}_{field}\right)=\oint_{\mathcal{S}(\mathcal{V})} da\,\hat{n}\cdot\underline{\underline{\mathcal{T}}} \quad\Longleftrightarrow\quad \frac{\partial}{\partial t}\left[\vec{p}_{mech}(\vec{r})+\vec{g}(\vec{r})\right]=\vec{\nabla}\cdot\underline{\underline{\mathcal{T}}}(\vec{r})} \tag{8.40}$$

We thus see that *the rate of change of the total (mechanical + field) linear momentum in a volume is given by the integral of the stress tensor over the surface, or that the rate of change of the total momentum density at a point is given by the divergence of the stress tensor at that point*. The stress tensor is thus seen to be the momentum current density in the same way that $\vec{J}$ is the electric current density and $\vec{S}$ is the energy current density (up to a sign): all satisfy local continuity equations.

The second equation can also be considered a generalized force law, where now we consider the rate of change of the momentum of both the particles and the fields, with $\vec{\nabla}\cdot\underline{\underline{\mathcal{T}}}$ being the "force" that acts on both.

When there is no change in mechanical momentum—*e.g.*, in empty space—we obtain the *continuity equation for the linear momentum of the electromagnetic field*:

$$\boxed{\vec{p}_{mech}=0: \quad \frac{d\vec{P}_{field}}{dt}=\oint_{\mathcal{S}(\mathcal{V})} da\,\hat{n}\cdot\underline{\underline{\mathcal{T}}} \quad\Longleftrightarrow\quad \vec{\nabla}\cdot\underline{\underline{\mathcal{T}}}(\vec{r})=\frac{\partial\vec{g}(\vec{r})}{\partial t}} \tag{8.41}$$

It is interesting to note that both $\vec{S}$ and $\underline{\underline{\mathcal{T}}}$ play two roles:

- $\vec{S}$ is the power per unit area transported by the electromagnetic field, while $\vec{g} = \epsilon_o \, \mu_o \, \vec{S}$ is the linear momentum per unit volume stored in the field. This intimate connection between energy and momentum for the electromagnetic field reflects the photon's masslessness in quantum field theory.
- Similarly, $\underline{\underline{\mathcal{T}}}$ plays two roles; both as a force per unit area (the stress) applied by the electromagnetic field as well as the momentum current density carried by the electromagnetic field (with a minus sign; units of momentum per unit area per unit time). This makes sense: for the electromagnetic field to exert a force, it must provide momentum.

Note this issue of the sign. If we wanted $\underline{\underline{\mathcal{T}}}$ to have a continuity equation like current and energy, where the rate of change of the conserved quantity is equal to the negative of the divergence of the current (loss of conserved quantity corresponds to outflow of current), we would have had to define $\underline{\underline{\mathcal{T}}}$ with the opposite sign. But the sign given ensures that $\underline{\underline{\mathcal{T}}}$ can be used to calculate forces without a sign flip. This makes sense: $\underline{\underline{\mathcal{T}}}$ pointing into a volume should have a positive surface integral so that it indicates it is adding momentum to the volume. The only way out of this choice would be if we wanted to flip the sign and interpret $\underline{\underline{\mathcal{T}}}$ as the force that the mechanical system exerts on the field (and then the continuity equation would behave the way we want), but that would be nonintuitive since we generally want to calculate the forces the field exerts on the mechanical system.

Example 8.2: Magnetic Force Between Two Spinning Charged Hemispheres (Griffiths Problem 8.3)

Given two hemispherical shells of radius R and uniform surface charge density σ spinning at angular frequency ω about the z axis, what is the magnetic force between the north and south hemispheres? (Griffiths Example 8.2 calculates the electrostatic force for a similar situation.)

We have calculated the magnetic field for a similar configuration when we calculated the field of the uniformly magnetized sphere, which was

$$\vec{B}(r \leq R) = \frac{2}{3}\,\mu_o\,M\,\hat{z} \tag{8.42}$$

$$\vec{B}(r \geq R) = \frac{\mu_o}{4\,\pi}\,\frac{3\,(\vec{m}\cdot\hat{r})\hat{r} - \vec{m}}{r^3} \quad \text{with} \quad \vec{m} = \frac{4}{3}\,\pi\,R^3 M\,\hat{z} \tag{8.43}$$

The surface current was $\vec{K} = \hat{\phi}\,M\sin\theta$. In the new problem, the surface current is $\vec{K} = \hat{\phi}\,\sigma\,\omega\,R\sin\theta$, so we just need to replace M with $\sigma\,\omega\,R$, giving

$$\vec{B}(r \leq R) = \frac{2}{3}\,\mu_o\,\sigma\,\omega\,R\,\hat{z} \tag{8.44}$$

$$\vec{B}(r \geq R) = \frac{\mu_o}{4\,\pi}\,\frac{3\,(\vec{m}\cdot\hat{r})\hat{r} - \vec{m}}{r^3} \quad \text{with} \quad \vec{m} = \frac{4}{3}\,\pi\,R^4\sigma\,\omega\,\hat{z} \tag{8.45}$$

To calculate the force, we would nominally expect to calculate the flux of the stress tensor over the hemisphere (the plane at $z = 0$ for $r < R$ and the hemispherical shell $r = R$ at $z > 0$). However, the derivation implies that any volume containing the matter on which we would like to calculate the force suffices for the calculation. So let's do the calculation more easily by setting the surface to be the $z = 0$ plane. The force will only be in the z direction by symmetry, so we need only the T_{3i} components. Moreover, because the plane we want to do the calculation for has a surface normal only in the z direction, we can restrict to the T_{33} component:

$$T_{33} = \frac{1}{2\,\mu_o}\,B_z^2 \quad \Longrightarrow \quad T_{33}(r < R, z = 0) = \frac{2}{9}\,\mu_o\,\sigma^2\omega^2\,R^2 \tag{8.46}$$

$$T_{33}(r > R, z = 0) = \frac{\mu_o\,\sigma^2\,\omega^2\,R^8}{18\,r^6} \tag{8.47}$$

We can do the area integral easily ($\widehat{n} = -\widehat{z}$ because we want the force on the upper half space and $-\widehat{z}$ is the outward surface normal):

$$F_z = -\int_0^{2\,\pi} d\phi \left[\int_0^R r\,dr\,T_{33}(r < R, z = 0) + \int_R^\infty r\,dr\,T_{33}(r > R, z = 0)\right] \tag{8.48}$$

$$= -2\,\pi\left[\frac{R^2}{2}\,\frac{2}{9}\,\mu_o\,\sigma^2\omega^2\,R^2 + \frac{\mu_o\,\sigma^2\,\omega^2\,R^8}{72\,R^4}\right] = -\frac{\pi}{4}\,\mu_o\,\sigma^2\,R^4\,\omega^2 \tag{8.49}$$

Conservation of Angular Momentum (Skip)

One can go back and write analogues of everything we did for linear momentum for the case of angular momentum. The key point is that the manipulations that led us to Equation 8.30 did not rely on any transformations of integrals; we just needed to manipulate the integrand. Those manipulations remain valid, but now with a $\vec{r}\times$ in front inside the integral. That is, we start with

$$\frac{d\vec{L}_{mech}}{dt} = \vec{N} = \vec{r}\times\vec{F} = \vec{r}\times q\left(\vec{E}+\vec{v}\times\vec{B}\right) \tag{8.50}$$

Again, we integrate over the charge and current density to obtain

$$\frac{d\vec{L}_{mech}}{dt} = \int_{\mathcal{V}} d\tau\, \vec{r}\times\left(\rho\,\vec{E}+\vec{J}\times\vec{B}\right) \tag{8.51}$$

Then we perform the same manipulations of the expression in parentheses as before, obtaining

$$\frac{d\vec{L}_{mech}}{dt} = \int_{\mathcal{V}} d\tau \left[\vec{r}\times\left(\vec{\nabla}\cdot\underline{\underline{\mathcal{T}}}\right) - \epsilon_o\,\mu_o\,\frac{\partial}{\partial t}\left(\vec{r}\times\vec{S}\right)\right] \tag{8.52}$$

Let's manipulate the expression $\vec{r} \times \left(\vec{\nabla} \cdot \underline{\underline{\mathcal{T}}} \right)$: we would obviously like to turn it into a pure divergence. Using Equation 8.32,

$$\vec{r} \times \left(\vec{\nabla} \cdot \underline{\underline{\mathcal{T}}}(\vec{r}) \right) = \vec{r} \times \sum_{i,j=1}^{3} \widehat{r}_j \frac{\partial T_{ij}}{\partial r_i} = \sum_{i,j,k=1}^{3} \widehat{r}_k \times \widehat{r}_j \, r_k \frac{\partial T_{ij}}{\partial r_i} \tag{8.53}$$

$$= \sum_{i,j,k=1}^{3} \widehat{r}_k \times \widehat{r}_j \left(\frac{\partial}{\partial r_i} r_k T_{ij} \right) - \sum_{i,j,k=1}^{3} \widehat{r}_k \times \widehat{r}_j \, T_{ij} \frac{\partial r_k}{\partial r_i} \tag{8.54}$$

$$= \sum_{i,j,k=1}^{3} \frac{\partial}{\partial r_i} \left(r_k \, \widehat{r}_k \times T_{ij} \, \widehat{r}_j \right) - \sum_{i,j,k=1}^{3} \widehat{r}_k \times \widehat{r}_j \, T_{ij} \, \delta_{ik} \tag{8.55}$$

$$= - \sum_{i,j,k=1}^{3} \frac{\partial}{\partial r_i} \left(T_{ij} \, \widehat{r}_j \times r_k \, \widehat{r}_k \right) - \sum_{i,j,k=1}^{3} \widehat{r}_k \times \widehat{r}_j \, T_{kj} \tag{8.56}$$

$$= \vec{\nabla} \cdot \left(-\underline{\underline{\mathcal{T}}}(\vec{r}) \times \vec{r} \right) = \vec{\nabla} \cdot \underline{\underline{\mathcal{M}}}(\vec{r}) \qquad \text{with} \quad \underline{\underline{\mathcal{M}}}(\vec{r}) = -\underline{\underline{\mathcal{T}}}(\vec{r}) \times \vec{r} \tag{8.57}$$

(we did not show the explicit $\vec{r}$ dependence for the intermediate steps for brevity) where the second term in the penultimate line vanishes because it is the product of quantities that are antisymmetric in j and k ($\widehat{r}_k \times \widehat{r}_j$) and symmetric in j and k (T_{kj}). $\underline{\underline{\mathcal{M}}}$ is the analogue of the stress tensor, but now for torque, which we will call the *torque tensor*. We reordered $\vec{r}$ and $\underline{\underline{\mathcal{T}}}$ to obtain $-\underline{\underline{\mathcal{T}}} \times \vec{r}$ rather than $\vec{r} \times \underline{\underline{\mathcal{T}}}$ so the coordinate index of the divergence matches up with the first coordinate index of $\underline{\underline{\mathcal{M}}}$.

Explicitly (undoing the reordering of $\underline{\underline{\mathcal{T}}}$ and $\vec{r}$ so we get $\vec{r}\times$, not $\times\vec{r}$, below, and not showing explicitly the position dependence for brevity):

$$\underline{\underline{\mathcal{M}}} = -\underline{\underline{\mathcal{T}}} \times \vec{r} = -\sum_{i,j,k=1}^{3} T_{ij}\,\hat{r}_i\,\hat{r}_j \times r_k\,\hat{r}_k = -\sum_{i,j,k,m=1}^{3} \hat{r}_i\,\hat{r}_m\,\epsilon_{mjk}\,T_{ij}\,r_k \tag{8.58}$$

$$= \sum_{i,j,k,m=1}^{3} \hat{r}_m\,\hat{r}_i\,r_k\,\epsilon_{mkj}\left(\epsilon_o\left[E_j\,E_i - \frac{1}{2}\,\delta_{ji}\,E^2\right] + \frac{1}{\mu_o}\left[B_j\,B_i - \frac{1}{2}\,\delta_{ji}\,B^2\right]\right) \tag{8.59}$$

$$= \epsilon_o\left[\vec{r}\times\vec{E}\right]\vec{E} + \frac{1}{\mu_o}\left[\vec{r}\times\vec{B}\right]\vec{B} - \frac{1}{2}\sum_{i,k,m=1}^{3} \hat{r}_m\,\hat{r}_i\,r_k\,\epsilon_{mki}\left[\epsilon_o E^2 + \frac{1}{\mu_o}B^2\right] \tag{8.60}$$

$$= \epsilon_o\left[\vec{r}\times\vec{E}\right]\vec{E} + \frac{1}{\mu_o}\left[\vec{r}\times\vec{B}\right]\vec{B} + \frac{1}{2}\left[\epsilon_o E^2 + \frac{1}{\mu_o}B^2\right]\begin{bmatrix} 0 & z & -y \\ -z & 0 & x \\ y & -x & 0 \end{bmatrix} \tag{8.61}$$

Aside: In Ph106a, one shows that angular momentum is more rigorously written as an antisymmetric second-rank (pseudo)tensor, but, because such an object has only 3 independent quantities, it can be reduced to a (pseudo)vector (first-rank tensor) using cross-product notation. That applies here to both $\vec{L}_{mech}$ and to $\vec{r}\times\vec{S}$. By extrapolation, $\underline{\underline{\mathcal{M}}}$ may be written as an a completely antisymmetric *third-rank* (pseudo)tensor. Since we do not use any of the transformation properties of these objects under rotations in this course, there is no need to use these higher-rank objects and so we stick with the less sophisticated vector notation for cross products. But this concept will return when we consider the relativistic generalization of $\underline{\underline{\mathcal{M}}}$ because the reduction to a second-rank tensor is only possible in three dimensions.

With our expression in terms of the divergence of $\underline{\underline{\mathcal{M}}}$, we may write the analogue of Equation 8.30 for torque:

$$\boxed{\frac{d\vec{L}_{mech}}{dt} = \int_{\mathcal{V}} d\tau \left[\vec{\nabla} \cdot \underline{\underline{\mathcal{M}}} - \epsilon_o \mu_o \frac{\partial}{\partial t}\left(\vec{r} \times \vec{S}\right)\right]} \tag{8.62}$$

Using the divergence theorem, we may rewrite as we did the force equation

$$\boxed{\frac{d\vec{L}_{mech}}{dt} = \oint_{\mathcal{S}(\mathcal{V})} da\, \hat{n}(\vec{r}) \cdot \underline{\underline{\mathcal{M}}}(\vec{r}) - \epsilon_o \mu_o \frac{d}{dt} \int_{\mathcal{V}} d\tau\, \vec{r} \times \vec{S}(\vec{r})} \tag{8.63}$$

We thus have a relation between *the rate of change of mechanical angular momentum and the flux of the torque tensor* $\underline{\underline{\mathcal{M}}}$ *into/out of the volume and the rate of the change of integral of the funny quantity containing the Poynting vector.*

Let's turn this into a differential version. We need to define the mechanical momentum density and the torque density:

$$\boxed{\vec{\ell}_{mech}(\vec{r}) \equiv \vec{r} \times \vec{p}_{mech}(\vec{r}) = \vec{r} \times \rho_m(\vec{r})\vec{v}(\vec{r}) \qquad \vec{n}_{torque}(\vec{r}) \equiv \vec{\nabla} \cdot \underline{\underline{\mathcal{M}}}(\vec{r}) - \epsilon_o\,\mu_o\,\frac{\partial}{\partial t}\left(\vec{r} \times \vec{S}(\vec{r})\right)} \tag{8.64}$$

Then we have

$$\int_{\mathcal{V}} d\tau\,\frac{\partial \vec{\ell}_{mech}}{\partial t} = \frac{d}{dt}\int_{\mathcal{V}} d\tau\,\vec{\ell}_{mech} = \frac{d\vec{L}_{mech}}{dt} = \int_{\mathcal{V}} d\tau\,\vec{n}_{torque} \tag{8.65}$$

$$\Longrightarrow \qquad \boxed{\frac{\partial \vec{\ell}_{mech}(\vec{r})}{\partial t} = \vec{n}_{torque}(\vec{r}) = \vec{\nabla} \cdot \underline{\underline{\mathcal{M}}}(\vec{r}) - \epsilon_o\,\mu_o\,\frac{\partial}{\partial t}\left(\vec{r} \times \vec{S}(\vec{r})\right)} \tag{8.66}$$

We thus obtain a local conservation law that relates *the rate of change of the local angular momentum density to the divergence of the local torque tensor and the rate of change of the rate of the change of the funny quantity containing the Poynting vector*. It can also be viewed as a local torque version of Newton's Second Law.

As before, it is natural to define a *field angular momentum density* and move it to the left side of the above equations:

$$\boxed{\vec{\ell}_{field}(\vec{r}) = \vec{r} \times \vec{g}(\vec{r}) = \epsilon_o \, \mu_o \, \vec{r} \times \vec{S}(\vec{r}) = \epsilon_o \, \vec{r} \times (\vec{E}(\vec{r}) \times \vec{B}(\vec{r})) \qquad \vec{L}_{field} = \int_{\mathcal{V}} d\tau \, \vec{\ell}_{field}} \tag{8.67}$$

With that definition, we obtain

$$\boxed{\frac{d}{dt}\left(\vec{L}_{mech} + \vec{L}_{field}\right) = \oint_{\mathcal{S}(\mathcal{V})} da \, \hat{n} \cdot \underline{\underline{\mathcal{M}}} \quad \Longleftrightarrow \quad \frac{\partial}{\partial t}\left(\vec{\ell}_{mech}(\vec{r}) + \vec{\ell}_{field}(\vec{r})\right) = \vec{\nabla} \cdot \underline{\underline{\mathcal{M}}}(\vec{r})} \tag{8.68}$$

Again, we obtain an integral conservation equation relating *the rate of change of the total angular momentum in a volume to the integral of the torque tensor over the surface* and a local conservation equation relating *the rate of change of the total* angular *momentum density to the divergence of a current density, here now the angular momentum current density* (which has units of angular momentum per unit area per unit time). The second equation is a generalized local "torque" equation.

Note the choice of sign for $\underline{\underline{\mathcal{M}}}$ follows the same convention as for the stress tensor: it gives a continuity equation with a sign flip but is the correct sign for torque. Be aware that this sign convention is the opposite of Jackson's (his Problem 6.9).

Note that field angular momentum is not the same as photon spin or circular polarization; we will come back to this later when we discuss polarization of EM waves.

Section 9
Electromagnetic Waves

Introduction and Study Guidelines

Maxwell's Equations have in them the seeds of self-propagating disturbances in the electromagnetic field: though time-varying charges and currents must generate the waves, they can propagate on their own once initiated. So, in this section, we will develop the theory of such waves propagating in either free space or linear dielectric media, without any free charges. Later on, we will discuss radiation, the process by which time-varying charges and currents generate electromagnetic waves.

We deviate from Griffiths' ordering of topics because you have seen the wave equation three times before, in Ph1c, Ph2/12a, and Ph106a, so we do not need to reintroduce it from scratch. Let's just launch into it and bring the formalism of waves in as we go.

Electromagnetic Waves in Vacuum

From Maxwell's Equations to the Wave Equation

As noted earlier, we will consider Maxwell's Equations in free space with no sources:

$$\vec{\nabla}\cdot\vec{E}=0 \qquad \vec{\nabla}\cdot\vec{B}=0 \qquad \vec{\nabla}\times\vec{E}+\frac{\partial\vec{B}}{\partial t}=0 \qquad \vec{\nabla}\times\vec{B}-\epsilon_o\,\mu_o\,\frac{\partial\vec{E}}{\partial t}=0 \tag{9.1}$$

These equations couple $\vec{E}$ and $\vec{B}$, so let's try to find uncoupled equations by eliminating $\vec{B}$ from the $\vec{\nabla}\times\vec{E}$ equation and $\vec{E}$ from the $\vec{\nabla}\times\vec{B}$ equation by taking the curl again and using one of our standard vector identities:

$$\vec{\nabla}\times\left(\vec{\nabla}\times\vec{E}\right)=\vec{\nabla}\times\left(-\frac{\partial\vec{B}}{\partial t}\right) \qquad \vec{\nabla}\times\left(\vec{\nabla}\times\vec{B}\right)=\epsilon_o\,\mu_o\,\vec{\nabla}\times\left(-\frac{\partial\vec{E}}{\partial t}\right)$$

$$\vec{\nabla}\left(\vec{\nabla}\cdot\vec{E}\right)-\nabla^2\vec{E}=-\epsilon_o\,\mu_o\,\frac{\partial}{\partial t}\,\vec{\nabla}\times\vec{B} \qquad \vec{\nabla}(\vec{\nabla}\cdot\vec{B})-\nabla^2\vec{B}=-\epsilon_o\,\mu_o\,\frac{\partial}{\partial t}\,\vec{\nabla}\times\vec{E}$$

$$\nabla^2\vec{E}=\epsilon_o\,\mu_o\,\frac{\partial^2\vec{E}}{\partial t^2} \qquad \nabla^2\vec{B}=\epsilon_o\,\mu_o\,\frac{\partial^2\vec{B}}{\partial t^2}$$

where $\vec{\nabla}\cdot\vec{E}=0$ because there is no charge density. These are copies of the *wave equation*.

Specifically, these are component-by-component versions of the equation

$$\nabla^2 f(\vec{r},t) = \frac{1}{v^2}\,\frac{\partial^2}{\partial t^2} f(\vec{r},t) \tag{9.2}$$

One can see by substitution that any function of the form

$$f(\vec{r},t) = g(w) \qquad \text{with} \qquad w = \vec{k}\cdot\vec{r} - \omega\, t \tag{9.3}$$

satisfies the wave equation. To prove this, we first need to calculate the derivatives:

$$\nabla^2 f(\vec{r},t) = \sum_{i=1}^{3} \frac{\partial^2 f}{\partial r_i^2} = \sum_{i=1}^{3} \frac{\partial}{\partial r_i}\frac{dg}{dw}\,\frac{\partial w}{\partial r_i} = \sum_{i=1}^{3} \frac{\partial}{\partial r_i}\frac{dg}{dw}\,k_i \tag{9.4}$$

$$= \sum_{i=1}^{3} k_i \frac{d^2 g}{dw^2}\frac{\partial w}{\partial r_i} = \sum_{i=1}^{3} k_i^2 \frac{d^2 g}{dw^2} = |\vec{k}|^2 \frac{d^2 g}{dw^2} \tag{9.5}$$

$$\frac{\partial^2}{\partial t^2} f(\vec{r},t) = \frac{\partial}{\partial t}\frac{dg}{dw}\,\frac{\partial w}{\partial t} = \frac{\partial}{\partial t}\frac{dg}{dw}\,(-\omega) = -\omega\,\frac{d^2 g}{dw^2}\frac{\partial w}{\partial t} = \omega^2\,\frac{d^2 g}{dw^2} \tag{9.6}$$

The wave equation is satisfied by the assumed form if

$$|\vec{k}|^2 = \frac{\omega^2}{v^2} \tag{9.7}$$

The above condition says we can rewrite the argument, eliminating either ω or $|\vec{k}|$, as

$$w = \pm\vec{k}\cdot\vec{r} - |\vec{k}|\,v\,t = \frac{\omega}{v}\left(\pm\widehat{k}\cdot\vec{r} - v\,t\right) \tag{9.8}$$

where we have chosen ω and v to be always nonnegative while $\vec{k}$ is allowed to take on any sign and direction. We can see that surfaces of constant w are given by

$$\delta w = 0 \qquad \Longrightarrow \qquad \omega\left(\pm\widehat{k}\cdot\delta\vec{r} - v\,\delta t\right) = 0 \qquad \Longrightarrow \qquad \pm\widehat{k}\cdot\frac{\delta\vec{r}}{\delta t} = v \tag{9.9}$$

That is, the surfaces of constant w propagate in space along the direction $\pm\widehat{k}$ at speed v. This implies that the "shape function" $g(w)$ propagates at this speed. Returning to our electromagnetic wave equations, we thus see that these waves in the electric and magnetic fields propagate at speed $v = 1/\sqrt{\epsilon_o\,\mu_o}$ which is now, by definition, the *speed of light*, denoted by c.

The interpretation of ω and $k = |\vec{k}|$ are not clear yet, and in fact they are no longer strictly necessary (the factor $\omega/v = k$ could be absorbed into $g(w)$ now that we know ω and k are not independent), but they will become so below when we consider sinusoidal waves.

Lecture 24:

Electromagnetic Waves II:

Electromagnetic Waves in Vacuum (cont.):
Generic Properties,
Time-Averaging, Complex Notation,
Types of Polarization

Electromagnetic Waves in Perfectly Nonconducting, Linear,
Isotropic Matter:
Generic Properties,
Reflection and Refraction

Date Revised: 2024/04/16 07:00
Date Given: 2024/04/16

General Properties of Solutions to the EM Wave Equations

We can use Maxwell's Equations to derive some general properties about electromagnetic waves. Many of these connect to the fact that $\vec{E}$ and $\vec{B}$ are vector quantities. We will begin by assuming the waves are sinusoidal solutions of the most general form allowed so far

$$\vec{E}(\vec{r},t) = \vec{E}_0 \cos\left(\vec{k}_E \cdot \vec{r} - \omega_E\, t + \delta_E\right) \tag{9.10}$$

$$\vec{B}(\vec{r},t) = \vec{B}_0 \cos\left(\vec{k}_B \cdot \vec{r} - \omega_B\, t + \delta_B\right) \tag{9.11}$$

where we have allowed different ω, propagation directions $\widehat{k}$, and phase shifts δ because nothing has restricted that freedom yet. (The sign freedom on $\vec{k}$ has been absorbed into $\widehat{k}$.) We have assumed sinusoidal solutions because they form a complete basis for solution of the wave equation, so any solution can be decomposed in terms of them.

This sinusoidal assumption now enables an interpretation of ω and k. The time dependence at a given point in space has *angular frequency* ω, *frequency* $\nu = \omega/2\,\pi$, and *period* $T = 1/\nu$. The quantity k is the *propagation constant*, and the spatial dependence implies a *wavelength* $\lambda = 2\,\pi/k = v/\nu$. When we consider more general solutions that are the sums of sinusoids, these interpretations fail again because the sum does not correspond to single ω and k values.

With the sinusoidal assumption, we can demonstrate the following:

- *Transversality*
  We can rewrite the divergence equations:

$$0 = \vec{\nabla} \cdot \vec{E} = \frac{d\vec{E}}{dw} \cdot \vec{\nabla} w = -\frac{\omega_E}{c} \, \hat{k}_E \cdot \vec{E}_0 \, \sin\left(\vec{k}_E \cdot \vec{r} - \omega_E \, t + \delta_E\right) \tag{9.12}$$

$$0 = \vec{\nabla} \cdot \vec{B} = \frac{d\vec{B}}{dw} \cdot \vec{\nabla} w = -\frac{\omega_B}{c} \, \hat{k}_B \cdot \vec{B}_0 \, \sin\left(\vec{k}_B \cdot \vec{r} - \omega_B \, t + \delta_B\right) \tag{9.13}$$

  For the above equations to hold at all points in space, it is necessary for $\vec{E}_0$ and $\vec{B}_0$ to be perpendicular to their respective propagation directions. EM waves are thus *transverse* waves: the field disturbance is in the direction perpendicular to propagation.

  It is somewhat remarkable that the same equations that relate fields to charges (and, in particular, imply there are no magnetic point charges) also show that, in the absence of charges, the fields are transverse. This turns out to be a very profound fact. You may recall from Ph106b PS1, when we considered a Coulomb law in which the potential included an exponential decay. In this case, $\vec{\nabla} \cdot \vec{E} \neq 0$ even when charges are not present. The non-vanishing of $\vec{\nabla} \cdot \vec{E}$ would permit a longitudinal component to the wave. In quantum field theory, a non-vanishing transverse component implies the particle corresponding to the field has mass and the static potential for the field is short range with decay length $\lambda \propto 1/M$.

- *Orthogonality and Equality of $\vec{k}$, ω, and δ*
  Let's write the curl equations. First, we take the necessary derivatives:

$$\begin{aligned}\vec{\nabla}\times\vec{E} &= \sum_{i,j,k=1}^{3}\epsilon_{ijk}\,\widehat{r}_i\,\frac{\partial E_k}{\partial r_j} = \sum_{i,j,k=1}^{3}\epsilon_{ijk}\,\widehat{r}_i\,\frac{dE_k}{dw}\,\frac{\partial w}{\partial r_j}\\ &= \sum_{i,j,k=1}^{3}\epsilon_{ijk}\,\widehat{r}_i\,\frac{dE_k}{dw}\,k_{E,j} = \vec{k}_E\times\frac{d\vec{E}}{dw} = \frac{\omega_E}{c}\,\widehat{k}_E\times\frac{d\vec{E}}{dw}\\ &= -\frac{\omega_E}{c}\,\widehat{k}_E\times\vec{E}_0\sin\left(\vec{k}_E\cdot\vec{r}-\omega_E\,t+\delta_E\right) \qquad (9.14)\end{aligned}$$

$$\frac{\partial\vec{E}}{\partial t} = \frac{d\vec{E}}{dw}\,\frac{\partial w}{\partial t} = -\omega_E\,\frac{d\vec{E}}{dw} = \omega_E\,\vec{E}_0\sin\left(\vec{k}_E\cdot\vec{r}-\omega_E\,t+\delta_E\right) \qquad (9.15)$$

$\vec{B}$ has similar derivatives. Plugging the above into Faraday's Law and Ampere's Law:

$$-\frac{\omega_E}{c}\,\widehat{k}_E\times\vec{E}_0\sin\left(\vec{k}_E\cdot\vec{r}-\omega_E\,t+\delta_E\right) = -\omega_B\,\vec{B}_0\,\sin\left(\vec{k}_B\cdot\vec{r}-\omega_B\,t+\delta_B\right) \qquad (9.16)$$

$$\frac{\omega_B}{c}\,\widehat{k}_B\times\vec{B}_0\sin\left(\vec{k}_B\cdot\vec{r}-\omega_B\,t+\delta_B\right) = -\frac{\omega_E}{c^2}\,\vec{E}_0\sin\left(\vec{k}_E\cdot\vec{r}-\omega_E\,t+\delta_E\right) \qquad (9.17)$$

We can conclude two things from these equations:

- In order for the equality to hold at all points in space and at all times, the arguments of the sin functions on the two sides must be the same, $\vec{k}_E = \vec{k}_B$, $\omega_E = \omega_B$, and $\delta_E = \delta_B$
- $\vec{k}$, $\vec{E}_0$, and $\vec{B}_0$ form a mutually orthogonal set of vectors

In the end, we therefore have the following relation between $\vec{E}$ and $\vec{B}$:

$$\vec{E}(\vec{r},t) = \vec{E}_0 \cos\left(\vec{k}\cdot\vec{r} - \omega\, t + \delta\right) \tag{9.18}$$

$$\vec{B}(\vec{r},t) = \vec{B}_0 \cos\left(\vec{k}\cdot\vec{r} - \omega\, t + \delta\right) \tag{9.19}$$

$$\vec{B}_0 = \frac{1}{c}\,\widehat{k}\times\vec{E}_0 \quad\Longleftrightarrow\quad \vec{E}_0 = -c\,\widehat{k}\times\vec{B}_0 \tag{9.20}$$

One note on the behavior of independent polarizations. Given a propagation direction $\widehat{k}$, we can pick two directions, $\widehat{n}_1$ and $\widehat{n}_2$, to form a basis for $\vec{E}$. A natural choice is to require $\widehat{n}_1 \times \widehat{n}_2 = \widehat{k}$, which also implies $\widehat{k} \times \widehat{n}_1 = \widehat{n}_2$. Then $\vec{E}_0 = E_1 \widehat{n}_1 + E_2 \widehat{n}_2$ and $\vec{B}_0 = B_1 \widehat{n}_1 + B_2 \widehat{n}_2$. E_1 and E_2 are the two possible *polarizations* of the electric field. (Of course, we can pick any $\widehat{n}_1$ we want; once $\widehat{n}_1$ has been picked, then the polarization directions are set.) The curl of $\vec{E}$ relation then implies

$$B_2 = \vec{B}_0 \cdot \widehat{n}_2 = \frac{1}{c} \vec{E}_0 \cdot \widehat{n}_1 = \frac{E_1}{c} \qquad B_1 = \vec{B}_0 \cdot \widehat{n}_1 = -\frac{1}{c} \vec{E}_0 \cdot \widehat{n}_2 = -\frac{E_2}{c} \tag{9.21}$$

We thus see that, aside from picking consistent $\widehat{k}$, $\widehat{n}_1$, and $\widehat{n}_2$, there is no connection between the (E_1, B_2) pair and the (E_2, B_1) pair. The waves in the two complementary polarizations can have different ω and thus different k. They are two completely independent waves. There is no fixed relationship between the waves in the two polarizations, and they can get out of phase with each other as they propagate if they have different ω.

If we consider two waves that have the same ω, then the two waves propagate together — their relative phase does not change with time or position. But there remains no condition connecting E_1 and E_2, or B_1 and B_2, so there is no requirement that the complementary polarizations have matching amplitude or phase (δ). We will see later that this independence of the two polarization amplitudes and phases can be used to generate a diverse set of possible polarizations: linear, circular, and elliptical.

Rewriting EM Waves using the Auxiliary Field

It is interesting to note at this point that $\vec{B}$ is not the most natural field quantity to work with: it is smaller than $\vec{E}$ by a factor c, which is large. If we instead use $\vec{H} = \vec{B}/\mu_o$, then we obtain

$$\vec{H}_0 = \frac{1}{Z_0}\,\widehat{k} \times \vec{E}_0 \qquad Z_0 = \sqrt{\frac{\mu_o}{\epsilon_o}} \tag{9.22}$$

The quantity $Z_0 \approx 377\ \Omega$ is known as the *impedance of free space* and has units of resistance (impedance). We see that $\vec{H}$ is now only a factor of 377 smaller than $\vec{E}$. We also recall that $\vec{H}$ has units of surface current density. This foreshadows the way $\vec{H}$ will be related to the surface currents that the electric field drives in polarizable/magnetizable media and in conductors.

Energy and Momentum in Electromagnetic Waves

The energy density in an electromagnetic wave, now using $|\vec{B}| = |\vec{E}|/c = |\vec{E}|\sqrt{\epsilon_o\,\mu_o}$, is

$$u = \frac{1}{2}\left(\epsilon_o\,E^2 + \frac{1}{\mu_o}\,B^2\right) = \frac{1}{2}\left(\epsilon_o\,E^2 + \frac{\epsilon_o\,\mu_o}{\mu_o}\,E^2\right) = \epsilon_o\,E^2 \tag{9.23}$$

The energy flux per unit area is the Poynting vector:

$$\vec{S} = \frac{1}{\mu_o}\,\vec{E}\times\vec{B} = \frac{E^2}{c\,\mu_o}\,\widehat{k} = c\,\epsilon_o\,E^2\,\widehat{k} = c\,u\,\widehat{k} \tag{9.24}$$

Thus, we see that the energy transported by the electromagnetic wave travels at the speed of light, just as the wave does. The momentum density vector is

$$\vec{g} = \epsilon_o\,\mu_o\,\vec{S} = \frac{u}{c}\,\widehat{k} = \frac{\epsilon_o\,E^2}{c}\,\widehat{k} \tag{9.25}$$

Note that the energy flux and momentum density differ by a factor of c^2, not just c (as we would expect for a relation between energy and momentum) because one is a *flux* (energy/area/time) and the other is a *density* (momentum/volume); the difference in spatial and temporal units introduces another factor of velocity between the two.

We can also write down the stress tensor. First, consider the special case $\widehat{k} = k\,\widehat{z}$ so $\widehat{n}_1 = \widehat{x}$ and $\widehat{n}_2 = \widehat{y}$. Consider a wave polarized along $\widehat{x}$. Then the fields are

$$\vec{E} = E\,\widehat{x} \qquad \vec{B} = B\,\widehat{y} = \frac{E}{c}\,\widehat{y} = \sqrt{\mu_o\,\epsilon_o}\,E\,\widehat{y} \tag{9.26}$$

The stress tensor is

$$T_{11} = \epsilon_o\left(E^2 - \frac{1}{2}\,E^2\right) + \frac{1}{\mu_o}\left(-\frac{1}{2}\,B^2\right) = 0 \tag{9.27}$$

$$T_{22} = \epsilon_o\left(-\frac{1}{2}\,E^2\right) + \frac{1}{\mu_o}\left(B^2 - \frac{1}{2}\,B^2\right) = 0 \tag{9.28}$$

$$T_{33} = \epsilon_o\left(-\frac{1}{2}\,E^2\right) + \frac{1}{\mu_o}\left(-\frac{1}{2}\,B^2\right) = -u \tag{9.29}$$

$$T_{12} = T_{13} = T_{23} = 0 \tag{9.30}$$

$$\Longrightarrow \qquad \underline{\underline{\mathcal{T}}} = -u\,\widehat{z}\,\widehat{z} \tag{9.31}$$

The stress tensor for the complementary polarization is the same.

It is reasonable to extrapolate from the above that the generic stress tensor (now making the time dependence explicit) is

$$\underline{\underline{\mathcal{T}}} = -\,\widehat{k}\,\widehat{k}\,\epsilon_o\,E_0^2\,\cos^2\left(\vec{k}\cdot\vec{r}-\omega\,t+\delta\right) \tag{9.32}$$

One explanation of the reason for the negative sign is that $\underline{\underline{\mathcal{T}}}$ is the *negative* of the momentum current density.

From the stress tensor, we can calculate the *radiation pressure*, the force per unit area that would be applied to an object that absorbs the electromagnetic wave. Recall that $\widehat{n}_1\cdot\underline{\underline{\mathcal{T}}}\cdot\widehat{n}_2$ gives the force acting in the $\widehat{n}_1$ direction on a surface element whose normal is in the $\widehat{n}_2$ direction. Since $\underline{\underline{\mathcal{T}}}\propto-\widehat{k}\,\widehat{k}$, the force is only nonzero (and positive) in the $\widehat{k}$ direction on an area element whose *outward* normal is in the $-\widehat{k}$ direction. (Recall how, in our example of using the stress tensor to calculate the force between the two spinning charged hemispheres, the surface normal was in the $-\widehat{z}$ direction for calculating the force on the hemisphere in the upper half-space.) The radiation pressure in the $\widehat{k}$ direction is then

$$\textit{Pressure} = \widehat{k}\cdot\underline{\underline{\mathcal{T}}}\cdot-\widehat{k} = E_0^2\,\cos^2\left(\vec{k}\cdot\vec{r}-\omega\,t+\delta\right) = u \tag{9.33}$$

We will see later that, if the wave is not absorbed but reflected, the wave maintains its amplitude $|\vec{E}|$ but its $\widehat{k}$ reverses sign, implying that the momentum transfer and thus the pressure are increased by a factor of 2.

It is not particularly useful to write down the angular momentum density and the angular momentum tensor for a plane wave. They are position dependent, reflecting the fact that, if a charged particle absorbs energy and momentum from a plane wave at some nonzero distance from the origin, it acquires a linear momentum $\vec{p}$ and thus an angular momentum $\vec{r} \times \vec{p}$. The latter carries no information beyond that of the former. Only if the wave has a nontrivial dependence of $\vec{E}$ and $\vec{B}$ on position — for example, $\vec{E} \times \vec{B} \propto \widehat{\phi}$ — is the angular momentum of the wave interesting. Such waves are beyond the scope of our current discussion of plane waves.

Time Averaging for Plane Waves

For sinusoidally oscillating plane waves, it is standard to take time averages of quantities. Obviously, the fields themselves time-average to zero. But energy and momentum do not:

$$\langle u(\vec{r}_0)\rangle = \left\langle \epsilon_o\, E_0^2\, \cos^2\left(\vec{k}\cdot\vec{r}_0 - \omega\, t + \delta\right)\right\rangle = \frac{1}{2}\,\epsilon_o\, E_0^2 \tag{9.34}$$

$$\left\langle \vec{S}(\vec{r}_0)\right\rangle = \frac{1}{2}\, c\,\epsilon_o\, E_0^2\,\widehat{k} \qquad \langle \vec{g}(\vec{r}_0)\rangle = \frac{1}{2}\,\frac{\epsilon_o\, E_0^2}{c}\,\widehat{k} \qquad \left\langle \underline{\underline{\mathcal{T}}}(\vec{r}_0)\right\rangle = -\frac{1}{2}\,\epsilon_o\, E_0^2\,\widehat{k}\,\widehat{k} \tag{9.35}$$

The average power per unit area transported by the wave is the *intensity*

$$\mathrm{I} = \left\langle |\vec{S}|\right\rangle = \frac{1}{2}\, c\,\epsilon_o\, E_0^2 \tag{9.36}$$

Note that the magnitude refers to the *vector magnitude*. Starting with the next slide, when we consider complex notation for fields, we will always define $\vec{S}$ in such a way as to be real, so the magnitude will continue to refer to vector magnitude even for complex fields.

Complex Notation for Plane Waves

For the sake of convenience in manipulation, we will from now on use *complex notation* for plane waves,

$$\vec{E}(\vec{r},t) = \widehat{n}\,\mathcal{R}\left[\,\widetilde{E}_0\, e^{i(\vec{k}\cdot\vec{r}-\omega\, t)}\right] \qquad \vec{B}(\vec{r},t) = \frac{\widehat{k}\times\widehat{n}}{c}\,\mathcal{R}\left[\,\widetilde{E}_0\, e^{i(\vec{k}\cdot\vec{r}-\omega\, t)}\right] \tag{9.37}$$

$$\widetilde{E}_0 = E_0\, e^{i\delta} \tag{9.38}$$

where $\widetilde{E}_0$ is now a complex number into which we have absorbed the phase factor $e^{i\delta}$ and $\mathcal{R}$ means "take the real part." We will not carry along tildes on the vectors $\vec{E}$ and $\vec{E}_0$ because it would be too cumbersome for the notation. It will be clear from context whether we mean the complex or real fields.

To calculate quadratic quantities like u, $\vec{S}$, $\vec{g}$, and $\underline{\underline{\mathcal{T}}}$ *with full space- and time-dependence* requires that one first take the real part and then apply the previously provided formulae for these quadratic quantities.

Fortunately, if one is only interested in time averages, there is a simple extension to the prescription for time averages involving complex conjugation and multiplication by a factor of 1/2:

$$\langle u(\vec{r}_0)\rangle = \frac{1}{4}\left(\epsilon_o \vec{E}^* \cdot \vec{E} + \frac{1}{\mu_o}\vec{B}^* \cdot \vec{B}\right) \overset{pw\ vac}{=} \frac{1}{2}\epsilon_o E_0^2 \tag{9.39}$$

$$\left\langle \vec{S}(\vec{r}_0)\right\rangle = c^2 \langle \vec{g}(\vec{r}_0)\rangle = \frac{1}{2\mu_o}\mathcal{R}\left(\vec{E}^* \times \vec{B}\right) \overset{pw\ vac}{=} \frac{1}{2} c\, \epsilon_o E_0^2\, \widehat{k} \tag{9.40}$$

$$\left\langle \left(\underline{\underline{\mathcal{T}}}\right)_{ij}(\vec{r}_0)\right\rangle = \frac{1}{2}\left[\epsilon_o\left(\mathcal{R}\left[E_i^* E_j\right] - \frac{1}{2}\delta_{ij}\,\vec{E}^* \cdot \vec{E}\right) + \frac{1}{\mu_o}\left(\mathcal{R}\left[B_i^* B_j\right] - \frac{1}{2}\delta_{ij}\,\vec{B}^* \cdot \vec{B}\right)\right] \tag{9.41}$$

$$= -\frac{1}{2}\epsilon_o E_0^2\, \widehat{k} \cdot \widehat{r}_i\, \widehat{k} \cdot \widehat{r}_j \tag{9.42}$$

where the first expression for each quantity is always valid (even later for conductors when $\vec{E}$ and $\vec{B}$ can be out of phase) while the final evaluation is only valid for the plane waves in vacuum we have been considering (indicated by the $\overset{pw\ vac}{=}$ notation). The factors of 1/2 in front come from time-averaging. Note that there may be spatial dependence remaining in the result if the wave amplitude has a dependence on position outside of the sinusoidal wave-propagation factor (not possible for plane EM waves, but it will happen for radiation). It is not necessary to take the real part for u and certain pieces of $\underline{\underline{\mathcal{T}}}$ because they are manifestly real.

Types of Polarization

So far, we have only discussed linear polarization, wherein the direction of $\vec{E}$ at a particular point in space stays constant over time (aside from sign flips). However, by combining two orthogonal linear polarizations with appropriate complex coefficients, one can obtain more "complex" behavior.

The simplest extension is to consider what happens when you add two orthogonal polarizations of the same amplitude but with a possible phase shift:

$$\vec{E}(\vec{r},t) = \frac{\widetilde{E}_0}{\sqrt{2}}\left(\widehat{n}_1 + \widehat{n}_2\, e^{i\delta}\right) e^{i\left(\vec{k}\cdot\vec{r}-\omega\, t\right)} \tag{9.43}$$

If $\delta = 0$ or $\delta = \pi$, then one just obtains a linear polarization in the direction $\left(\widehat{n}_1 \pm \widehat{n}_2\right)/\sqrt{2}$.

But, if $\delta = \pm\pi/2$, then the wave polarized along $\widehat{n}_2$ is $\pi/2$ out of phase with the wave polarized along $\widehat{n}_1$: when the $\widehat{n}_1$ mode has zero amplitude, the $\widehat{n}_2$ mode has maximum amplitude and vice versa. If we take the real part, it is clear what is going on:

$$\mathcal{R}\left[\vec{E}(\vec{r},t)\right] = \frac{E_0}{\sqrt{2}}\left[\widehat{n}_1\,\cos\left(\vec{k}\cdot\vec{r}-\omega\,t+\delta_0\right)+\widehat{n}_2\,\cos\left(\vec{k}\cdot\vec{r}-\omega\,t+\delta_0\pm\frac{\pi}{2}\right)\right] \tag{9.44}$$

$$= \frac{E_0}{\sqrt{2}}\left[\widehat{n}_1\,\cos\left(\vec{k}\cdot\vec{r}-\omega\,t+\delta_0\right)\mp\widehat{n}_2\,\sin\left(\vec{k}\cdot\vec{r}-\omega\,t+\delta_0\right)\right] \tag{9.45}$$

The polarization vector maintains an amplitude $E_0/\sqrt{2}$ but it sweeps around in a circle with period $T = 2\,\pi/\omega$: this is *circular polarization*. To understand which direction the polarization vector rotates, let's look *into* the wave (toward $-\widehat{k}$) while sitting at a fixed point in space (fixed $\vec{r}$). The time-varying component of the arguments of the sinusoids is $-\omega\,t$, and thus, as time evolves positively, the arguments of the sinusoids evolve negatively. For $\delta = +\pi/2$, the sign on the second term is negative, the rotation is counterclockwise, and the wave is called *left circularly polarized*. Conversely, $\delta = -\pi/2$ yields clockwise rotation and is called *right circularly polarized*. One also speaks in terms of helicity, in which case one considers the rotation of the polarization *relative to the direction of motion* using the right-hand rule. The left circularly polarized wave has *positive helicity* because the polarization vector rotates around $+\widehat{k}$ according to the right-hand rule (thumb along $+\widehat{k}$). The right circularly polarized wave has *negative helicity* because it obeys the left-hand rule.

The next possibility is to allow unequal coefficients:

$$\vec{E}(\vec{r},t) = \widetilde{E}_0 \left(\alpha\,\widehat{n}_1 + \beta\,\widehat{n}_2\,e^{i\delta}\right) e^{i\left(\vec{k}\cdot\vec{r}-\omega\,t\right)} \qquad \alpha^2+\beta^2 = 1 \qquad (9.46)$$

When $\delta = 0$ or $\delta = \pi$, we again obtain linear polarization, but now making an angle $\theta = \tan^{-1}(\pm\beta/\alpha)$ with the $\widehat{n}_1$ axis (sign is the same as sign of $e^{i\delta} = \pm 1$).

If we now consider $\delta = \pm\pi/2$ and $\alpha \neq \beta$, we obtain an *elliptically polarized* wave: at a fixed point, the polarization vector sweeps out an ellipse whose semimajor and semiminor axes are along $\widehat{n}_1$ and $\widehat{n}_2$. If δ is an arbitrary value, then the semimajor and semiminor axes are rotated from the $\widehat{n}_1$–$\widehat{n}_2$ system by an angle related to δ. We also obtain an elliptically polarized wave if we consider $\delta \neq 0$, $\pm\pi/2$, or π but $\alpha = \beta$,

It turns out that elliptically polarized waves are easier to analyze if they are rewritten in terms of the two helicities (or circular polarizations). That is, if we take as our polarization basis and field decomposition

$$\widehat{n}_\pm = \frac{1}{\sqrt{2}}\left(\widehat{n}_1 \pm e^{i\,\pi/2}\,\widehat{n}_2\right) \qquad \vec{E}(\vec{r},t) = \left(\widetilde{E}_+\,\widehat{n}_+ + \widetilde{E}_-\,\widehat{n}_-\right) e^{i\left(\vec{k}\cdot\vec{r}-\omega\,t\right)} \qquad (9.47)$$

then the parameters of the ellipse traced out by the polarization vector are:

$$r\,e^{i\,\theta} = \frac{\widetilde{E}_-}{\widetilde{E}_+} \qquad \frac{\text{semiminor axis}}{\text{semimajor axis}} = \left|\frac{1-r}{1+r}\right| \qquad \text{angle wrt } \widehat{n}_1 = \frac{\theta}{2} \qquad (9.48)$$

where the angle is measured *looking into the wave* (*i.e.*, looking in the $-\widehat{k}$ direction).

It is interesting to note that a circularly polarized plane wave is, in terms of angular momentum, no different from a linearly polarized plane wave according to the definition of the angular momentum density, Equation 8.67:

$$\vec{\ell}_{field} = \epsilon_o \, \vec{r} \times \left(\vec{E} \times \vec{B} \right) = \frac{\epsilon_o \, E_0^2}{c} \, \vec{r} \times \widehat{k} \tag{9.49}$$

One can see that the angular momentum has to do with the relative orientation of the propagation direction and the position vector, not with the nature of the polarization. This reflects the fact that, in quantum mechanics, the helicity of the wave becomes the intrinsic *spin angular momentum* of the photon, while the quantity calculated above is the *orbital angular momentum* of the photon and has to do with the *spatial distribution* of the EM wave, in much the same way that orbital angular momentum in quantum mechanics is determined by the spatial distribution of the wavefunction and is unassociated with the particle's spin.

Electromagnetic Waves in Perfectly Nonconducting, Linear, Isotropic Matter

Propagation in Linear, Isotropic Media

Maxwell's Equations in matter in the absence of free charges and currents are

$$\vec{\nabla}\cdot\vec{D}=0 \qquad \vec{\nabla}\cdot\vec{B}=0 \qquad \vec{\nabla}\times\vec{E}+\frac{\partial\vec{B}}{\partial t}=0 \qquad \vec{\nabla}\times\vec{H}-\frac{\partial\vec{D}}{\partial t}=0 \qquad (9.50)$$

As noted earlier, we need relations between $\vec{D}$ and $\vec{E}$ and between $\vec{H}$ and $\vec{B}$ to make use of these. If we assume linear, isotropic media (ϵ and μ scalars, not tensors)

$$\vec{D}=\epsilon\,\vec{E} \qquad \vec{B}=\mu\,\vec{H} \qquad (9.51)$$

then the equations reduce to

$$\vec{\nabla}\cdot\vec{E}=0 \qquad \vec{\nabla}\cdot\vec{B}=0 \qquad \vec{\nabla}\times\vec{E}+\frac{\partial\vec{B}}{\partial t}=0 \qquad \vec{\nabla}\times\vec{B}-\epsilon\,\mu\,\frac{\partial\vec{E}}{\partial t}=0 \qquad (9.52)$$

These are the same as our equations in vacuum, leading to the same kinds of waves, but with the modification

$$v=\frac{1}{\sqrt{\epsilon\,\mu}}=\frac{c}{n} \qquad \text{with} \qquad n=\sqrt{\frac{\epsilon\,\mu}{\epsilon_o\,\mu_o}}=\textit{index of refraction} \qquad (9.53)$$

This mathematical transformation reflects a remarkable fact: the complicated polarization and magnetization of the medium occurring as the wave passes through it do nothing except change its speed and, we shall see, affect the wave amplitude. This is a consequence of the linearity of the medium we assume.

Most materials in which waves can propagate (as we will see, this means materials that do not have high conductivities) have $\mu \approx \mu_o$, so $n \approx \sqrt{\epsilon/\epsilon_o} = \sqrt{\epsilon_r}$. Since $\epsilon_r > 1$ in general (there are very few paraelectric materials that enhance the field rather than act to decrease it), light generally goes more slowly in dielectrics. Though, creation of metamaterials in which the effective index of refraction is less than unity (over a limited frequency range via use of resonant structures) is an area of active research!

The relation between $\vec{B}$ and $\vec{E}$, Equation 9.20, is modified in the obvious manner:

$$\vec{B}(\vec{r},t) = \frac{1}{v}\,\widehat{k} \times \vec{E}(\vec{r},t) \tag{9.54}$$

As we did for free space, we also have that $\vec{H}$ and $\vec{E}$ are related by an impedance, $Z = \sqrt{\mu/\epsilon}$, which we now call the *wave impedance*. With it, we have

$$\vec{H}(\vec{r},t) = \frac{\vec{B}(\vec{r},t)}{\mu} = \frac{1}{Z}\,\widehat{k} \times \vec{E}(\vec{r},t) \tag{9.55}$$

Recall that $\vec{H}$ carries units of surface current density.

Energy density, the Poynting vector, the momentum vector, and the stress tensor take on unsurprising forms given the above modification:

$$u = \frac{1}{2}\left(\epsilon E^2 + \frac{1}{\mu} B^2\right) = \epsilon E^2 \qquad \vec{S} = \frac{1}{\mu}\vec{E}^* \times \vec{B} = v\,\epsilon E^2\,\widehat{k} = v\,u\,\widehat{k} \tag{9.56}$$

$$\vec{g} = \epsilon\,\mu\,\vec{S} = \frac{\epsilon E^2}{v}\,\widehat{k} = \frac{u}{v}\,\widehat{k} \qquad \underline{\underline{\mathcal{T}}} = -\epsilon E^2\,\widehat{k}\,\widehat{k} = -u\,\widehat{k}\,\widehat{k} \tag{9.57}$$

The time averages for a sinusoidal wave of (real) amplitude E_0 are (now including intensity):

$$\langle u \rangle = \frac{1}{2}\,\epsilon E_0^2 \qquad \left\langle \vec{S} \right\rangle = \frac{1}{2\,\mu}\left\langle \vec{E}^* \times \vec{B} \right\rangle = \frac{1}{2}\,v\,\epsilon E_0^2\,\widehat{k} = v\,\langle u \rangle\,\widehat{k} \tag{9.58}$$

$$\mathrm{I} = \left\langle |\vec{S}| \right\rangle = \frac{1}{2}\,v\,\epsilon E_0^2 = v\,\langle u \rangle \tag{9.59}$$

$$\langle \vec{g} \rangle = \epsilon\,\mu \left\langle \vec{S} \right\rangle = \frac{1}{2}\,\frac{\epsilon E_0^2}{v}\,\widehat{k} = \frac{\langle u \rangle}{v}\,\widehat{k} \qquad \left\langle \underline{\underline{\mathcal{T}}} \right\rangle = -\frac{1}{2}\,\epsilon E_0^2\,\widehat{k}\,\widehat{k} = -\langle u \rangle\,\widehat{k}\,\widehat{k} \tag{9.60}$$

Since $v > c$ is possible, we cannot so easily interpret the above equations as implying that energy propagates at speed v. You know from Ph2/12 that energy in a wave propagates with the group velocity $v_g = d\omega/dk$ and so it becomes important to know $v(\omega)$. We will build an approximate physical model for $v(\omega)$ in §**??**.

Boundary Conditions

Recalling our boundary conditions for linear media (Equations 2.61, 4.23, **??**, **??**) and applying our assumption of no free currents, we have ($\widehat{n}$ = normal from 1 to 2, $\widehat{s}$ = any tangential vector at interface):

$$\widehat{n}\cdot\epsilon_1\,\vec{E}_1=\widehat{n}\cdot\epsilon_2\,\vec{E}_2 \qquad \widehat{n}\cdot\vec{B}_1=\widehat{n}\cdot\vec{B}_2 \qquad \widehat{s}\cdot\vec{E}_1=\widehat{s}\cdot\vec{E}_2 \qquad \widehat{s}\cdot\frac{\vec{B}_1}{\mu_1}=\widehat{s}\cdot\frac{\vec{B}_2}{\mu_2} \quad (9.61)$$

We will apply these to calculate the reflection and refraction of EM waves at the interface between different linear media. We will write the magnetic field boundary condition in terms of $\vec{H}$ because it makes them look like the electric field boundary conditions, which will be convenient during our discussion of reflection and refraction:

$$\widehat{n}\cdot\mu_1\,\vec{H}_1=\widehat{n}\cdot\mu_2\,\vec{H}_2 \qquad \widehat{s}\cdot\vec{H}_1=\widehat{s}\cdot\vec{H}_2 \quad (9.62)$$

Reflection and Refraction: General Considerations

We will skip past the case of normal incidence, which you studied in Ph1c, but we will consider it as a special case of our generic results.

Assume we have a wave with propagation vector $\vec{k}_i$ propagating in medium 1 and incident on an interface with medium 2, with normal $\widehat{n}$ pointing from 1 into 2. We expect there to be reflected and transmitted waves. We write these all as

$$\vec{E}_i(\vec{r},t) = \vec{E}_{0,i}\, e^{i\left(\vec{k}_i\cdot\vec{r}-\omega t\right)} \qquad \vec{H}_i(\vec{r},t) = \frac{1}{Z_1}\,\widehat{k}_i \times \vec{E}_i \tag{9.63}$$

$$\vec{E}_r(\vec{r},t) = \vec{E}_{0,r}\, e^{i\left(\vec{k}_r\cdot\vec{r}-\omega t\right)} \qquad \vec{H}_r(\vec{r},t) = \frac{1}{Z_1}\,\widehat{k}_r \times \vec{E}_r \tag{9.64}$$

$$\vec{E}_t(\vec{r},t) = \vec{E}_{0,t}\, e^{i\left(\vec{k}_t\cdot\vec{r}-\omega t\right)} \qquad \vec{H}_t(\vec{r},t) = \frac{1}{Z_2}\,\widehat{k}_t \times \vec{E}_t \tag{9.65}$$

We use $\vec{H}$ instead of $\vec{B}$ because the boundary conditions are more easily written in terms of $\vec{H}$. We have already applied the condition that the frequencies of the three waves are identical. This is necessary for any boundary conditions connecting them to be applicable at all time. Then

$$k_i\, v_1 = k_r\, v_1 = k_t\, v_2 = \omega \qquad \Longrightarrow \qquad k_r = k_i \qquad k_t = \frac{v_1}{v_2}\, k_i = \frac{n_2}{n_1}\, k_i \tag{9.66}$$

Since $k = 2\,\pi/\lambda$, this implies the wavelength differs by n_2/n_1 in the two media!

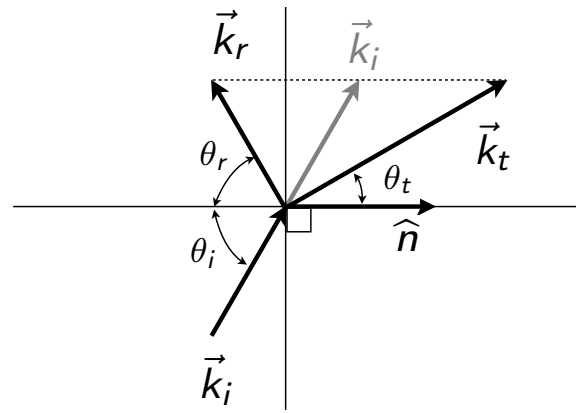
Now, consider the kinds of matching conditions we will write down. They must always hold over all $\vec{r}$ in the interface. In particular, if we take the interface to be the $z = 0$ plane, with $\widehat{n} = \widehat{z}$, then the matching conditions that must hold at all x and y are of the form

$$(\ )\, e^{i\left(\vec{k}_i \cdot (x\widehat{x}+y\widehat{y})-\omega t\right)} + (\ )\, e^{i\left(\vec{k}_r \cdot (x\widehat{x}+y\widehat{y})-\omega t\right)} = (\ )\, e^{i\left(\vec{k}_t \cdot (x\widehat{x}+y\widehat{y})-\omega t\right)} \tag{9.67}$$

In order for these to hold at arbitrary x and y, it must be that

$$\widehat{x} \cdot \vec{k}_i = \widehat{x} \cdot \vec{k}_r = \widehat{x} \cdot \vec{k}_t \tag{9.68}$$

$$\widehat{y} \cdot \vec{k}_i = \widehat{y} \cdot \vec{k}_r = \widehat{y} \cdot \vec{k}_t \tag{9.69}$$



These conditions imply that if you project all three propagation vectors into the plane of the interface (whose normal is $\widehat{n}$), then their projections in that plane are equal. Furthermore, there is a plane formed by this common xy projection of the propagation vectors and the normal $\widehat{n}$ (which is normal to the projection plane), and all three vectors lie in this plane.

Each propagation vector makes an angle with the interface normal, $\widehat{n}$. We label them θ_i, θ_r, and θ_t. The projection of a given wavevector parallel to the normal is therefore $|\vec{k}| \cos\theta$ while the projection perpendicular to the normal (in the plane of the interface) is $|\vec{k}| \sin\theta$. Since we have argued that these projections into the plane of the interface are equal, we thus have

$$k_i \sin\theta_i = k_r \sin\theta_r = k_t \sin\theta_t \tag{9.70}$$

Now, using our relations $k_r = k_i$ and $k_t = \frac{n_2}{n_1} k_i$, we may conclude:

$$\textit{law of reflection}: \qquad \theta_i = \theta_r \tag{9.71}$$

$$\textit{law of refraction (Snell's Law)}: \qquad n_1 \sin\theta_i = n_2 \sin\theta_t \tag{9.72}$$

Snell's Law tells us that the light ray bends toward the normal if $n_2 > n_1$, and it bends away from the normal if $n_2 < n_1$.

Total internal reflection occurs when $n_2 < n_1$ and $\sin\theta_i > n_2/n_1$: in this case, $\sin\theta_t > 1$ and there is no solution for k_t. This happens because, when $n_2 < n_1$, then the magnitude $k_t < k_i$, so then the projection $k_i \sin\theta_i$ can be (but doesn't have to be, depending on the value of $\sin\theta_i$) too large for k_t to match. We will study this case in more detail in a homework problem.

Note that none of these results depending on knowing anything about Maxwell's Equations or boundary conditions: these are generic properties of any waves, which is why they hold for sound waves, phonons, etc. They result only from matching time and space dependences at the boundary.

Lecture 25:

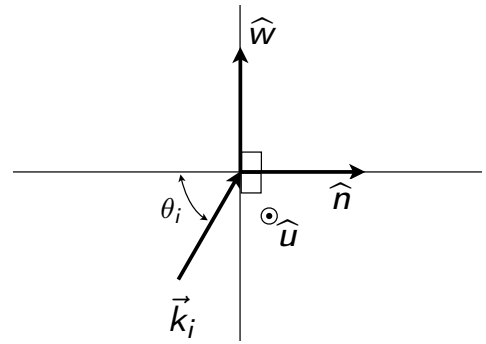
Electromagnetic Waves III:
Electromagnetic Waves in Perfectly Nonconducting, Linear, Isotropic Matter (cont.):
Reflection and Refraction (cont.)

Date Revised: 2024/04/18 06:30
Date Given: 2024/04/18

Reflected and Transmitted Field Relations

Now, let's apply our electromagnetic boundary conditions to figure out how the amplitudes and energies are related. We may drop all the exponential factors because we have established that they are identical at the interface. We need to define two new unit vectors:

$$\widehat{u} \propto \widehat{n} \times \widehat{k}_i = \widehat{n} \times \widehat{k}_r = \widehat{n} \times \widehat{k}_t$$
$$\widehat{w} = \widehat{u} \times \widehat{n} \qquad (9.73)$$



It doesn't matter which $\widehat{k}$ we use to define $\widehat{u}$ because we argued earlier that all of them lie in the same plane with $\widehat{n}$, so the direction normal to the plane containing them and $\widehat{n}$ is independent of which one is used. (The equality of the three cross-products follows from Equations 9.68 and 9.69.) Clearly, then, $\widehat{u}$ is perpendicular to this plane in which $\widehat{n}$ and the propagation vectors lie. $\widehat{w}$ is then the obvious third direction, and it and $\widehat{n}$ define the plane that the $\widehat{k}$ live in.

With these definitions, our boundary conditions can be written as

$$\widehat{n}\cdot\epsilon_1\left(\vec{E}_{0,i}+\vec{E}_{0,r}\right)=\widehat{n}\cdot\epsilon_2\,\vec{E}_{0,t} \qquad \begin{bmatrix}\widehat{u}\\ \widehat{w}\end{bmatrix}\cdot\left(\vec{E}_{0,i}+\vec{E}_{0,r}\right)=\begin{bmatrix}\widehat{u}\\ \widehat{w}\end{bmatrix}\cdot\vec{E}_{0,t} \qquad (9.74)$$

$$\widehat{n}\cdot\mu_1\left(\vec{H}_{0,i}+\vec{H}_{0,r}\right)=\widehat{n}\cdot\mu_2\vec{H}_{0,t} \qquad \begin{bmatrix}\widehat{u}\\ \widehat{w}\end{bmatrix}\cdot\left(\vec{H}_{0,i}+\vec{H}_{0,r}\right)=\begin{bmatrix}\widehat{u}\\ \widehat{w}\end{bmatrix}\cdot\vec{H}_{0,t} \qquad (9.75)$$

where the stacking of $\widehat{u}$ and $\widehat{w}$ is just meant to indicate that those equations apply with either $\widehat{u}$ on both sides or $\widehat{w}$ on both sides.

Now, we must consider two cases for the polarization of the incoming wave: *in (parallel to) the plane of incidence* or *perpendicular to the plane of incidence*, also termed *transverse magnetic (TM)* and *transverse electric (TE)* for obvious reasons:

in (parallel to) the plane of incidence or transverse magnetic (TM)

$$\vec{E}_{0,i}\cdot\widehat{u}=0 \qquad \vec{H}_{0,i}\cdot\begin{bmatrix}\widehat{n}\\ \widehat{w}\end{bmatrix}=0 \qquad (9.76)$$

perpendicular to the plane of incidence or transverse electric (TE)

$$\vec{E}_{0,i}\cdot\begin{bmatrix}\widehat{n}\\ \widehat{w}\end{bmatrix}=0 \qquad \vec{H}_{0,i}\cdot\widehat{u}=0 \qquad (9.77)$$

In (parallel to) the plane of incidence (TM):
When $\vec{E}_{0,i}$ is in the plane of incidence, we can decompose it into pieces along $\widehat{w}$ and along $\widehat{n}$. There is freedom on the sign convention, and we choose

$$\vec{E}_{0,i} = \widetilde{E}_{0,i}\,(\widehat{w}\,\cos\theta_i - \widehat{n}\sin\theta_i) \tag{9.78}$$

We pick the conventions for $\vec{E}_{0,r}$ and $\vec{E}_{0,t}$ so that all three electric field vectors align for normal incidence. (The $\vec{H}$ orientations are then defined by this choice and the direction of the corresponding $\widehat{k}$ vectors.) With these choices, we then have (remember, $\vec{H} = \widehat{k} \times \vec{E}/Z$):

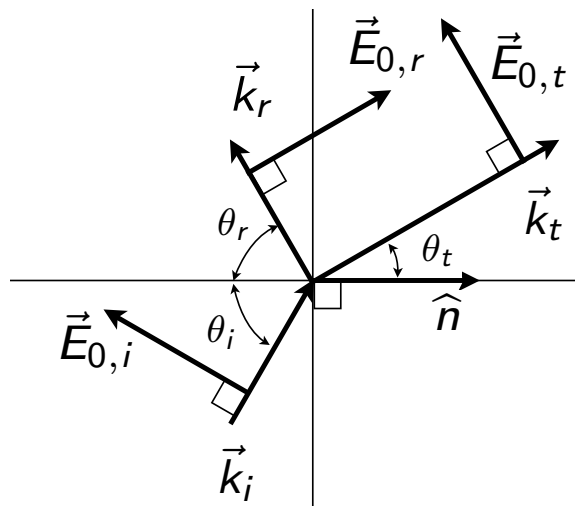
$$\vec{E}_{0,r} = \widetilde{E}_{0,r}\,(\widehat{w}\,\cos\theta_r + \widehat{n}\sin\theta_r) \tag{9.79}$$

$$\vec{E}_{0,t} = \widetilde{E}_{0,t}\,(\widehat{w}\,\cos\theta_t - \widehat{n}\sin\theta_t) \tag{9.80}$$

$$\vec{H}_{0,i} = \frac{\widetilde{E}_{0,i}}{Z_1}\,\widehat{u} \tag{9.81}$$

$$\vec{H}_{0,r} = -\frac{\widetilde{E}_{0,r}}{Z_1}\,\widehat{u} \tag{9.82}$$

$$\vec{H}_{0,t} = \frac{\widetilde{E}_{0,t}}{Z_2}\,\widehat{u} \tag{9.83}$$



So, restricting to the boundary conditions with information, we obtain

$$\widehat{n}: \qquad \epsilon_1\left(-\widetilde{E}_{0,i}\,\sin\theta_i + \widetilde{E}_{0,r}\,\sin\theta_r\right) = -\epsilon_2\widetilde{E}_{0,t}\,\sin\theta_t \tag{9.84}$$

$$\widehat{w}: \qquad \widetilde{E}_{0,i}\,\cos\theta_i + \widetilde{E}_{0,r}\,\cos\theta_r = \widetilde{E}_{0,t}\,\cos\theta_t \tag{9.85}$$

$$\widehat{u}: \qquad \frac{1}{Z_1}\left(\widetilde{E}_{0,i} - \widetilde{E}_{0,r}\right) = \frac{1}{Z_2}\,\widetilde{E}_{0,t} \tag{9.86}$$

Since the incident wave amplitude must be allowed to be arbitrary, we expect to only be able to determine the ratios $\widetilde{E}_{r,0}/\widetilde{E}_{i,0}$ and $\widetilde{E}_{t,0}/\widetilde{E}_{i,0}$. Only two of the above equations can therefore be independent. (One can easily see that the $\widehat{n}$ and $\widehat{u}$ equations are equivalent via Snell's Law.) Picking the last two because they are easiest to work with, we obtain *Fresnel's Equations in (parallel to) the plane of incidence (for TM waves)*:

$$\alpha = \frac{\cos\theta_t}{\cos\theta_i} \qquad \beta = \frac{Z_1}{Z_2} \tag{9.87}$$

$$\boxed{\frac{\widetilde{E}_{0,r}}{\widetilde{E}_{0,i}} = \left(\frac{\alpha-\beta}{\alpha+\beta}\right) \qquad \frac{\widetilde{E}_{0,t}}{\widetilde{E}_{0,i}} = \left(\frac{2}{\alpha+\beta}\right) \qquad \textit{Fresnel's Equations in (parallel to) the plane of incidence (TM)}} \tag{9.88}$$

Perpendicular to the plane of incidence (TE):
When $\vec{E}_{0,i}$ is perpendicular to the plane of incidence, it must be parallel to $\widehat{u}$ because $\widehat{u}$ defines the normal to that plane (again, up to a sign choice.) Now $\vec{H}_{0,i}$ is in the plane of incidence. We have

$$\vec{E}_{0,i} = \widetilde{E}_{0,i}\widehat{u} \tag{9.89}$$

We again pick the convention so that all three electric field vectors align for normal incidence. With these choices, we then have (again, $\vec{H} = \widehat{k} \times \vec{E}/Z$):

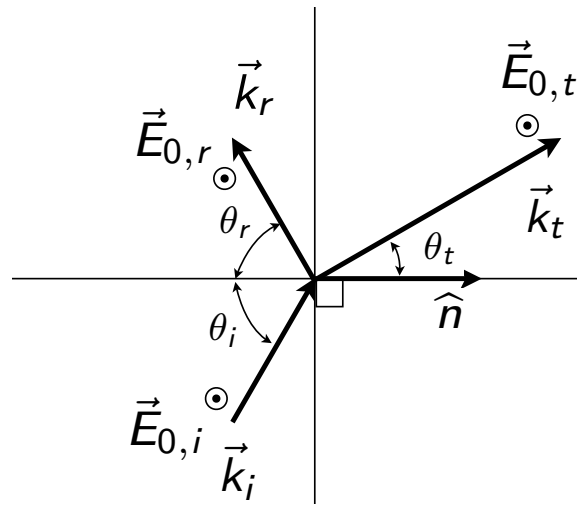
$$\vec{E}_{0,r} = \widetilde{E}_{0,r}\widehat{u} \tag{9.90}$$

$$\vec{E}_{0,t} = \widetilde{E}_{0,t}\widehat{u} \tag{9.91}$$

$$\vec{H}_{0,i} = \frac{\widetilde{E}_{0,i}}{Z_1}\left(-\widehat{w}\,\cos\theta_i + \widehat{n}\,\sin\theta_i\right) \tag{9.92}$$

$$\vec{H}_{0,r} = \frac{\widetilde{E}_{0,r}}{Z_1}\left(\widehat{w}\,\cos\theta_r + \widehat{n}\,\sin\theta_r\right) \tag{9.93}$$

$$\vec{H}_{0,t} = \frac{\widetilde{E}_{0,t}}{Z_2}\left(-\widehat{w}\,\cos\theta_t + \widehat{n}\,\sin\theta_t\right) \tag{9.94}$$



Restricting to the boundary conditions with information, we obtain

$$\widehat{n}: \qquad \frac{\mu_1}{Z_1}\left(\widetilde{E}_{0,i}\,\sin\theta_i + \widetilde{E}_{0,r}\,\sin\theta_r\right) = \frac{\mu_2}{Z_2}\,\widetilde{E}_{0,t}\,\sin\theta_t \tag{9.95}$$

$$\widehat{w}: \qquad \frac{1}{Z_1}\left(\widetilde{E}_{0,i}\,\cos\theta_i - \widetilde{E}_{0,r}\,\cos\theta_r\right) = \frac{1}{Z_2}\,\widetilde{E}_{0,t}\,\cos\theta_t \tag{9.96}$$

$$\widehat{u}: \qquad \widetilde{E}_{0,i} + \widetilde{E}_{0,r} = \widetilde{E}_{0,t} \tag{9.97}$$

Again, we only need two of the equations, so we use the latter two because the $\widehat{n}$ equation can be reduced to the $\widehat{u}$ equation via Snell's Law. We solve to obtain *Fresnel's Equations perpendicular to the plane of incidence (TE)*:

$$\boxed{\frac{\widetilde{E}_{0,r}}{\widetilde{E}_{0,i}} = \left(\frac{1-\alpha\,\beta}{1+\alpha\,\beta}\right) \qquad \frac{\widetilde{E}_{0,t}}{\widetilde{E}_{0,i}} = \left(\frac{2}{1+\alpha\,\beta}\right) \qquad \textit{Fresnel's Equations perpendicular to the plane of incidence (TE)}} \tag{9.98}$$

with α and β as defined earlier.

Reflected and Transmitted Energy

The energy flux (intensity) at a particular point is

$$\mathrm{I}_j = \left\langle |\vec{S}_j| \right\rangle = \frac{1}{2}\,\epsilon_j\, v_j\, E_j^2 \cos\theta_j = \frac{1}{2}\,\frac{c}{Z_j}\, E_j^2 \cos\theta_j \tag{9.99}$$

We can calculate from this the reflected and transmitted energy or power ratios:

$$\mathcal{R} = \frac{\mathrm{I}_r}{\mathrm{I}_i} = \left(\frac{\widetilde{E}_{0,r}}{\widetilde{E}_{0,i}}\right)^2 = \left(\frac{\alpha-\beta}{\alpha+\beta}\right)^2 \qquad \text{parallel (TM)} \tag{9.100}$$

$$= \left(\frac{1-\alpha\,\beta}{1+\alpha\,\beta}\right)^2 \qquad \text{perpendicular (TE)} \tag{9.101}$$

$$\mathcal{T} = \frac{\mathrm{I}_t}{\mathrm{I}_i} = \frac{Z_1}{Z_2}\left(\frac{\widetilde{E}_{0,t}}{\widetilde{E}_{0,i}}\right)^2 \frac{\cos\theta_t}{\cos\theta_i} = \alpha\,\beta\left(\frac{2}{\alpha+\beta}\right)^2 \qquad \text{parallel (TM)} \tag{9.102}$$

$$= \alpha\,\beta\left(\frac{2}{1+\alpha\,\beta}\right)^2 \qquad \text{perpendicular (TE)} \tag{9.103}$$

By calculating $\mathcal{R}+\mathcal{T}$ explicitly, one can see that $\mathcal{R}+\mathcal{T}=1$ always in both cases. Notice the $\alpha\,\beta$ prefactor for $\mathcal{T}$.

Normal Incidence

Let's summarize the results for normal incidence, $\theta_i = \theta_r = \theta_t = 0$, for which things simplify substantially: $\alpha = 1$ and thus the TM and TE cases are equivalent, yielding:

$$\frac{\widetilde{E}_{0,r}}{\widetilde{E}_{0,i}} = \frac{1-\beta}{1+\beta} = \frac{1-\frac{Z_1}{Z_2}}{1+\frac{Z_1}{Z_2}} \qquad \mathcal{R} = \left(\frac{1-\beta}{1+\beta}\right)^2 = \left(\frac{1-\frac{Z_1}{Z_2}}{1+\frac{Z_1}{Z_2}}\right)^2 \tag{9.104}$$

$$\frac{\widetilde{E}_{0,t}}{\widetilde{E}_{0,i}} = \frac{2}{1+\beta} = \frac{2}{1+\frac{Z_1}{Z_2}} \qquad \mathcal{T} = \beta\left(\frac{2}{1+\beta}\right)^2 = \frac{Z_1}{Z_2}\left(\frac{2}{1+\frac{Z_1}{Z_2}}\right)^2 \tag{9.105}$$

We will see that we get similar equations for transmission lines and waveguides. The concept of "impedance matching" becomes apparent: when the wave impedances of the two media match, then $\mathcal{R} = 0$.

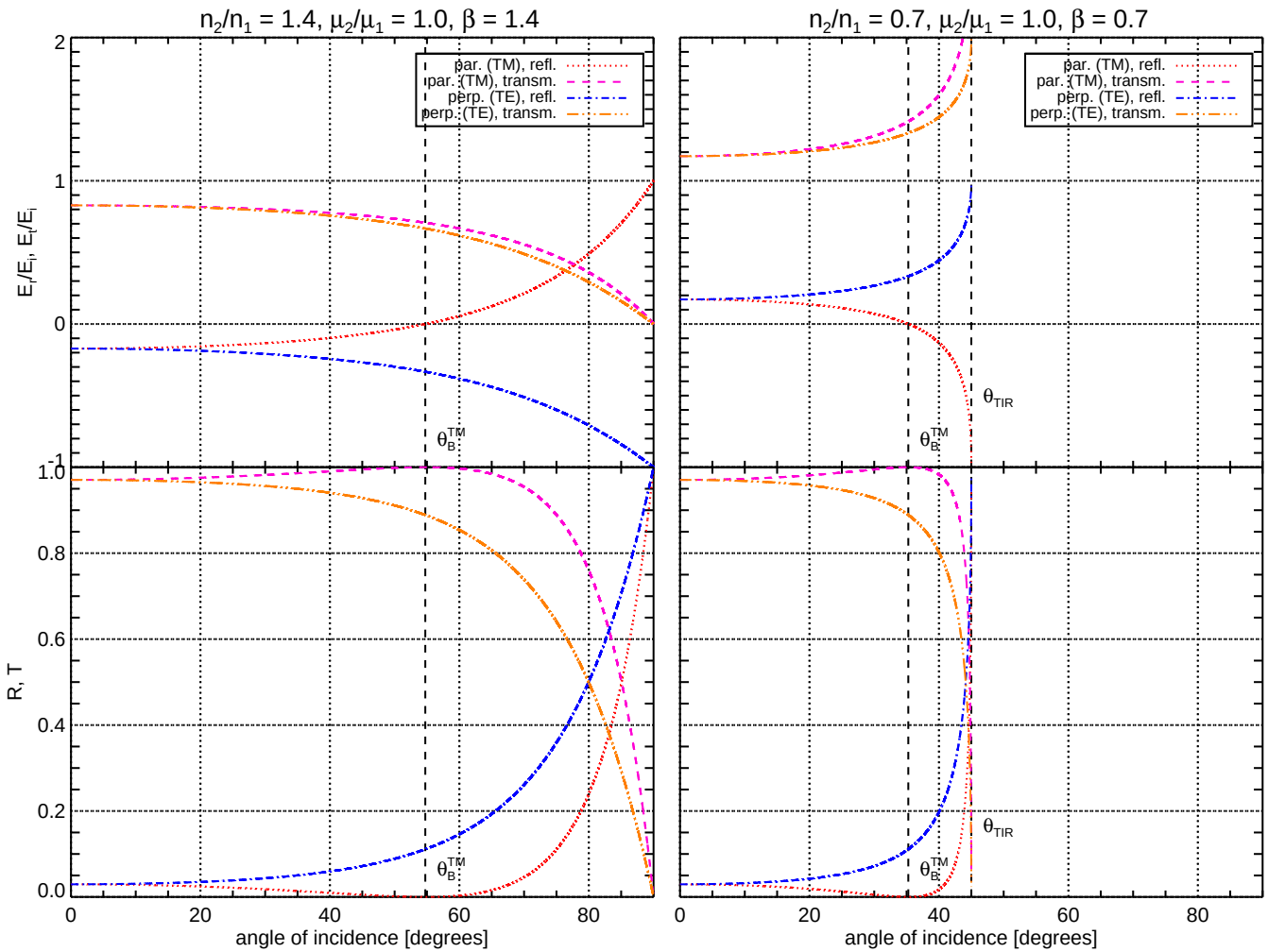
For the special but frequently encountered case $\mu_1 = \mu_2 = \mu_0$, $\beta = Z_1/Z_2$ reduces to $n_2/n_1 = v_1/v_2$ and we recover the results for a wave on a string from Ph2/12:

$$\frac{\widetilde{E}_{0,r}}{\widetilde{E}_{0,i}} = \frac{n_1-n_2}{n_1+n_2} = \frac{v_2-v_1}{v_2+v_1} \qquad \mathcal{R} = \left(\frac{n_2-n_1}{n_2+n_1}\right)^2 = \left(\frac{v_2-v_1}{v_2+v_1}\right)^2 \tag{9.106}$$

$$\frac{\widetilde{E}_{0,t}}{\widetilde{E}_{0,i}} = \frac{2\,n_1}{n_1+n_2} = \frac{2\,v_2}{v_2+v_1} \qquad \mathcal{T} = \frac{n_2}{n_1}\left(\frac{2\,n_1}{n_1+n_2}\right)^2 = \frac{v_1}{v_2}\left(\frac{2\,v_2}{v_2+v_1}\right)^2 \tag{9.107}$$

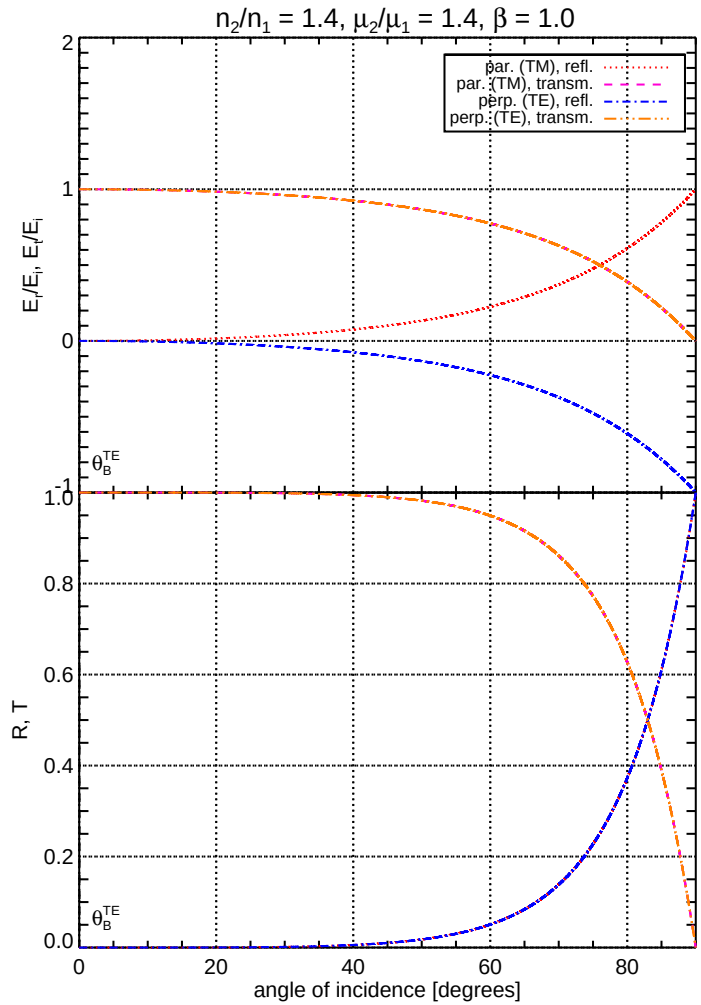
Typical Behavior

We first show the simplest case $\mu_1/\mu_2 = 1$, which implies $n_2/n_1 = (Z_2/Z_1)^{-1} = \beta$.

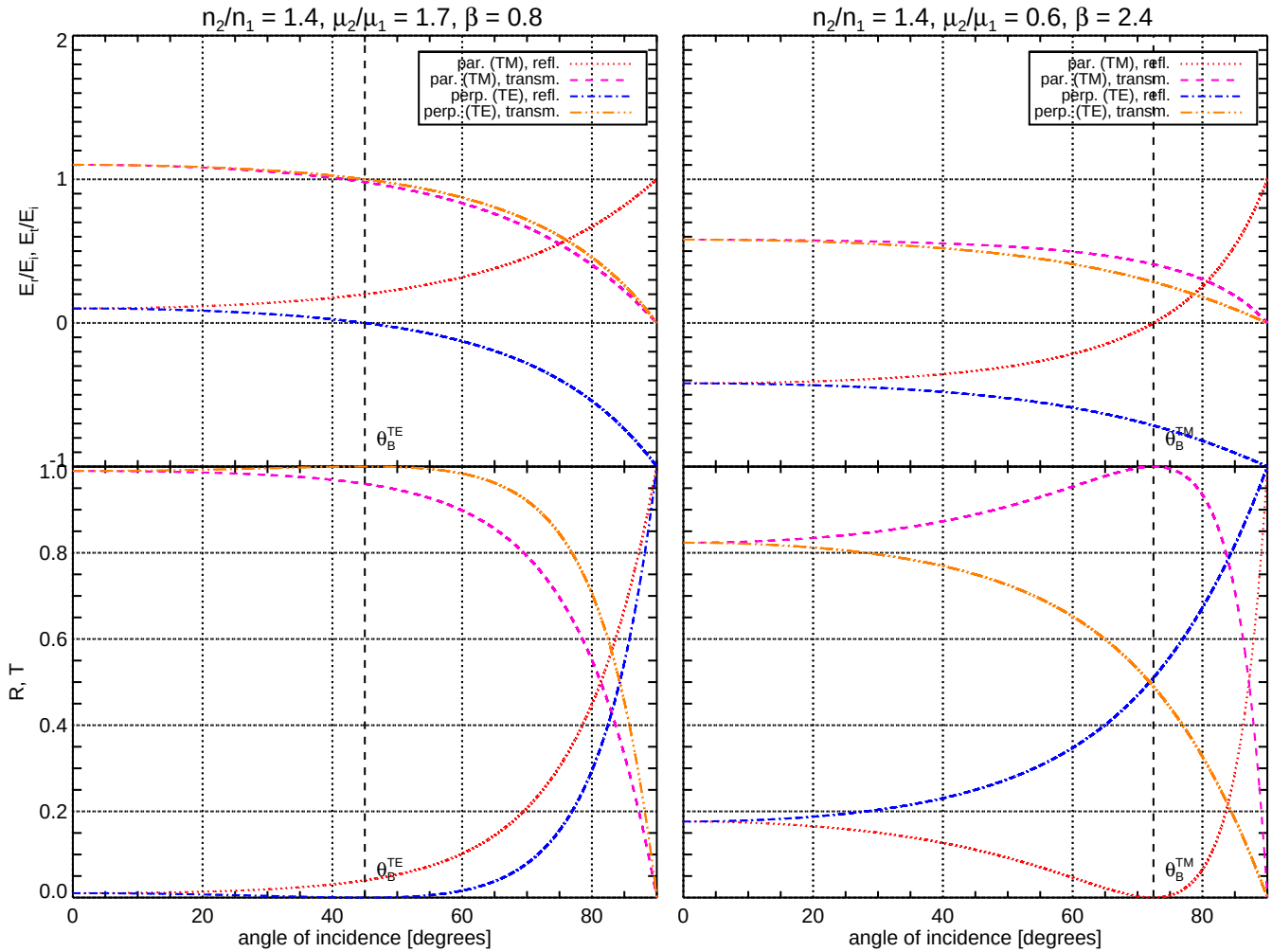


Notice how, for the $n_2/n_1 < 1$ case, the transmitted field amplitude exceeds unity but the transmitted power $\mathcal{T}$ does not and, in particular, how the transmitted field amplitude appears to diverge but the transmitted energy vanishes at $\theta_{TIR} = \sin^{-1} n_2/n_1$. These behaviors are simultaneously possible because of the $\alpha\beta$ pre-factor in $\mathcal{T}$. For both cases of n_2/n_1, notice also the nonmonotonic behavior in $\mathcal{R}$ and $\mathcal{T}$ near the Brewster Angle (θ_B, to be defined later) for the parallel incidence case: this behavior occurs because the reflected field, while monotonic, passes through zero, requiring the squared quantity $\mathcal{R}$ to be positive, zero, and then positive again. Note also that the transmitted field does not pass through unity magnitude even though $\mathcal{T}$ does. Both this behavior, and $\mathcal{T}$'s non-monoticity, occur because of the α prefactor, which starts at 1 at normal incidence and then grows with θ_i.

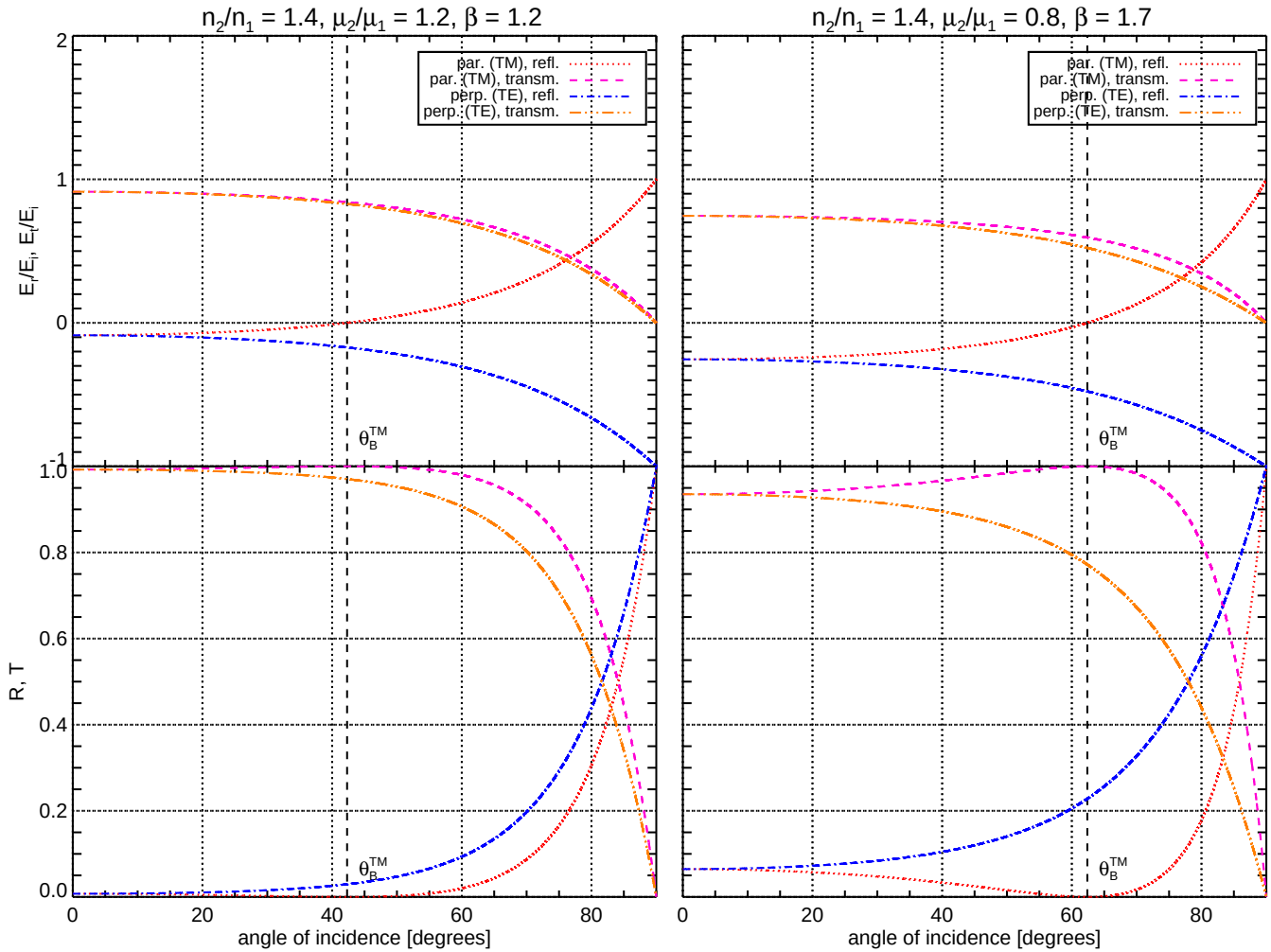
Now we allow $\mu_2/\mu_1 \neq 1$. In general, $\beta = Z_1/Z_2 = (n_2/n_1)\,/\,(\mu_2/\mu_1)$, so we first take a value for $n_2/n_1 > 1$ and consider different cases for μ_2/μ_1 such that all three cases for $\beta - 1$ are explored, and then we do the same for $n_2/n_1 < 1$. In the $\beta = 1.0$ case ($\mu_2/\mu_1 = n_2/n_1$), we have impedance matching and thus unity transmission at normal incidence, but the deviation of μ_2/μ_1 from unity results in nontrivial behavior away from normal incidence ($\alpha \neq 1$). Notice how $\theta_B^{TM} = \theta_B^{TE} = 0$ for $\beta = 1.0$ and how the TE and TM behavior in energy are identical (though not in field). There is no θ_{TIR} because $n_2/n_1 > 1$.



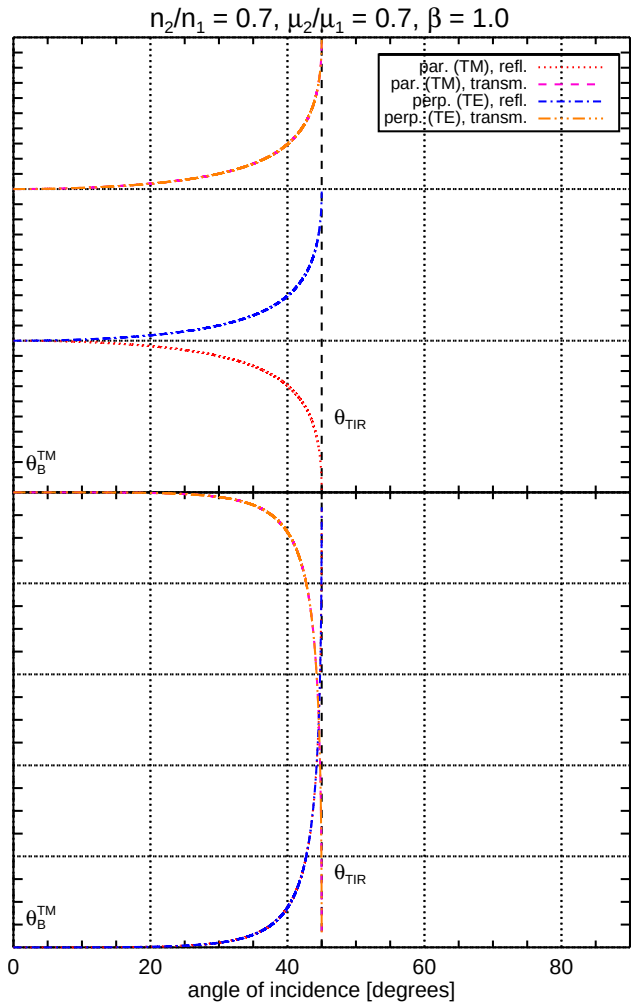
We now consider values of μ_2/μ_1 that change the sign of $\beta - 1$ for the chosen value of n_2/n_1. θ_B^{TE} appears for $\beta = Z_1/Z_2 < 1$ and θ_B^{TM} for $\beta > 1$. Again, we have no θ_{TIR} because $n_2/n_1 > 1$.



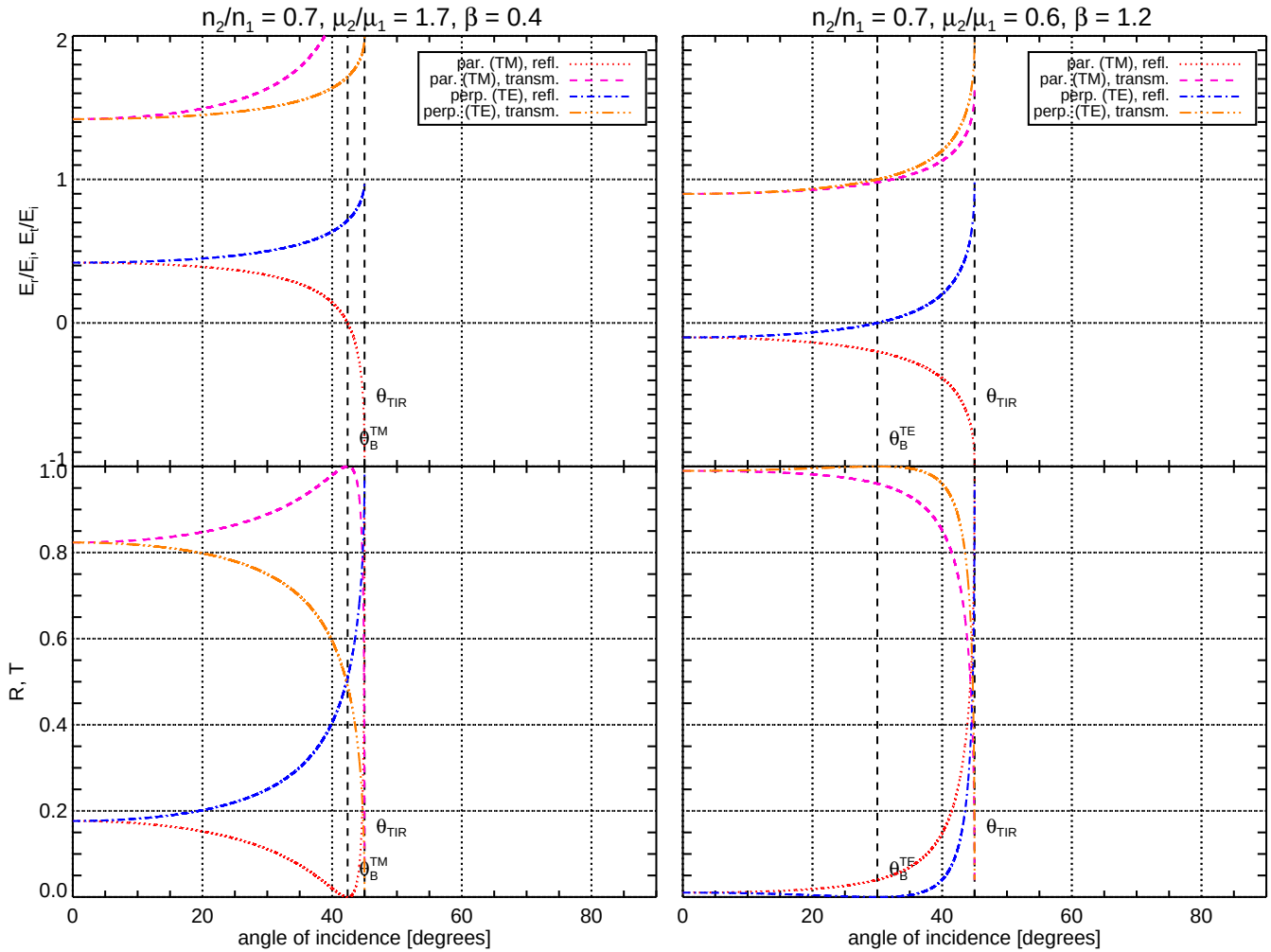
In this case, we hold the sign of $\beta - 1$ fixed but allow the sign of $\mu_2/\mu_1 - 1$ to change. We see it is not μ_2/μ_1 that matters but rather it is $\beta = Z_1/Z_2$.



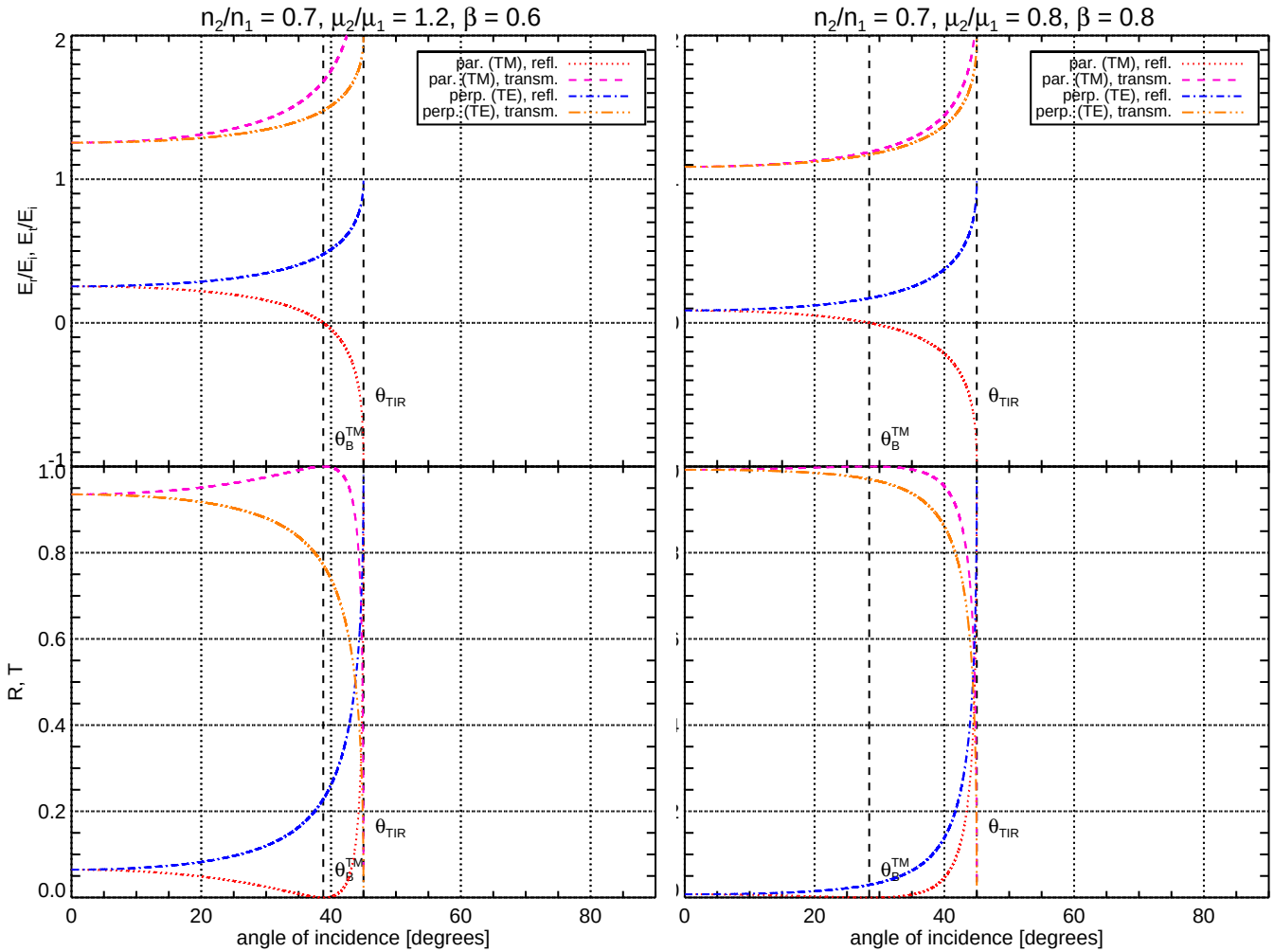
We now take $n_2/n_1 < 1$, which yields a θ_{TIR} for these and the remaining cases. Here, we again consider the $\beta = 1$ special case, for which we again see $\theta_B^{TM} = \theta_B^{TE} = 0$.



We now change the sign of $\mu_2/\mu_1 - 1$ to change the sign of $\beta - 1$, yielding θ_B^{TM} for $\beta < 1$ and θ_B^{TE} for $\beta > 1$.



Here we again see how β is more important than μ_2/μ_1: we change the sign of $\mu_2/\mu_1 - 1$ while holding the sign of $\beta - 1$ fixed and we see the behavior is qualitatively unchanged.



Understanding the Signs and Magnitudes of the Transmitted and Reflected Fields

- *Sign of transmitted wave*
  To understand the sign of the transmitted wave, we just need to notice that α and β are always positive numbers. α is always positive because θ_i and θ_r are restricted to the first quadrant. Therefore, all the quantities in the expressions for $\widetilde{E}_{0,t}$ are positive, and thus the transmitted wave always has the same sign electric field as the incident wave.

  Since $\widehat{k}_i \cdot \widehat{n}$ and $\widehat{k}_t \cdot \widehat{n}$ have the same sign and $\vec{H} \propto \widehat{k} \times \vec{E}$, we may also conclude that the sign of the magnetic field of the transmitted wave is unchanged.

- *Sign of electric field of reflected wave (general case) and Brewster's Angle*
  The sign of the reflected wave depends on the sizes of α and β. What general statements can we make?

  We know that, for a given pair of media, either $\sin\theta_t/\sin\theta_i < 1$ or $\sin\theta_t/\sin\theta_i > 1$ is true for all angles because this ratio is set by Snell's Law, $\sin\theta_t/\sin\theta_i = n_1/n_2$. Since sin and cos are monotonic over the first quadrant, we can conclude that $\alpha = \cos\theta_t/\cos\theta_i$ is also either smaller than or greater than 1 for all angles, with the case being determined by n_1/n_2:

$$\frac{n_1}{n_2} > 1 \iff \frac{\sin\theta_t}{\sin\theta_i} > 1 \iff \frac{\cos\theta_t}{\cos\theta_i} < 1 \iff \alpha < 1 \qquad (9.108)$$

$$\frac{n_1}{n_2} < 1 \iff \frac{\sin\theta_t}{\sin\theta_i} < 1 \iff \frac{\cos\theta_t}{\cos\theta_i} > 1 \iff \alpha > 1 \qquad (9.109)$$

  However, $\beta = Z_1/Z_2 = (n_2/n_1)/(\mu_2/\mu_1)$, so the size of n_2/n_1 relative to unity does not completely determine the size of β relative to unity. No generic statement can be made about the relative size of α, β, and $1/\beta$.

That said, we can make conditional statements that depend on whether we are considering the parallel (TM) or perpendicular (TE) incidence cases and whether β is smaller or larger than unity. We break this up into two steps: what is the sign of the reflection at normal incidence, and whether that sign changes as the angle of incidence changes:

- Sign of reflected wave at normal incidence, $\theta_i = \theta_t = 0$ (parallel (TM) and perpendicular (TE) incidence degenerate at this angle)
  The sign of the denominator of the expression for the reflected amplitude is always positive, and $\alpha = 1$ because $\theta_i = \theta_t = 0$ at normal incidence, so the sign of the reflection at normal incidence is set by the sign of the numerator $1 - \beta$, which is set by the ratio of the wave impedances:

$$\text{if} \quad \beta = \frac{n_2/n_1}{\mu_2/\mu_1} = \frac{Z_1}{Z_2} > 1 \quad \Longrightarrow \quad \text{sign}\left(\frac{\widetilde{E}_{0,r}}{\widetilde{E}_{0,i}}\right)_{\theta_i=0} < 0 \qquad (9.110)$$

$$\text{if} \quad \beta = \frac{n_2/n_1}{\mu_2/\mu_1} = \frac{Z_1}{Z_2} < 1 \quad \Longrightarrow \quad \text{sign}\left(\frac{\widetilde{E}_{0,r}}{\widetilde{E}_{0,i}}\right)_{\theta_i=0} > 0 \qquad (9.111)$$

- Does the sign of the reflected wave change with angle of incidence?
  We can answer this question by determining whether there is an angle at which zero reflection occurs. If so, the sign of the reflected wave will change from its sign at normal incidence to the opposite sign. Whether there is such an angle depends on the relative sizes of α and β. We will call this angle *Brewster's Angle*, though that nomenclature generally only applies for the case $\mu_1 = \mu_2 = \mu_o$, which we will consider shortly.

- *Parallel incidence (TM)*
  The condition for zero reflection is $\alpha = \beta$. If we square this requirement, use a trigonometric identity, and use Snell's law, we find

$$\frac{1 - \frac{n_1^2}{n_2^2}\sin^2\theta_B}{1 - \sin^2\theta_B} = \beta^2 \quad \Longleftrightarrow \quad \sin^2\theta_B = \frac{\beta^2 - 1}{\beta^2 - \frac{n_1^2}{n_2^2}} \tag{9.112}$$

  It is clear that the condition is driven by $\beta = Z_1/Z_2$, the ratio of the wave impedances, and n_1/n_2, the ratio of the indices of refraction. We have to consider multiple cases to determine when there is a solution for θ_B, which can only happen if the above expression takes on a value between 0 (numerator and denominator have same sign) and 1 (numerator smaller in magnitude than denominator):

$$\text{if} \quad \beta = \frac{n_2/n_1}{\mu_2/\mu_1} = \frac{Z_1}{Z_2} > 1 \implies \text{need } \frac{n_2}{n_1} > 1,\ \frac{n_2}{n_1} > \frac{\mu_2}{\mu_1} \tag{9.113}$$

$$\text{if} \quad \beta = \frac{n_2/n_1}{\mu_2/\mu_1} = \frac{Z_1}{Z_2} < 1 \implies \text{need } \frac{n_2}{n_1} < 1,\ \frac{n_2}{n_1} < \frac{\mu_2}{\mu_1} \tag{9.114}$$

- *Perpendicular incidence (TE)*
  The condition for zero reflection is $\alpha = 1/\beta$. Repeating the same process to find a solution, we have

$$\frac{1 - \frac{n_1^2}{n_2^2}\sin^2\theta_B}{1 - \sin^2\theta_B} = \frac{1}{\beta^2} \quad \Longleftrightarrow \quad \sin^2\theta_B = \frac{\frac{1}{\beta^2} - 1}{\frac{1}{\beta^2} - \frac{n_1^2}{n_2^2}} \tag{9.115}$$

  This is basically just the converse of the TM case, so the same set of requirements for a solution for θ_B implies

$$\text{if} \quad \beta = \frac{n_2/n_1}{\mu_2/\mu_1} = \frac{Z_1}{Z_2} > 1 \implies \text{need } \frac{n_2}{n_1} < 1,\ \frac{n_2}{n_1} > \frac{\mu_2}{\mu_1} \tag{9.116}$$

$$\text{if} \quad \beta = \frac{n_2/n_1}{\mu_2/\mu_1} = \frac{Z_1}{Z_2} < 1 \implies \text{need } \frac{n_2}{n_1} > 1,\ \frac{n_2}{n_1} < \frac{\mu_2}{\mu_1} \tag{9.117}$$

Note how the pairing of the two n_2/n_1 conditions changes between the TM and TE cases. The above behavior was seen in the plots prior to this section.

- *Sign of electric field of reflected wave, and Brewster's Angle, for $\mu_1 = \mu_2$*
  This case is typical for everyday experience; there are few light-transmitting yet also $\mu \neq \mu_0$ materials. We can be much more specific in this case because now $\beta = Z_1/Z_2 = n_2/n_1$ and there is a clear relationship between α and β, which we may now rewrite as

$$\frac{n_1}{n_2} > 1 \quad \Longleftrightarrow \quad \beta = \frac{n_2}{n_1} < 1 \quad \text{and} \quad \alpha = \frac{\cos\theta_t}{\cos\theta_i} < 1 \tag{9.118}$$

$$\frac{n_1}{n_2} < 1 \quad \Longleftrightarrow \quad \beta = \frac{n_2}{n_1} > 1 \quad \text{and} \quad \alpha = \frac{\cos\theta_t}{\cos\theta_i} > 1 \tag{9.119}$$

  Considering the two cases separately:

  - *Parallel incidence (TM), $\mu_1 = \mu_2$*
    Since α and β are both either < 1 or > 1, it is possible for $\alpha = \beta$ to be true and therefore the sign may depend on the angle. At $\theta_i = \theta_t = 0$, we have $\alpha = 1$ identically, so the sign of the reflected wave is $1 - \beta$, which is positive if $\beta < 1$ and negative if $\beta > 1$. This sets the polarity of the reflected wave for small θ_i.

    To understand the polarity of the reflected wave for larger $\theta_i \neq 0$, we then have to ask whether it is possible for the polarity of $\alpha - \beta$ to change as α varies with θ_i, which would happen if there is a zero in $\alpha - \beta$.

To solve for this angle where $\alpha = \beta$ (*Brewster's Angle* again, θ_B), we can square the ratio of cosines so it can be written in terms of sines, then use Snell's Law to obtain:

$$\frac{1 - \frac{n_1^2}{n_2^2}\sin^2\theta_B}{1 - \sin^2\theta_B} = \frac{n_2^2}{n_1^2} \quad \Longleftrightarrow \quad \sin^2\theta_B = \frac{n_2^2}{n_1^2 + n_2^2} \tag{9.120}$$

$$\Longleftrightarrow \quad \boxed{\tan\theta_B = \frac{n_2}{n_1} = \beta = \frac{Z_1}{Z_2} \qquad \begin{array}{l}\textit{Brewster's}\\ \textit{Angle for}\\ \mu_1 = \mu_2\end{array}} \tag{9.121}$$

Note that $0 < \theta_B < \pi/4$ for $n_2 < n_1$ and $\pi/4 < \theta_B < \pi/2$ for $n_2 > n_1$. Therefore, we may summarize this case as (using $1 - \beta = (n_1 - n_2)/n_1$):

sign of reflected wave for parallel incidence (TM) and $\mu_1 = \mu_2$:

$$0 < \theta_i < \theta_B : \quad \text{sign}\left(\frac{\widetilde{E}_{0,r}}{\widetilde{E}_{0,i}}\right) = -\text{sign}\left(\frac{n_2}{n_1} - 1\right) \tag{9.122}$$

$$\theta_B < \theta_i < \theta_{max} : \quad \text{sign}\left(\frac{\widetilde{E}_{0,r}}{\widetilde{E}_{0,i}}\right) = \text{sign}\left(\frac{n_2}{n_1} - 1\right) \tag{9.123}$$

$$\theta_{max} = \sin^{-1}\left(\min\left(1, \frac{n_2}{n_1}\right)\right) \tag{9.124}$$

- *Perpendicular incidence (TE), $\mu_1 = \mu_2$*
  The perpendicular incidence case is easier to analyze. At normal incidence, $\alpha = 1$ again and the formulae become identical to the parallel incidence case, giving us the same behavior: the reflected wave is positive if $\beta < 1$ and negative if $\beta > 1$. This common behavior must hold, as the parallel and perpendicular cases are degenerate for normal incidence.

  To see if there is an angle at which the sign of the reflected wave can flip, we need to know if there is an angle at which the reflected wave vanishes. It is easy to see there is not: for $1 - \alpha\,\beta$ to vanish, we require $\alpha = 1/\beta$. For the case $\mu_1 = \mu_2$, we thus require $\alpha = n_1/n_2$. But we saw above that, when $1/\beta = n_1/n_2 > 1$ we have $\alpha < 1$ and when $1/\beta = n_1/n_2 < 1$ we have $\alpha > 1$. Thus, there is never a zero in the reflected wave, and the sign can never flip.

  Therefore, we may summarize this case as:

  sign of reflected wave for perpendicular incidence (TE) and $\mu_1 = \mu_2$:

$$0 < \theta_i < \theta_{max}: \quad \text{sign}\left(\frac{\widetilde{E}_{0,r}}{\widetilde{E}_{0,i}}\right) = -\text{sign}\left(\frac{n_2}{n_1} - 1\right) \tag{9.125}$$

$$\theta_{max} = \sin^{-1}\left(\min\left(1, \frac{n_2}{n_1}\right)\right) \tag{9.126}$$

- *Practical Implications of Brewster's Angle*
  Because the reflected amplitude for parallel incidence goes through a zero at θ_B, it is small for angles near θ_B. For the everyday materials water ($n = 1.33$) and glass ($n \approx 1.5$), this angle is 53° and 56°, which is a typical viewing angle. This explains why polarized sunglasses reduce glare: by blocking the reflected polarization that is perpendicular to the plane of incidence (TE), they block the only component of the reflected wave that has appreciable amplitude. They are designed to pass the parallel component because it has no reflection near Brewster's angle. Be sure to get the orientation right: if the interface is a horizontal surface, then that plane is actually parallel to the surface for light coming from above, so the plane of parallel incidence is vertical for the viewer.

- *Relation between sign of reflected wave and possibility of total internal reflection*
  Both sets of equations imply that total internal reflection occurs only if $n_2 < n_1$, which also implies that the sign of the reflected wave at normal incidence is equal to the sign of the incident wave. Thus, the two conditions are equivalent:

no sign flip at normal incidence $\iff$ *total internal reflection possible*

- *Sign of magnetic field of reflected wave*
  For any of these cases, we can obtain the sign of the *magnetic field* of the reflected wave by applying the rule $\vec{H} \propto \widehat{k} \times \vec{E}$ to the reflected wave. From this, and from the sign flip of $\widehat{k}_r \cdot \widehat{n}$ relative to $\widehat{k}_i \cdot \widehat{n}$, we can conclude that the magnetic field of the reflected wave has the opposite behavior as the electric field in both the parallel and perpendicular cases: if the electric field receives a sign flip, the magnetic field does not, and vice versa.